

Essays on the Dynamic Decisions of Homeowners and Retail

by

Eduardo Jardim

Department of Economics
Duke University

Date: _____

Approved:

Patrick Bayer, Supervisor

Allan Collard-Wexler

Modibo Sidibe

Christopher Timmins

Arnaud Maurel

Dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in the Department of Economics
in the Graduate School of Duke University
2016

ABSTRACT

Essays on the Dynamic Decisions of Homeowners and Retail

by

Eduardo Jardim

Department of Economics
Duke University

Date: _____

Approved:

Patrick Bayer, Supervisor

Allan Collard-Wexler

Modibo Sidibe

Christopher Timmins

Arnaud Maurel

An abstract of a dissertation submitted in partial fulfillment of the requirements for
the degree of Doctor of Philosophy in the Department of Economics
in the Graduate School of Duke University
2016

Copyright © 2016 by Eduardo Jardim
All rights reserved except the rights granted by the
Creative Commons Attribution-Noncommercial Licence

Abstract

Urban problems have several features that make them inherently dynamic. Large transaction costs all but guarantee that homeowners will do their best to consider how a neighborhood might change before buying a house. Similarly, stores face large sunk costs when opening, and want to be sure that their investment will pay off in the long run. In line with those concerns, different areas of Economics have made recent advances in modeling those questions within a dynamic framework. This dissertation contributes to those efforts.

Chapter 2 discusses how to model an agent's location decision when the agent must learn about an exogenous amenity that may be changing over time. The model is applied to estimating the marginal willingness to pay to avoid crime, in which agents are learning about the crime rate in a neighborhood, and the crime rate can change in predictable (Markovian) ways.

Chapters 3 and 4 concentrate on location decision problems when there are externalities between decision makers. Chapter 3 focuses on the decision of business owners to open a store, when its demand is a function of other nearby stores, either through competition, or through spillovers on foot traffic. It uses a dynamic model in continuous time to model agents' decisions. A particular challenge is isolating the contribution of spillovers from the contribution of other unobserved neighborhood attributes that could also lead to agglomeration. A key contribution of this chapter is showing how we can use information on storefront ownership to help separately

identify spillovers.

Finally, chapter 4 focuses on a class of models in which families prefer to live close to similar neighbors. This chapter provides the first simulation of such a model in which agents are forward looking, and shows that this leads to more segregation than it would have been observed with myopic agents, which is the standard in this literature. The chapter also discusses several extensions of the model that can be used to investigate relevant questions such as the arrival of a large contingent high skilled tech workers in San Francisco, the immigration of hispanic families to several southern American cities, large changes in local amenities, such as the construction of magnet schools or metro stations, and the flight of wealthy residents from cities in the Rust belt, such as Detroit.

I would like to dedicate this to my parents, whose support made all of this possible, and my partner, Katya, who was with me every step of the way.

Contents

Abstract	iv
List of Tables	x
List of Figures	xiii
Acknowledgements	xv
1 Introduction	1
2 Estimating the MWP to Avoid Crime in Changing Neighborhoods	5
2.1 Introduction	5
2.2 Data	9
2.3 House Prices and Past Crimes	16
2.3.1 Identification Strategy	18
2.3.2 Initial results	22
2.4 Model	32
2.4.1 Crimes	34
2.4.2 Agents Expectations	35
2.4.3 Price Equation	37
2.4.4 Estimation Procedure	39
2.5 Results	42
2.5.1 Comparing structural and reduced form estimates	50
2.6 Conclusion	53

3	All in the Mix: Spillovers and the Agglomeration of Neighborhood Retail	55
3.1	Introduction	55
3.1.1	Literature Review	60
3.2	A Static Example of Entry	61
3.3	Data	65
3.3.1	Description	65
3.3.2	Business Clusters	68
3.3.3	Descriptive Statistics	70
3.3.4	Reduced form evidence	73
3.4	A Model of Business Location	77
3.4.1	The Profit Function	80
3.4.2	Landlord's Value Function	81
3.4.3	Business' Value Function	83
3.4.4	Match surplus	84
3.4.5	Rent	86
3.4.6	Entry and exit probabilities	87
3.4.7	Identification	88
3.5	Estimation	90
3.6	Results	93
3.6.1	Spillovers versus unobservables	97
3.7	The effect of joint ownership on the mix of businesses	100
3.8	Subsidies and business density	104
3.9	Conclusions	107
4	A Dynamic Model of Neighborhood Segregation	110
4.1	Introduction	110

4.2	Model	112
4.2.1	Value function approximation	115
4.2.2	The consequences of forward looking behavior	117
4.3	Simulations	118
4.4	Model extensions	121
4.4.1	Changes in demographics	121
4.4.2	Changes in population density	123
4.4.3	Changes in amenities	124
4.5	Conclusion	126
	Bibliography	135
	Biography	139

List of Tables

2.1	Summary of explanatory variables	11
2.2	Crimes in a 0.4 mile buffer	13
2.3	Interquartile transition matrix for total crime Census Tract, 2003-2011	14
2.4	Interquartile transition matrix for violent crime (except robberies) Census Tract, 2003-2011	14
2.5	Summary statistics for yearly crimes Census Tract, 2003-2011	15
2.6	Transition Matrix for Violent Crimes, 0.4 mile, 2 states	42
2.7	Initial Distributions for Violent Crimes, 0.4 mile, 2 states	42
2.8	Lambda for Violent Crimes, 0.4 mile, 2 states	42
2.9	Transition Matrix for Violent Crimes, 0.4 mile, 3 states	43
2.10	Initial Distributions for Violent Crimes, 0.4 mile, 3 states	43
2.11	Lambda for Violent Crimes, 0.4 mile, 3 states	43
2.12	Transition Matrix for Violent Crimes, 0.4 mile, 4 states	43
2.13	Initial Distributions for Violent Crimes, 0.4 mile, 4 states	43
2.14	Lambda for Violent Crimes, 0.4 mile, 4 states	44
2.15	Transition Matrix, 12 months, for Violent Crimes, 0.4 mile, 4 states .	44
2.16	Transition Matrix for Property Crimes, 0.4 mile, 2 states	45
2.17	Initial Distributions for Property Crimes, 0.4 mile, 2 states	45
2.18	Lambda for Property Crimes, 0.4 mile, 2 states	45
2.19	Transition Matrix for Property Crimes, 0.4 mile, 3 states	45

2.20	Initial Distributions for Property Crimes, 0.4 mile, 3 states	45
2.21	Lambda for Property Crimes, 0.4 mile, 3 states	46
2.22	Transition Matrix for Property Crimes, 0.4 mile, 4 states	46
2.23	Initial Distribution for Property Crimes, 0.4 mile, 4 states	46
2.24	Lambda for Property Crimes, 0.4 mile, 4 states	46
2.25	Transition Matrix, 12 months, for Property Crimes, 0.4 mile, 4 states	46
2.26	MWP to avoid a violent crime (x100), coef and t-stat	48
2.27	MWP to avoid a property crime (x100), coef and t-stat	48
2.28	Contribution to price when certain to be in state (%)	49
2.29	MWP to avoid violent crime today (x100), assuming no state transi- tions, coef and t-stat	50
3.1	Number of business-parcel matches by distance (feet)	67
3.2	Total business observations, 2002-2011	67
3.3	Cluster definitions, total storefronts and area (sqft)	70
3.4	Number of surrounding businesses at the time of entry	71
3.5	Description of control variables	71
3.6	Control variables at time of entry	72
3.7	Time until a vacant storefront is leased, by entrant type (in years) . .	72
3.8	Share of storefronts in Census Tracts ranked by different criteria . . .	74
3.9	Regression of number of businesses within 300 ft on entrant type . . .	74
3.10	Cox Proportional Hazard model, 300 ft Exponential of coefficients . .	76
3.11	Cox Proportional Hazard control variables, 300 ft	76
3.12	Cox Proportional Hazard model, 300 ft and 500 ft, exponential of coefficients	78
3.13	Cox Proportional Hazard control variables, 300 ft and 500 ft	79
3.14	Estimates of structural parameters	95

3.15	Intercepts and shares of cluster types	95
3.16	Cox Proportional Hazard model, 300 ft, simulated data	97
3.17	Business type concentration within clusters (HHI)	99
3.18	Difference in share of businesses across simulations, relative to baseline (p.p.)	99
3.19	Neighbors creating positive spillovers to business (%)	100
3.20	Average expected profits and rents for different concentrations of store- front ownership	102
3.21	Average expected profits for baseline and type 3 cluster	103

List of Figures

2.1	Number of Interquartile Transitions by Census Tract, Total Crimes 2003-2011	15
2.2	Number of Interquartile Transitions by Census Tract, Violent Crimes 2003-2011	16
2.3	Absolute change in crimes conditional on changing quartiles, Total Crimes 2003-2011	17
2.4	Absolute change in crimes conditional on changing quartiles, Violent Crimes 2003-2011	18
2.5	Property crimes at a 0.4 mile radius, no FE	22
2.6	Violent crimes at a 0.4 mile radius, no FE	23
2.7	Property crimes at a 0.4 mile radius, Census Block Group by quarter FE	24
2.8	Violent crimes at a 0.4 mile radius, Census Block Group by quarter FE	25
2.9	Property crimes by distance, Census Block Group by quarter FE . . .	27
2.10	Violent crimes by distance, Census Block Group by quarter FE	28
2.11	Crimes at a 0.4 mile radius, Census Block Group by quarter FE . . .	30
2.12	Crimes at a 0.4 mile radius, Census Block Group by quarter FE. (cont.)	31
2.13	Timeline of events	36
2.14	Violent crimes at a 0.4 mile radius, Census Block Group by quarter FE	52
3.1	Entry probabilities as a function of storefront ownership	64
3.2	Three business clusters in San Diego	69
3.3	Shares of neighbors with positive spillovers	103

3.4	Average difference in occupancy rate relative to control simulation . .	105
3.5	Occupancy rate for increasing spillovers, before and after subsidies . .	106
3.6	Occupancy rate for increasing entry costs, before and after subsidies .	107
4.1	Example of a city with a random allocation of agents	128
4.2	City at the start of the simulation	129
4.3	City after 1 cycle	130
4.4	City after 5 cycles	131
4.5	City after 10 cycles	132
4.6	Evolution of utility and share of same-sign neighbors for +1/-1	133
4.7	Evolution of utility and share of same-sign neighbors for +2/-2	134

Acknowledgements

First and foremost, I would like to thank my advisor, Pat Bayer, for guiding me. Pat always had the best questions.

I would also like to thank the other members of my committee, Modibo Sidibe, Allan Collard-Wexler, Chris Timmins, and Arnaud Maurel, for all the comments and suggestions.

I also received great comments from the participants of the Public/IO lunch group, both from professors and from students, and close friends from Duke. I appreciate the time and attention they gave me.

I must thank the Duke Graduate School for summer research support, and the Duke Economics Department for data acquisition funding.

Finally, I would like to thank my family for always being there for me, even if a continent away, and my partner, Katya Roshchina, both for the uncountable ideas and discussions, and for the moments of peace and rest.

1

Introduction

There are several features of the housing market that make buying a house an inherently dynamic process. According to Bayer et al. (2011, pp. 2), those include “(i) large transactions costs that make moves relatively rare, (ii) changing household tastes and needs over the life-cycle, and (iii) evolving local amenities and housing prices that give neighborhoods a dynamic character”. Thus, in light of those features, it is expected that all homebuyers will make their decision with an eye at the future of their neighborhood.

The decision to open a restaurant or a retail store encounters many of the same circumstances. There are high entry costs involved, much of it sunk costs, and long leases, which mean high exit costs as well. Demand in a neighborhood may change over time, which can just as easily benefit or break a business. Finally, the composition of nearby retail stores will also matter for business entry decisions, since neighboring businesses can either bring more demand to the store if there are synergies, or create additional competition. Hence, business owners also face a dynamic problem when deciding where and when to open a store.

In the past decade, several attempts were made to develop dynamic models for

both the housing market and retail entry. Bayer et al. (2011); Bishop and Murphy (2011) present a dynamic sorting model and a dynamic hedonic model. Arcidiacono et al. (2012); Cosman (2014) show how to model dynamic entry decisions of a large retail chain (Wal-Mart), and small nightlife venues, in continuous time. With this thesis, I contribute to this growing literature.

In chapter 2, I discuss how to value agents' revealed preferences for an amenity when (i) there is large, high frequency information available about this amenity, and (ii) there is a concern that the amenity may change over time. I use crime data for the city of Chicago and real estate sales data from Dataquick to show that the detailed information regarding the time and location of a crime can be leveraged to help isolate the effect of crime on house prices from other confounding factors correlated to both.

I propose a method to estimate agent's marginal willingness to pay to avoid crime that allows the probability distribution of crime in neighborhoods to change over time. The method relies on a Latent Markov Model to model the changing distribution of crime at the neighborhood level, and a learning model to describe agent's expectations about a neighborhood at any point in time. I find that agents' willingness to pay for safety is large. A house believed to be in the most dangerous area of the city will sell for 10% less than a house in the safest area, holding all other characteristics constant.

In chapter 3, I look into the retail owner's decision to open a shop. A key feature of the model is that shops can profit from or lose with the presence of other stores. That is, there are spillovers between stores, which will lead some types of stores to locate next to each other, and other types to avoid each other. A few examples of these spillovers are the possibility of sharing foot traffic with other stores, or similar retail stores colocating as a strategy to reduce search costs to costumers (Wolinsky, 1983; Dudey, 1990).

The main challenge in estimating the spillovers is separately identifying them from other neighborhood attributes that will benefit certain types of stores in detriment to others. My work contributes to the literature by showing how information on landlords can be used to help separate spillovers from unobserved neighborhood attributes. The key intuition is that comparing the behavior of two different of landlords in the same location, one who owns multiple storefronts and one who does not, uncovers information on spillovers, while keeping unobserved attributes constant. This follows from the fact that both of these landlords are equally influenced by the neighborhood attributes, but they will be internalizing the spillovers produced by a new tenant differently.

My estimates show that spillovers are particularly important between similar retail stores, and weaker or non-existent between service stores and across stores of different types. These results give credence to the theories suggesting that spillovers are the result of reduced search costs. I use these estimates to study under which conditions the city can use public subsidies to store openings to permanently raise the occupancy rate of storefronts in a neighborhood. I find that subsidized neighborhoods are not better off than non-subsidized ones in the long run. Interestingly, I show that the reason for this is not that spillovers are too weak, but that entry costs are too low. As a consequence, there are always enough stores willing to open and wait to see if other stores will come, and there are no coordination failures for the subsidy to solve.

Finally, in chapter 4, co-authored with Patrick Bayer, Aurel Hizmo, and Marcus Casey, we use a model of household location decisions where families have preferences over their neighbors' identities to revisit the results from Schelling (1971). In his seminal paper, Schelling used a simulation with two types of agents and simple behavioral algorithms to show that even slight preferences for neighbors of the same type could lead to stark segregation patterns in the city. We contribute to this

literature by developing the first version of Schelling's problem with forward looking agents. That is, families make decisions to move or not according not only to the current composition of the neighborhood, but also according to their expectations of how the neighborhood might change in the future.

We show that forward looking agents segregate more and faster than myopic ones. Our model is suitable to discuss the effect of different trends on neighborhood sorting in American urban centers. Examples of those trends are the arrival of high skilled tech workers in San Francisco, the immigration of hispanic families to several southern American cities, large changes in local amenities, such as the construction of magnet schools or metro stations, and the flight of wealthy residents from cities in the Rust belt, such as Detroit. We finish the chapter by discussing several extensions of the model that can be used to study each of those trends.

Estimating the MWP to Avoid Crime in Changing Neighborhoods

2.1 Introduction

Even though crime is one of the most pressing concerns for home buyers in large cities, few papers have attempted to estimate the marginal willingness to pay to avoid crimes. This is not a reflection of low interest in the topic: several papers have looked into the effects of crimes on house prices (see Thaler (1978), Lynch and Rasmussen (2001), Tita et al. (2006), Gibbons (2004), Pope and Pope (2012)). Most likely, the reason these papers avoid using their results to discuss the MWP is that the connection between the coefficients they estimate and the MWP is not obvious at all. Agents value a house according to (their expectation of) the future sequence of crimes for that house. This is in contrast to the estimates in these papers, which use total crimes or change in crimes the year the house was sold. Without a model that can tell us how agents use observed crimes to form expectations about the future, we cannot attempt to estimate the MWP to avoid crime.

This problem is further complicated by two issues. First, observed crimes in

a neighborhood at a single point in time provide very little information about the expected crime for that neighborhood. Thus a panel data of observed crimes is necessary for each house. Second, the expected number of crimes in a neighborhood may change over time. Rational, forward looking buyers take this possibility into account, and so should econometricians.

This paper extends the literature by showing how a panel data of crime at the house level can be used to estimate the MWP to avoid crime.¹ We achieve this while allowing the distribution of crime around each house to change over time. We also show that this assumption is crucial for reproducing some of the patterns we find in the data.

The paper is divided in two parts. We start by estimating the effect of past crimes on house prices. We overcome several difficulties that the previous literature has faced. One of those is omitted variable bias. Violent crimes are more present in poorer areas, with worse street lighting, and where houses are not as well maintained. Property crimes, on the other hand, are more common in wealthier neighborhoods (or at least are reported more frequently there). These are just a few examples of confounders that, if not properly controlled for, will bias coefficient estimates (Gibbons (2004) and Pope and Pope (2012) are two papers that try to address endogeneity concerns).

Other two problems are a result of previous data limitations. Until recently, crime data was available only for large geographical areas, such as census tract and police beat. As a consequence, researchers were forced to use very aggregated data, which introduces attenuation bias if agents only derive dis-utility from more local crimes.²

¹ Bishop and Murphy (2011) use a dynamic hedonic model to estimate the MWP that includes agents expectations, but the model assumes a constant long run probability distribution of crime for all houses.

² Gibbons (2004) is the only exception. The paper only has data on criminal damage and burglaries though.

Similarly, most papers have relied on yearly crime data in their regressions. Time aggregation introduces another measurement error because the crimes before the house sale (observable to an individual) are lumped together with the crimes which happened after the sale and thus, supposedly, should not affect the price directly.

This paper uses new publicly available data on reported crimes in Chicago to overcome such difficulties. We use very detailed information on the date and location of both crimes and house transactions to control for omitted variables and estimate the effect of past crimes and the MWP to avoid crime in Chicago. We use a strategy akin to a placebo treatment to verify that omitted variables do not seem to be driving any of our results, given our controls.

Additionally, we contribute to the literature by presenting two new results. We show that the effect of crimes on property values diminishes with the distance to the property. Violent crimes do not affect the price of houses more than 0.4 mile away in Chicago. Also, we not only focus exclusively on crimes before the sale, as we actually estimate the impact of crimes on a monthly basis. We find that the effect of crimes on property values clearly fades with time.

In the second part of the paper we present a model of the home buying decision and estimate the MWP to avoid crime. The results uncovered in the first part are crucial in this stage. That omitted variables do not interfere with estimates in our identification strategy reassures us that our estimate of the MWP will not be affected either. Moreover, the heterogeneous effect of crime across time differences provides both a guide for constructing our model and a standard which the model estimates can be tested against.

There are two main ideas behind our model. First, the characteristic that describes most precisely a house's safeness is the probability distribution of crime around that house, not crimes in the past month or even past year - these are at most a very noisy proxy for the distribution of crimes. When buying a house,

agents would like to know the sequence of future monthly crimes around the house. Since it is impossible to know this sequence in advance, they will need to learn the probability distribution of crimes to form expectations.

Second, the distribution of crime may change over time, making some houses safer and others more dangerous. Agents will have to account for this accordingly. As a consequence, recent crimes will contain more reliable information about the current distribution of crime than older ones, and thus will have a stronger effect on house prices.

We model crimes as a Latent Markov Model (henceforth LMM; see Bartolucci et al. (2013) and Zucchini and MacDonald (2009)). The LMM allows us to write a simple statistical model of crime composed of three components: a vector with the probability distribution function of crime for a pre-defined set of states, a transition matrix with the probability of transitions between states, and a vector with initial probabilities of each state. This framework has been used traditionally as a statistical model for dealing with measurement errors (see Bartolucci et al. (2013), chapter 1). To our knowledge, this is the first paper to use LMM to model changing amenities with the goal of estimating the MWP.

The model is completed with a stylized version of the buyer's decision problem, where buyers must evaluate each house according to their beliefs over the distribution of crime surrounding the house. Although the LMM and learning components of the model could be easily adapted to more complex dynamic models, such as Bayer et al. (2011), we choose a simpler framework because the estimating equation is closer to what is commonly used in the hedonics literature. This way we can better illustrate the gains to using a LMM for modeling other changing amenities besides crimes, such as water and air quality, or even terrorism.³

³ Leggett and Bockstael (2000) and Arbel et al. (2010) are examples of papers which use datasets suitable for employing the methods developed in this paper.

Assuming four states for violent crimes in Chicago, we find that the probability a state will change between two months is between 0.7% and 1.9%. This implies probabilities of 7.5% to 19.9% over a 12 months period. The MWP to avoid one violent crime is small but significant, and it is enough to explain sizable differences in house prices: a house that agents believe to be in the most dangerous state will sell for 10 percentage points less than a house believed to be in the safest state. Also, we use data simulated from the estimates for the LMM and the MWP to show that the model reproduces the patterns we see between past crimes and house prices in Chicago.

The paper is organized as follows. In section 2.2 we present the crime and house transaction datasets and explain how the data is set up. We also provide evidence that the level of crime in a neighborhood changes over time. We move on to explain our identification strategy for estimating the effect of past crimes on house prices and present results in section 2.3. In section 2.4 we introduce the model and in section 2.5 we present the estimates for the LMM and the MWP. We conclude in section 2.6.

2.2 Data

This paper makes use of two high frequency datasets, with daily information on crime and real estate transactions for Chicago. These datasets are indispensable for both the identification strategy employed in section 2.3 and for estimating the model introduced in section 2.4.

The crime data is made publicly available by the City of Chicago, and lists all reported crimes in the city starting in 2001.⁴ For each reported crime, the dataset

⁴ This is the same data used in Chicago's crime map: <http://gis.chicagopolice.org/> The data can be download at <https://data.cityofchicago.org/>

has detailed information on the time and location of the crime.⁵ This allows us to match crimes to real estate transactions in a much more refined way than earlier papers. Besides the day and hour of the crime and its geographic coordinates, the data also includes information on approximated address, some additional details about the crime (type of weapon used, for instance), beat, ward, and a dummy for whether it was a domestic crime.

We will restrict our attention to the following types of crime: aggravated assault, robbery, sexual crime, homicide, motor vehicle theft (MVT), burglary, and larceny-theft. Certain types of crime can be pretty rare. For this reason in most regressions we aggregate crimes into violent crimes (aggravated assault, homicide, and sexual crime) and property crimes (burglary, MVT, and larceny-theft).⁶ By doing so we can get more precise estimates, at the cost of mixing together crimes of different severity. We still present a regression looking at the effect of each crime type in section 2.3.

A common concern with reported crime data is under-reporting. Comparisons between reported crime and victimization datasets show that under-reporting may be considerable, specially for less serious crimes and in poorer neighborhoods (Skogan (1999)). Whether this will bias our estimates depends on what is the information set available to buyers and sellers. If, for instance, their source of information is Chicago's Crime Map, then they should be reacting only to crimes listed in our data and there is no bias. Bias becomes a concern only if agents have information on some of the unreported crimes.

The data on real estate transactions is called Dataquick. It is a private dataset with a comprehensive list of real estate transactions for several cities starting in the

⁵ Geographic coordinates are approximated by rounding the street number to the nearest hundred. That is, if a crime happened at 135 13th Street, the coordinates will be for 100 13th Street. Since we will focus on crimes at medium distances from a house, this approximation should not be a problem.

⁶ Robberies would usually fit into violent crimes. We decided not to include it though, since it does not seem to affect houses prices in our sample.

late 80s. We will restrict the data to the city of Chicago and to the years of 2003 through 2012. The reason we drop the years 2001 and 2002 is because we want to ensure we have at least 24 months of crime data prior to the date the house is sold, for reasons that will become clear in section 2.4.

For each transaction, Dataquick has information on the sale price, date, characteristics of the house (of which we use number of bathrooms and stories, lot area, constructed area, exterior of the house, roof type, and if there is a fireplace), name of buyer and seller, whether it was an arms length transaction, if it is a single family home, address, and more. Table 2.1 has some descriptive statistics for the non-categorical variables we use and for the area of the census block group that each house belongs to.

Table 2.1: Summary of explanatory variables

	Mean	S.D.	Min.	1st Q	Median	3rd Q	Max.
log(Price)	12.220	0.879	2.303	11.850	12.320	12.740	17.430
N. Bathrooms	2.027	1.277	1	1	2	2	84
N. Stories	1.653	0.593	1	1	2	2	3
Lot Size (sqft)	3808	1532.893	233	3125	3732	4191	122500
Sq. Footage (sqft)	1806	1113.956	400	1080	1508	2250	63200
Census BG Area (sqmi)	0.1200	0.2401	0.0051	0.0620	0.0795	0.1260	8.3560

We keep only arms length transactions and single family homes. Because we think companies and banks may respond differently to crime when buying real estate, we drop several transactions where the buyer’s name indicates a bank, real estate agency, or any other companies. These sales represent a share of 9.3% of the data.⁷ We also drop houses missing information for any of the house characteristics. Unfortunately, this is a considerable share of the remaining data, 38.3%, leaving us with 170886

⁷ Indeed, when we run the regressions from section 2.3 in this subset of our sample, we do not find any patterns, independent of the significance level.

observations.

We should note that some coordinates do not refer to the exact location of the house. For some houses, coordinates refer to the zip-4 centroid. The zip-4 is a more detailed form of the zip code that has 9 instead of the standard 5 digits. Although this situation is not ideal, the area of a zip-4 is small enough that it never covers more than one block, and for big cities it is very common for a single building (or even floors in the building) to get it's own zip-4 number.

We merge the two datasets in the following way. For each house sold, we count the number of crimes of each type that fall within a predetermined distance of the house, for several time intervals of 30 days, defined relative to the transaction date. We count crimes from 749 days (25 months) prior to the sale to 120 days (4 months) after. We index each of the 30 days intervals with $t \in \{4, \dots, -24\}$; $t = 4$ are crimes from 120 to 91 days *after* the sale, $t = 0$ are crimes from 0 to 29 days *before* the sale, and $t = -24$ are crimes from 720 to 749 days *before* the sale. In consonance with how the data is set up, our treatment of time in this paper will always be relative to the day of the sale, unless we say otherwise. This is important not only for interpreting the regressions, but also for the model in section 2.4.

We repeat this process for several different distances (from 0.2 to 0.7 miles, with increments of 0.1), but we will focus mostly at a 0.4 mile radius around the house. The reason we choose 0.4 mile is to increase variation in monthly crimes, even if at the cost of diluting the effect a little. In section 2.3 we show that the effect of crime on house prices does not decrease much with distance up to 0.4 mile, but disappears for any farther crimes.⁸

Table 2.2 shows descriptive statistics for total crimes at a 0.4 mile radius the month preceding the sale. Violent crimes are pretty rare events. The average house

⁸ The area of a 0.4 mile radius circle is around 0.5 square miles, six times the median of a census block group in Chicago.

has only 1.18 aggravated assaults, 0.33 homicides and 0.12 sexual assaults reported monthly. We see from this table that aggravated assaults will be driving most of the variation in violent crimes. Since we do not see a clear effect of robberies on house prices, the decision of excluding it from violent crimes seems correct considering that the median for robberies is 4 times larger than for aggravated assault. Property crimes, on the other hand, are much more common. A house in the 25th percentile already faces 14 property crimes in a month.

Table 2.2: Crimes in a 0.4 mile buffer

	Mean	S.D.	Min.	1st Q	Median	3rd Q	Max.
Assault	1.1800	1.5172	0	0	1	2	17
Homicide	0.3319	0.6428	0	0	0	1	7
Sexual Assault	0.1198	0.3818	0	0	0	0	6
Robbery	3.7090	3.8086	0	1	3	6	37
Violent	5.3410	5.0721	0	1	4	8	45
Violent, no Rob	1.6320	1.9105	0	0	1	2	18
Burglary	6.2190	5.0619	0	2	5	9	62
MVT	4.8760	3.8837	0	2	4	7	33
Theft	16.6600	12.3462	0	8	14	22	226
Property	27.7500	17.6861	0	14	25	38	253

One of our main claims is that neighborhoods change, agents take this into account, and so should econometricians. Before moving on, we show some evidence of this. Instead of using our main dataset, we aggregate reported crimes by year, from 2003 to 2011, for all census tracts in Chicago. This dataset is closer to what most of the literature has used. We start by dividing census tracts into quartiles according to reported crimes each year. Since this is done on a yearly basis, breakpoints change every year. This way the transition matrices will account for any city wide trends in crime.⁹

Tables 2.3 and 2.4 show interquartile transition matrices for total crimes and vio-

⁹ The breakpoints for total crimes are 114, 174, and 268 in 2003 and fall to 90, 138, and 211 in 2011. For violent crimes, they are 3, 6, and 12 in 2003 and fall slightly to 2, 5, and 10 in 2011.

Table 2.3: Interquartile transition matrix for total crime Census Tract, 2003-2011

$t \setminus t + 1$	1	2	3	4
1	0.8426	0.1556	0.0018	0.0000
2	0.1586	0.6760	0.1648	0.0006
3	0.0000	0.1672	0.7358	0.0971
4	0.0000	0.0006	0.0968	0.9026

Table 2.4: Interquartile transition matrix for violent crime (except robberies) Census Tract, 2003-2011

$t \setminus t + 1$	1	2	3	4
1	0.6998	0.2326	0.0632	0.0044
2	0.3366	0.3806	0.2537	0.0291
3	0.0831	0.2413	0.4729	0.2027
4	0.0059	0.0224	0.2163	0.7554

lent crimes, respectively.¹⁰ It is clear that there is a lot of movement of census tracts between different quartiles. Observed crime in one year is a not a good predictor for crime in the next year: between 10% and 33% of census tracts will move to a different quartile of total crimes, depending on their current one. The numbers are much higher for violent crimes.

Part of this movement may be attributed to census tracts that are very close to the breakpoints, and thus change quartiles with only a small increase/decrease in crime. We check how important is this explanation by plotting the histograms of absolute change in total crimes and violent crimes between two years, restricted to combinations of census tract and year that change quartiles. These histograms are presented in figures 2.1 and 2.2. For total crimes, we present the changes in groups of ten. That is, the first bar is for 0 to 9 crimes, the second for 10 to 19, and so on.

More than half of the census tracts that changed total crime quartiles did so

¹⁰ Table 2.4 does not include robberies. Including them does not change the transitions much, resulting in a table more similar to table 2.3.

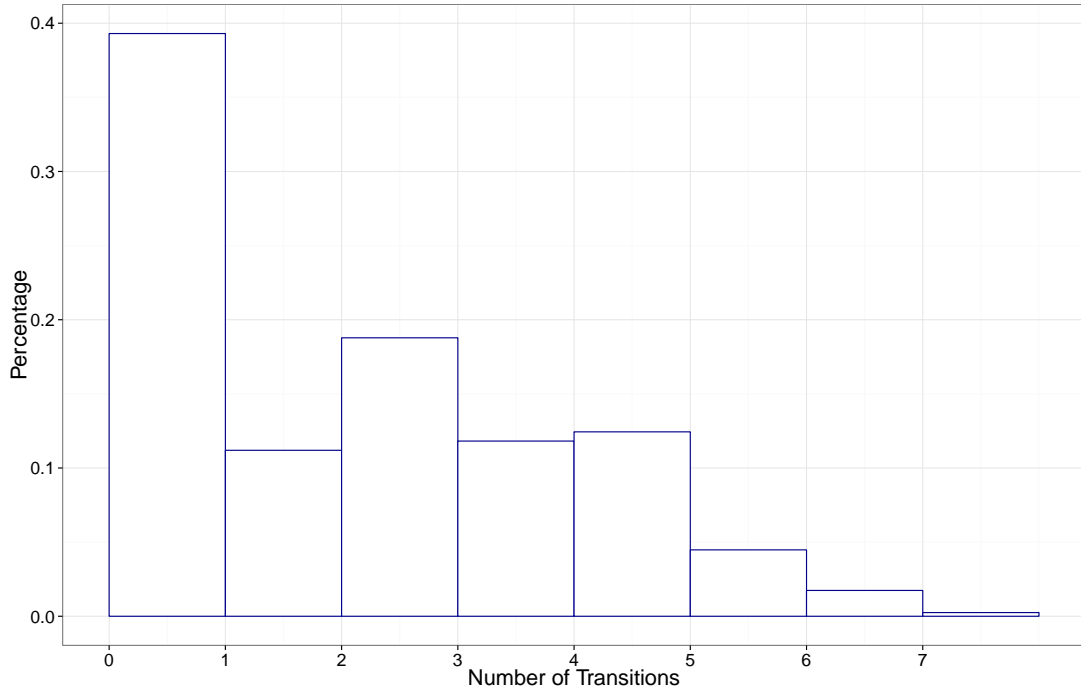


FIGURE 2.1: Number of Interquartile Transitions by Census Tract, Total Crimes 2003-2011

Table 2.5: Summary statistics for yearly crimes Census Tract, 2003-2011

Variable	Mean	S.D.	Min.	1st Q	Median	3rd Q	Max.
Violent Crimes	8.47	8.41	0	3	6	12	71
Total Crimes	189.70	148.30	0	99	153	240	2110

due to an increase or decrease of between 10 and 50 crimes, and more than 10% of them due to a difference of more than 50 crimes. Table 2.5 provides summary statistics for total and violent crimes for comparison. As for violent crimes, almost half of the census tracts had changes of more than 5 crimes between two years, and a considerable percentage had differences larger than 10 crimes. This is very high, specially if compared to a mean of 8.5 violent crimes for the whole sample, and a median of 6.

We finish by showing that most census tracts move to a different quartile at some point. Figure 2.3 shows a histogram for the number of interquartile transitions for

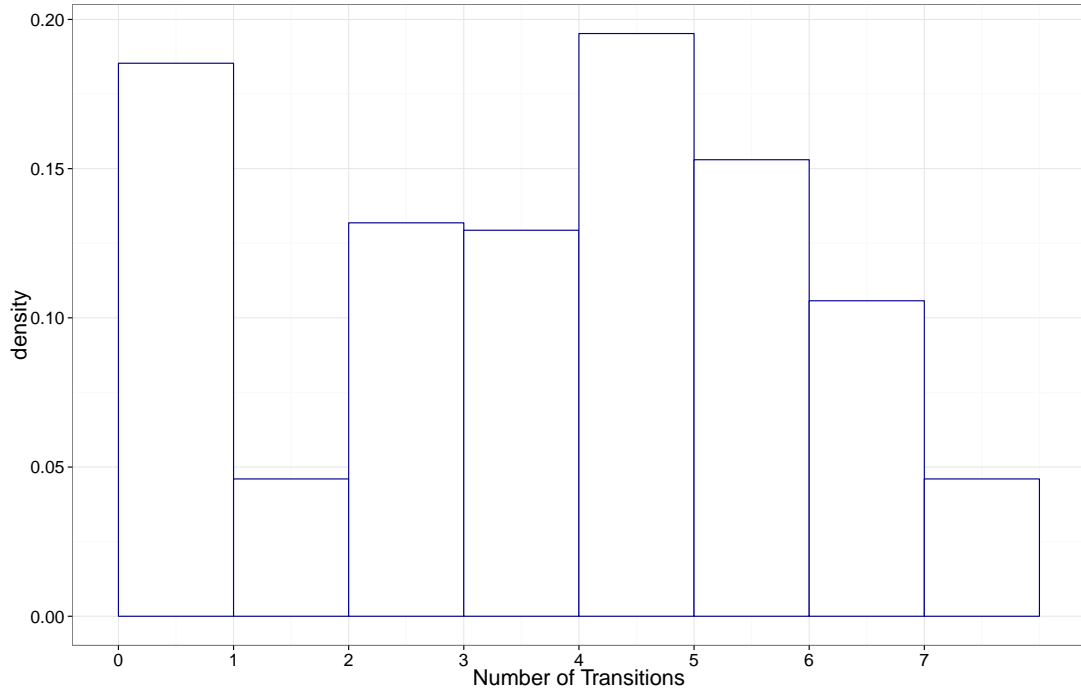


FIGURE 2.2: Number of Interquartile Transitions by Census Tract, Violent Crimes 2003-2011

total crimes. During the nine years we use, less than 40% of the census tracts remain in the same quartile. The majority moves to different quartiles from one to seven times. Figure 2.4 shows the same figure for violent crimes. As we would expect from the high transition rates we saw in table 2.4, less than 20% of census tracts go through the nine years without switching quartiles.

2.3 House Prices and Past Crimes

In this section we provide some reduced form evidence on the relationship between crimes and house prices. We will use very detailed information on the timing and location of both crimes and real estate transactions to build an identification strategy that allow us to estimate the effect of crimes on house prices without the interference of omitted variables.

We do not interpret the coefficients in this section as the MWP to avoid crime,

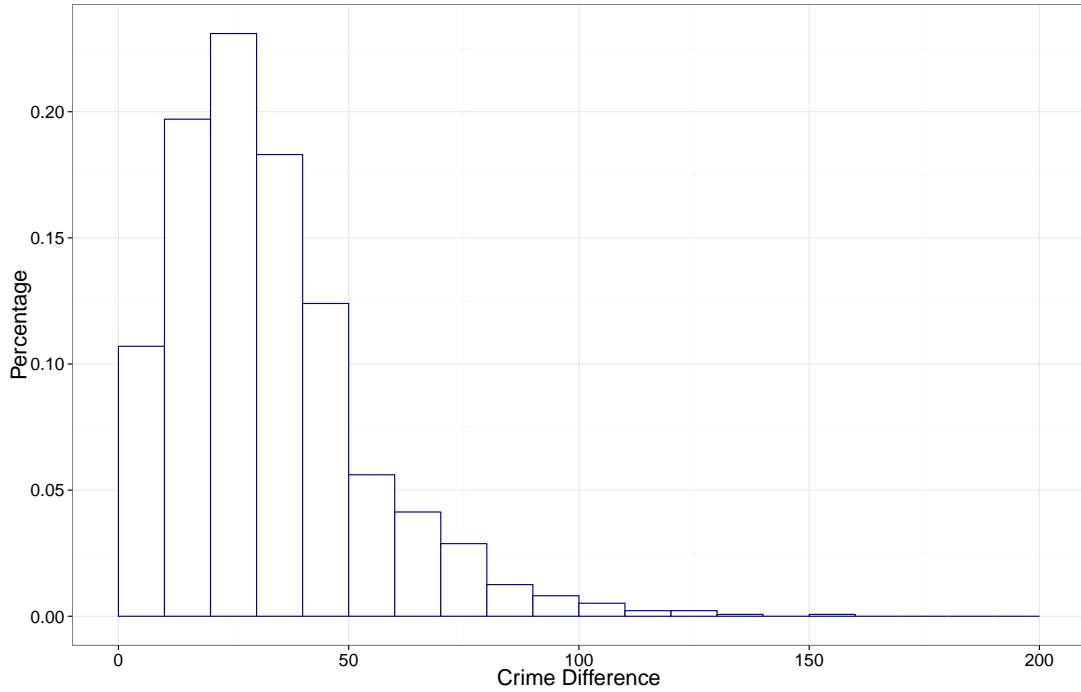


FIGURE 2.3: Absolute change in crimes conditional on changing quartiles, Total Crimes 2003-2011

but this does not diminishes the importance of these results. First of all, we do believe these estimates of the capitalization of crimes in house prices are interesting by themselves. We are able to provide the first evidence on how monthly crimes affect house prices, and specially how the effect of one additional crime dissipates over time. This finding will guide some of our model assumptions in section 2.4. The results also show very clearly the role played by omitted variable bias and the efficacy of our identification strategy. Finally, in section 2.5 we will be able to test our model by comparing its predictions to some of the results in this section.

We begin this section by explaining our identification strategy for handling omitted variables, and then present the results.

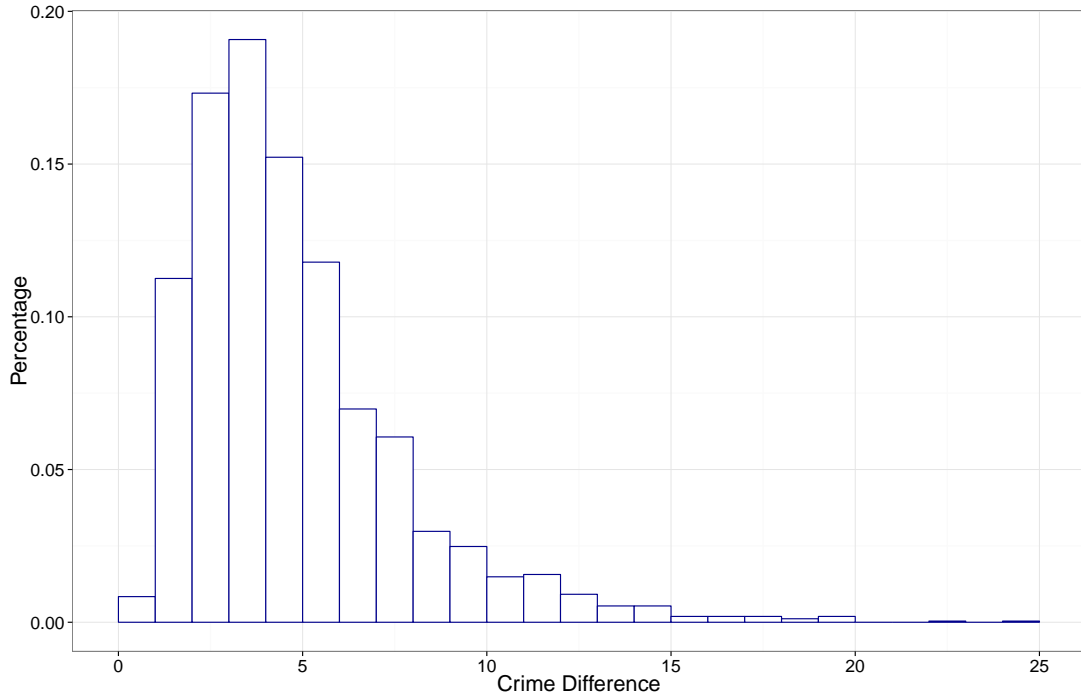


FIGURE 2.4: Absolute change in crimes conditional on changing quartiles, Violent Crimes 2003-2011

2.3.1 Identification Strategy

We want our identification strategy to be robust to two different problems. One is the omitted variable bias. We expect the distribution of crime in a city to be correlated with several (dis)amenities. Poor schools, houses of worse (unobserved) quality, lower income, and poor street lighting are all examples of house and neighborhood characteristics that may be correlated with crime. As we will show later in this section, not controlling for these endogenous variables would lead us to overestimate the effect of violent crimes and find a positive effect for property crimes on house prices.

Another problem is that most studies have not been able to account for how localized crimes are (Gibbons (2004) is an exception). Due to data limitations, most have looked at the effect of crimes aggregated at the census tract, police beat or zip

code. This leads to a treatment that is heterogeneous across houses if agents value closer crimes differently from distant ones. First, because different census tracts, police beats or zip codes all differ in area. As a consequence, an average crime will have a different impact in a small versus a large area. Second, because the number of crimes in any of spaces will be much more informative of a house in its center than one in its boundary, introducing some measurement error. In addition, there is the possibility that these areas are too wide, and by using them we would count crimes that do not affect house prices because they are too distant from the house. This could further bias the estimates towards zero.

Our strategy uses very detailed information on time and location of crimes to circumvent those difficulties. Our data allow us to focus only on crimes at a predetermined distance from each house. This way we can understand what is the trade-off between increasing the maximum distance (and thus variation) and weakening the average crime's influence on house prices. It is then easier to deal with any concerns regarding the correct spatial aggregation of crime.

We will focus on the effects of crimes at most 0.4 miles from the house. One could worry that this is already too far. Research looking at the consequences to house prices of having sex-offenders moving nearby have found that the effects disappear after 0.1 mile (see Linden and Rockoff (2008), and Pope (2008)). This points to what could possibly be a very local effect of crimes on house prices.

Due to data limitations, we cannot estimate effects of crime at such a short distance.¹¹ But we will still show that effects of crime in house prices decrease with distance at a slower rate than the effect of sex-offenders. Indeed, it only disappears after 0.4 mile. Our reason for choosing this as our preferred distance is because it gives us the most variation without including crimes that agents seem to ignore.

¹¹ As we explained in section 2.2, the exact location of each crime is not disclosed by the police departments so as to protect victim's privacy. The address is approximated to the closest hundred street number. Also, we would hardly have enough variation at such a short range.

With such a detailed data we can also think more carefully about timing of the treatment. To our knowledge no paper so far was able to focus only on crimes that happened before the house is sold. Instead, the literature has used crimes in the year the house was sold as the variable of interest, thus lumping together crimes that happened before and after the sale. By counting only crimes before the sale, we get a more clear treatment, and by aggregating crimes on a monthly basis, instead of yearly, we can show how the effect of an additional crime dissipates with time. We believe this result is interesting in itself, but we will also use it in section 2.5 to test the fit of our model.

We should note that using total crimes in a year as opposed to crimes before the house is sold can be rationalized by arguing that yearly crimes are a proxy for some latent variable that describes the underlying safety of a neighborhood. If this is the case though, the correct strategy would be to model the relationship between latent and observed variable and account for possible changes in the latent variable. This will be our strategy in sections 2.4 and 2.5.

Thus, the first regression of interest is

$$\log(\text{price}_{it}) = \sum_{j \in \text{Crimes}} \sum_{\tau=4}^{-12} \alpha_{j\tau}^{4/10} Cr_{ij\tau}^{4/10} + \beta X_i + u_i,$$

where $Cr_{ij\tau}^{4/10}$ is the number of crimes of type j , no more than 0.4 miles from house i , τ months before the house was sold (negative for before the sale and positive for after), and X_i is a vector of house characteristics and a constant.

We have not made any references to omitted variable bias thus far. We address this problem by including census block group by year by quarter fixed effects. The identifying assumption is that any omitted variables that are correlated with crime affect the whole Census Block Group with the same intensity. Since the fixed effects are interacted with year and quarter, we also account for any trends in house

prices and for any omitted variables that may be changing over time (as long as the identifying assumption still holds).

Once again we can see the importance of the detailed crime data. Because we can match crimes to house transactions at pre-specified distances and relative to the exact date of the sale, we have variation in number of crimes within the census block group. This is why we are able to include fixed effects that are so detailed spatially and across time.

In hindsight, we cannot guarantee that the fixed effects are effective in controlling for omitted variables. We test this hypothesis by including placebo treatments. More specifically, we include in all regressions the number of crimes from one to four months after the sale. These crimes were not observed by buyers nor sellers, and should not influence price in the absence of endogeneity. If these coefficients are statistically significant, that can be taken as a signal of omitted variable bias.

The second specifications is:

$$\log(\text{price}_{it}) = \sum_{j \in \text{Crimes}} \sum_{\tau=4}^{-12} \alpha_{j\tau}^{4/10} Cr_{ij\tau}^{4/10} + \beta X_i + \delta_{it} + u_i,$$

where δ_{it} is the census block group by year by quarter fixed effect.

Before presenting the results, we should say why we will not interpret α 's here as the MWP to avoid crime. How much the price of a house is updated following a certain amount of crime depends on two things. First, it depends on agent's preferences for avoiding crime, that is, their MWP. But it also depends on how agents are learning with new signals. Agents are trying to learn how safe each house is, and in so doing they value past crimes because these crimes tell them something about future crimes. Since these regressions do not account for this learning aspect of the problem, we cannot expect an estimate of the MWP from them.

2.3.2 Initial results

We start by showing results from our first specification, without fixed effects. Figures 2.5 and 2.6 present coefficient estimates and 95% confidence intervals using a 0.4 mile radius for property crimes and violent crimes, respectively.¹² The y-axis is the coefficient estimate. The dependent variable is in log, so a coefficient of -0.01 indicates that one additional crime in that month decreases house prices in 1%. The x-axis is number of months from the sale, going from 4 (91 to 120 days after the sale) to -12 (360 to 389 days before the sale).

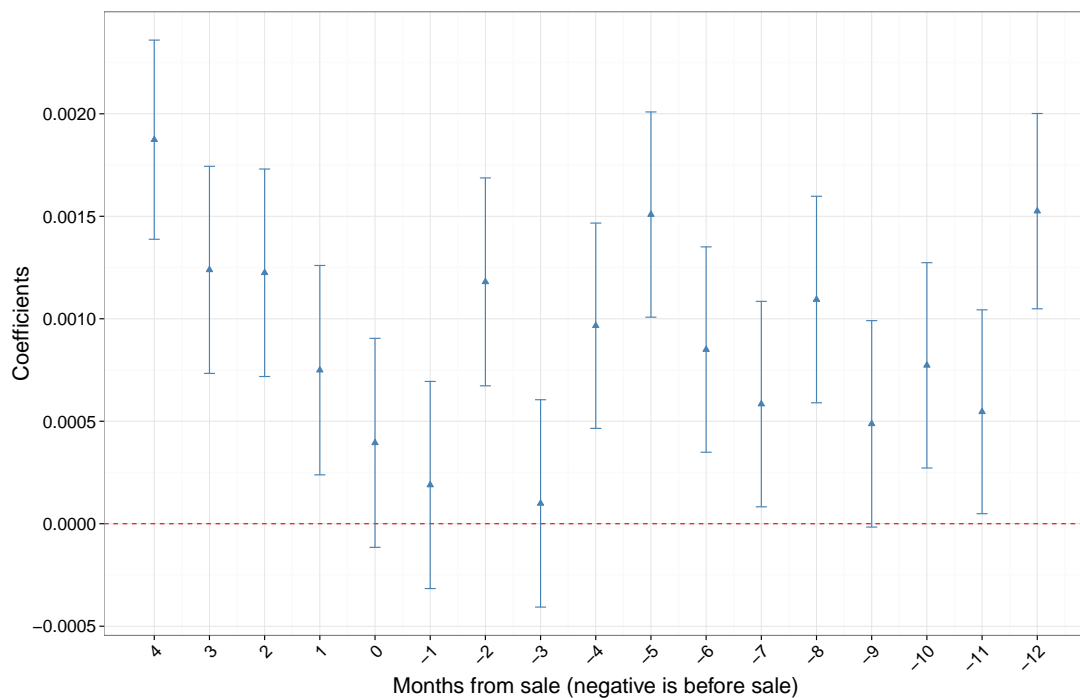


FIGURE 2.5: Property crimes at a 0.4 mile radius, no FE. Dependent variable is log-price. Each dot is the percentage effect of an additional property crime. Negative months indicate crimes before the sale.

These regressions make evident the important role played by omitted variables. Property crimes are positively correlated with house prices in all months, including

¹² Coefficients in both figures were estimated in a single regression. Standard errors were clustered at the Census Block Group.

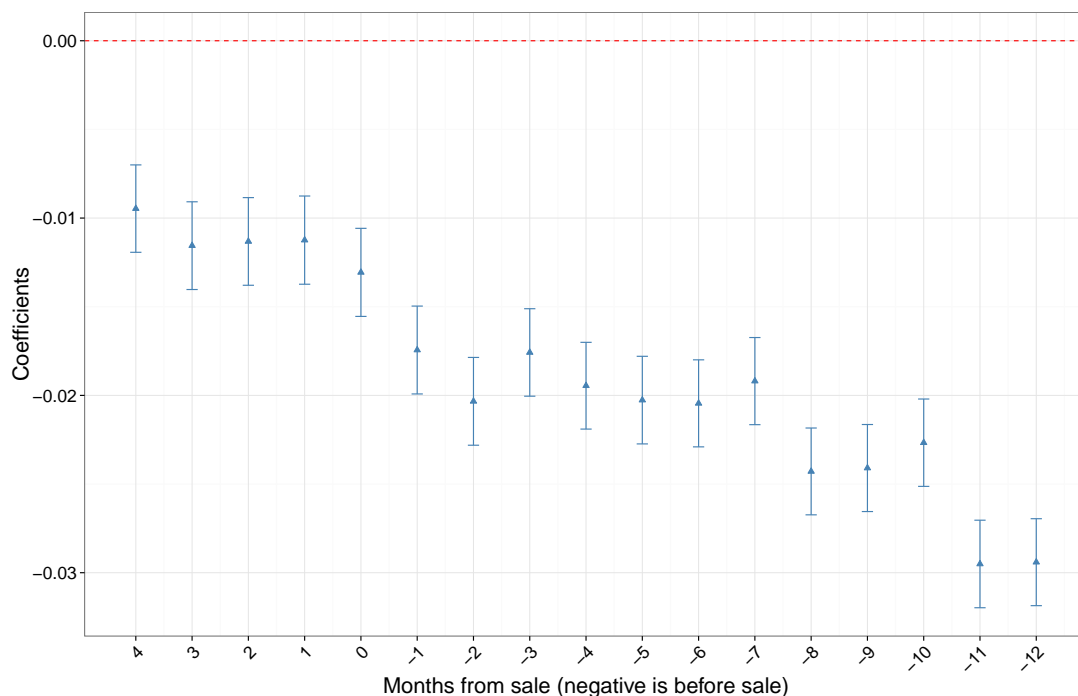


FIGURE 2.6: Violent crimes at a 0.4 mile radius, no FE. Dependent variable is log-price. Each dot is the percentage effect of an additional violent crime. Negative months indicate crimes before the sale.

those after the sale. Violent crimes on the other hand have very large and negative correlation. One additional violent crime would seem to decrease house prices by 1% to 2%, depending on the month.¹³ What these coefficient estimates are really capturing is the large concentration of reported property crimes in wealthier neighborhoods and of violent crimes in poorer neighborhoods, and tell us close to nothing about the actual effect of crimes on house prices.

Figures 2.7 and 2.8 introduce the census block group by year by quarter fixed effects in the regressions. Crimes after sale do not affect house prices anymore, indicating that omitted variables are no longer an issue. The positive correlation between property crimes and house prices is gone, but now the coefficients show no clear pattern, with almost all of them statistically insignificant.

¹³ These findings are similar to those for Jacksonville, FL, in Lynch and Rasmussen (2001). The authors do not control for endogeneity either.

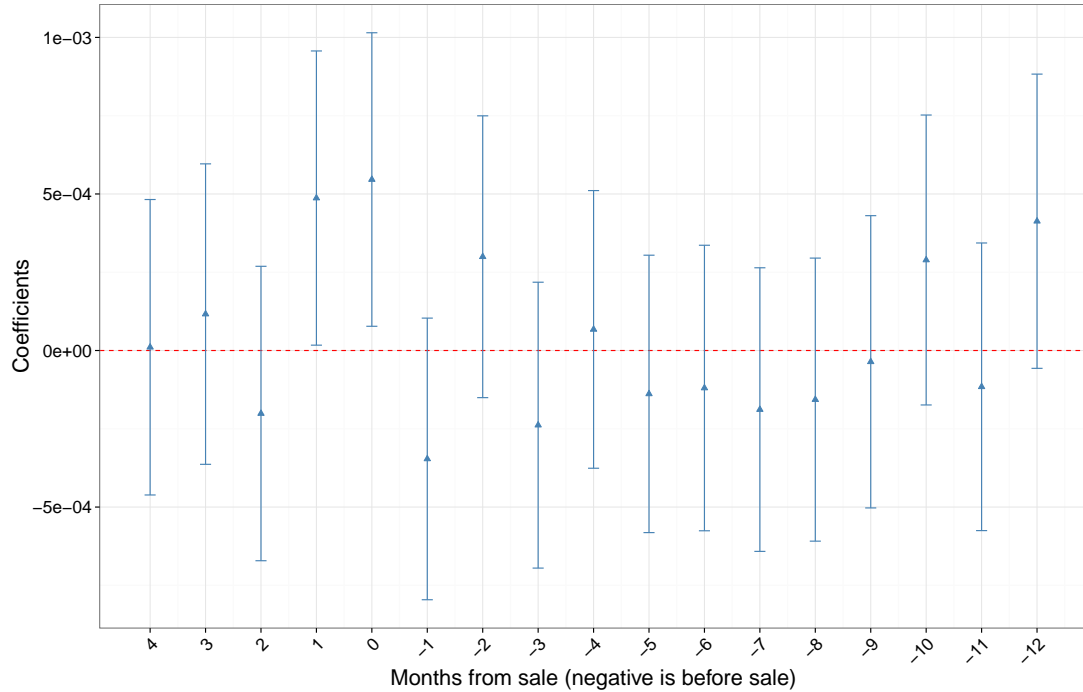


FIGURE 2.7: Property crimes at a 0.4 mile radius, Census Block Group by quarter FE. Dependent variable is log-price. Each dot is the percentage effect of an additional property crime. Negative months indicate crimes before the sale.

Violent crimes show more interesting results. Coefficients remain insignificant for crimes from $t = 4$ to $t = -1$, move abruptly to -0.5% at $t = -2$, and then slowly fade away. The seemingly weird timing, with crimes at $t = 0$ and $t = -1$ having no effect, is explained by the lag between deciding to buy a house and the sale becoming official. During these two events, buyers must contract a mortgage and handle legal obligations. Since the price is decided before any of these steps take place, it stands to reason that crimes right before the sale should have no impact in house prices. This pattern is robust to disaggregating crimes by distance to the house or crime type, as we will see.

It is important to take note of how the effects of crime fade with time. There are two possible interpretations for this. One is that somehow agents have less information about older crimes. This could arise if either the cost of collecting

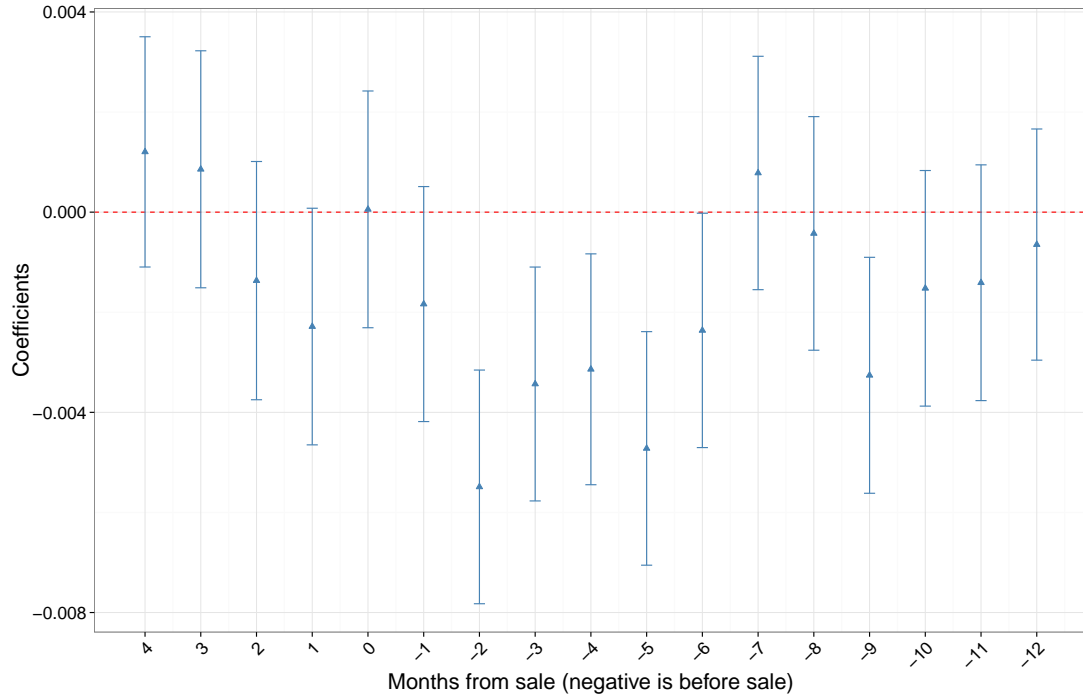


FIGURE 2.8: Violent crimes at a 0.4 mile radius, Census Block Group by quarter FE. Dependent variable is log-price. Each dot is the percentage effect of an additional violent crime. Negative months indicate crimes before the sale.

information increases the farther we look into the past, or if the information about the crime becomes noisier with time. This would be a reasonable hypothesis in the past, when information on crimes would have to be acquired either through old newspapers or neighbors. It does not sound as credible today because information on past crimes is now easily available through different websites that provide crime maps for the city of Chicago, including Chicago PD’s own website. These websites are all fairly simple to use, and the cost of observing crimes for different periods is the same.

The alternative interpretation takes us back to learning. As we discussed earlier, buyers are most likely using past crimes to assess the expected number of future crimes around the house. As a new signal arrives, prices are updated to reflect new beliefs. A standard learning model cannot explain how the effect an additional crime

decreases with time though. If agents believed that the crime level around each house is constant, then all observed crimes would carry the same amount of information, independent of how far in the past they were. Consequently, all coefficients would have to be roughly of the same size. Moreover, as agents observe increasingly more monthly crimes, the contribution of each crime to house prices will converge to zero.

This result changes if the crime level surrounding houses is allowed to change. If this is indeed the case, then recent crimes will carry more information than not so recent ones. The value of information depreciates with time because it becomes progressively more likely that the crime level changed since the crimes were observed. Thus, figure 2.8 corroborates the learning model we develop in section 2.4.

We can calculate the effect of a one standard deviation increase in monthly violent crimes for all 12 months prior to the sale ($t = 0$ through $t = -11$). The standard deviation of monthly violent crimes is 1.91 after excluding robberies, and thus the total effect of such an increase in violent crimes on house prices is around -5.11%.

This number should be not taken at face value though. The coefficients in this regression are estimated from within census block group-quarter monthly variations in crimes. It does not account for a possible build up in the effect of crime on house prices when a consistent increase in crime is observed, and as such could be understating the effects. It is different for an agent learning about a house if he observes one more crime in a specific month and if he observes one more crime than expected several months in a row.

One benefit of our data is that we can disaggregate crime by distance to the house. This is done in figures 2.9 and 2.10. Each figure is divided in six panels, for crimes up to 0.2 mile, from 0.2 to 0.3 mile, and so on up to 0.7 mile.¹⁴ We keep the same scale for all panels within a figure.

¹⁴ Again, all coefficients in both figures were estimated jointly.

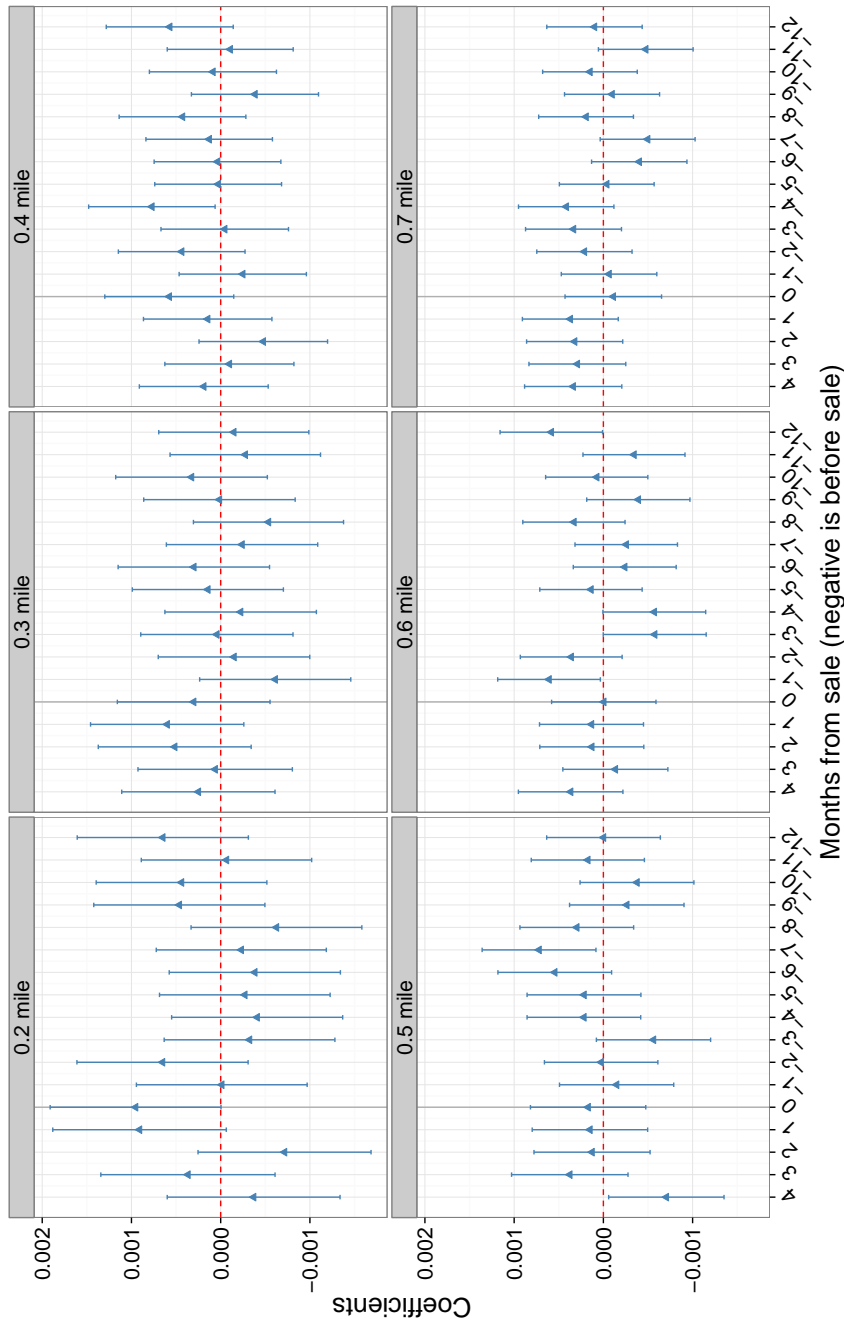


FIGURE 2.9: Property crimes by distance, Census Block Group by quarter FE. Dependent variable is log-price. Each dot is the percentage effect of an additional property crime. Negative months indicate crimes before the sale. 0.2 mile are all crimes less than 0.2 mile from the house, 0.3 mile are crimes from 0.2 to 0.3 mile, and so on.

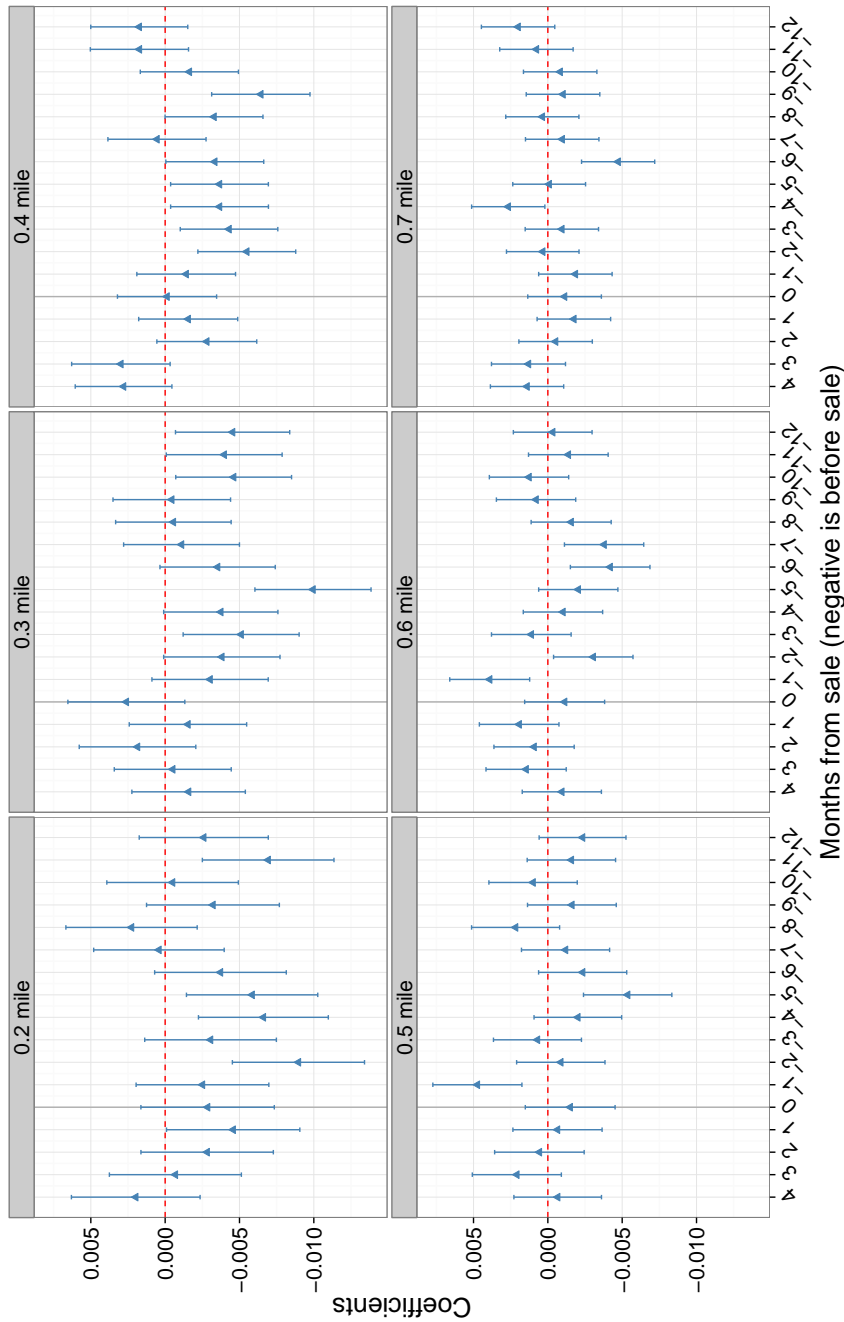


FIGURE 2.10: Violent crimes by distance, Census Block Group by quarter FE. Dependent variable is log-price. Each dot is the percentage effect of an additional violent crime. Negative months indicate crimes before the sale. 0.2 mile are all crimes less than 0.2 mile from the house, 0.3 mile are crimes from 0.2 to 0.3 mile, and so on.

Results for property crimes are robust across distances, with coefficients not significantly different from zero. That is not the case for violent crimes: although estimates are less precise now, there is a clear difference between the effect of crimes up to 0.4 mile and from 0.4 to 0.7 mile. The later do not show any distinct patterns. It seems that crimes at such distances no longer affect house prices. Crimes up to 0.4 mile, on the other hand, present the same pattern we saw before, the first significant coefficient at $t = -2$, followed by a slow fading. This indicates that for Chicago 0.4 mile is a reasonable distance for studying violent crimes. The effects does seem to weaken with distance a little, as we see by comparing panels for 0.2 mile and 0.3 to 0.4 mile, although the estimates are not statistically different.

We finish this section by disaggregating crimes. Figures 2.11 and 2.12 present the coefficients for violent and property crimes respectively. Again, because of the smaller variation, estimates are less precise, specially for homicides and sexual assaults, which are the least common types. Even so, it is clear that the impact of an additional homicide on house prices is considerably stronger than other violent crimes. While aggravated assault, which is responsible for most of the variation in the aggregated data, has a coefficient of -0.005 at $t = -2$, the coefficient at the same period for homicide is closer to -0.015 , three times the size for aggravated assaults. Alternatively, we can compare the effects of an increase in one standard deviation in homicides (0.64) versus one standard deviation in aggravated assaults (1.52). In this case, homicides would lead to a decrease in house prices in the order of 0.96%, compared to 0.75% for aggravated assaults. That is, the impact of an additional homicide is large enough to compensate for its rarity.

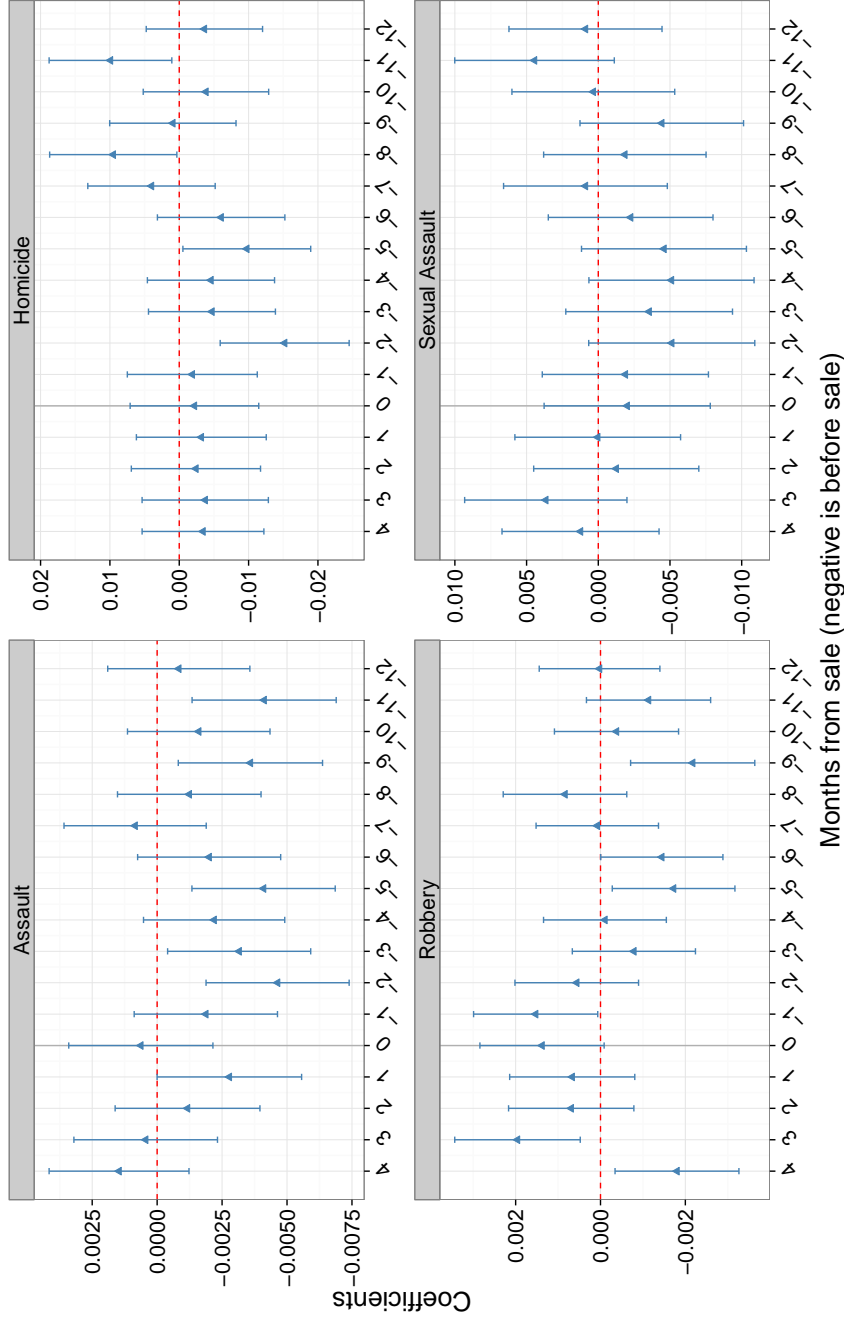


FIGURE 2.11: Crimes at a 0.4 mile radius, Census Block Group by quarter FE. Dependent variable is log-price. Each dot is the percentage effect of an additional crime. Negative months indicate crimes before the sale. Effects for all 7 crimes were estimated in the same regression.

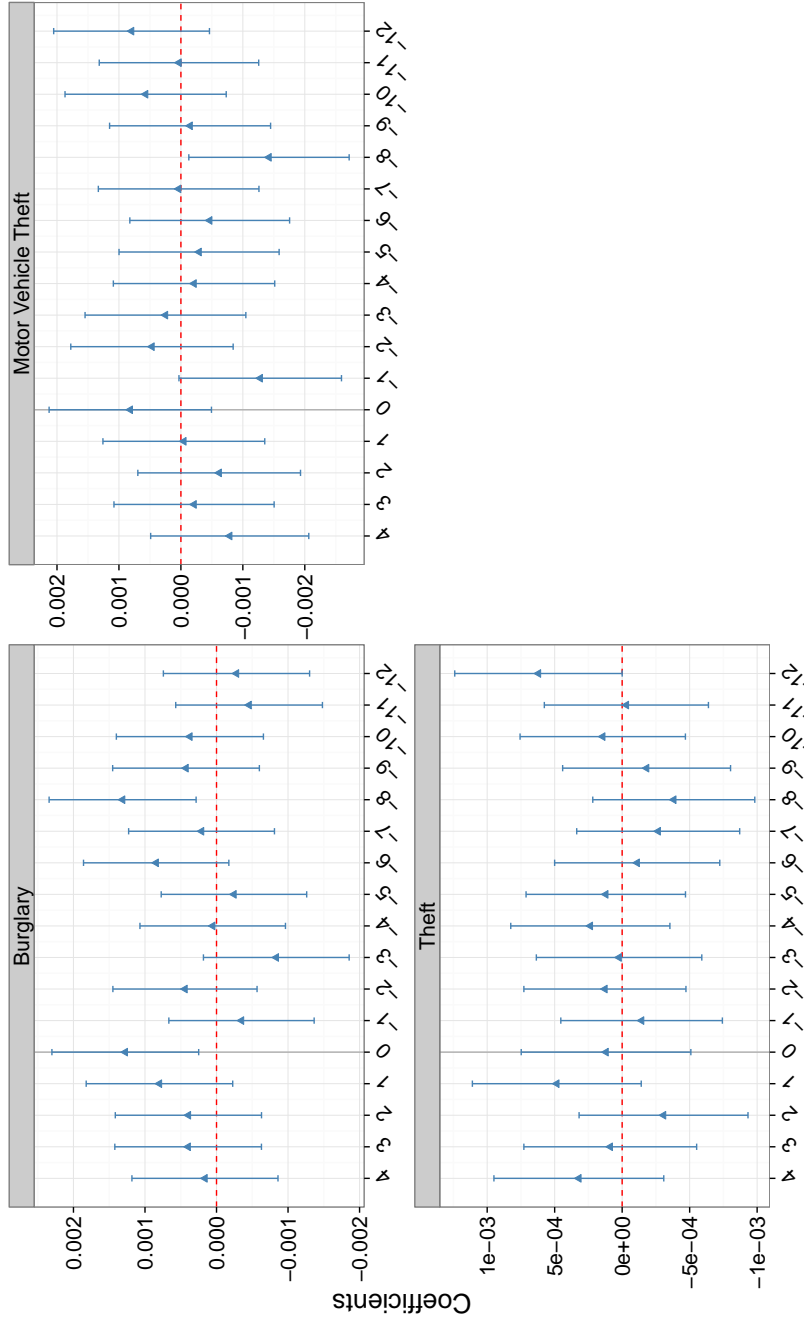


FIGURE 2.12: Crimes at a 0.4 mile radius, Census Block Group by quarter FE. (cont.)

The coefficients for sexual assault are not significant. This is most likely a consequence of the smaller variation. They nonetheless show a similar pattern as aggravated assault and homicide, with a larger impact at $t = -2$, and coefficients hovering around -0.005 for about 4 months.

Robberies, on the other hand, do not show a clear pattern, even if a few coefficients are significant, and we see a similar result for property crimes. We can interpret these result in two ways: either the MWP to avoid such crimes is not large enough, or an extra crime brings little additional information so that agents will not update their beliefs.¹⁵

2.4 Model

In this section we develop a model to estimate the MWP to avoid crime from our monthly crime data. In addition to provide a clear strategy to identify the MWP, we also want a model that is able to reproduce the patterns we saw in the reduced form regressions, especially how the effect of crime in house prices fades with time.

Our model has two main features. First, we assume that agents are forward looking. The value of each house is a function of the sequence of future crimes, which is not observed. Agents must use data on past crimes to form expectations about the future. One consequence of this assumption is that the amenity agents value when choosing a house is not total crimes in the year or past monthly crimes, but his belief of what the probability distribution of crime for the house is at the present and in the future.

Second, we allow the distribution of crime to change over time. This is a reasonable assumption by itself, but it plays a very important role in the model. If the distribution of crime was fixed, monthly crimes would have the same amount of

¹⁵ Although we do not show the regressions in the paper, robberies do not have statistically significant effects at shorter distances.

information, whether they took place one month or one year earlier. Consequently, house prices would be affected by monthly crimes in the same way, independent of when these crimes were observed. By allowing the distribution to change, we break this result. Since the odds of the distribution changing over a one year period is much higher than over a month, recent crimes will contain more reliable information than older crimes and consequently will affect house prices more.

We choose to model crimes as a Latent Markov Model. This is a very flexible framework that enables us to estimate a statistical model of crime in the first stage. One nice feature of our model is that both the LMM and agents beliefs are estimated without any reference to the agents' decision problem. Several specifications of the price function can be estimated without having to re-estimate the LMM. Moreover, the LMM can be easily introduced even into more complex dynamic structural models such as Bayer et al. (2011) and Bishop and Murphy (2011).

We use a stylized model of the real estate market. There are two main assumptions: agents have homogeneous preferences over crime and that after choosing a house, buyers keep it indefinitely. Under these assumptions, the price equation is a linear function of house characteristics and the present discounted value of the disutility from crime. This equation can be estimated by employing standard statistical packages. The reason we choose this simple model is because the resulting price function is closer to what most papers have estimated in the hedonic price literature (see ? for a review of the literature). This way we can illustrate how even less structural papers can benefit from using the LMM.

We start this section by explaining how a LMM can be used to model crime. Next, we explain how we use the LMM and observed crime sequences to form agents beliefs about crime for each house. We then present agents' decision problem and derive the price function. We finish by discussing the estimation process.

2.4.1 Crimes

We first develop a statistical model of crime. We will model monthly crimes surrounding a house as a LMM.¹⁶ In this model, the probability distribution of crime around houses will be allowed to change over time within a fixed set of probability distribution. This set will be the same for all houses in the city of Chicago. Agents will use this model and the observed sequence of crimes to form beliefs about different houses, which will then be used to estimate the MWP.

Assume there are K states a house can be in any time, and the house's state may change from period to period. Each state is defined by a probability distribution function of total monthly crimes, $P(Cr|s = k)$, where k is the state and Cr is total monthly crimes. Neither the econometrician nor the agents observe the sequence of states for a house, but they do observe monthly crimes. The LMM is a probabilistic model of the joint probability of $\{\{Cr_0^i, \dots, Cr_T^0, s_1^i, \dots, s_T^i\}_{i=1}^I\}$, that is, the history of states and crimes for all $i \in [1, I]$ houses for periods 0 through T .

There are two main independence assumptions underlying the LMM:

Assumption 1. $P\left(s_{t+1}^i \mid \{s_t^j, Cr_t^j, \dots, s_0^j, Cr_0^j\}_{j=1}^I\right) = P(s_{t+1}^i | s_t^i)$

Assumption 2. $P\left(Cr_t^i \mid \{s_T^j, Cr_T^j, \dots, s_0^j, Cr_0^j\}_{j=1}^I \setminus \{Cr_t^i\}\right) = P(Cr_t^i | s_t^i)$

Assumption 1 states that conditional on house i 's state in period t , the probability distribution function of states in period $t + 1$ for house i does not depend on its past history, or on the past and current history of any other houses. This is the standard assumption of any Markov model. Assumption 2 states that conditional on house i 's current state, the distribution of crimes in i does not depend on the past or future history of i or in the history of any other house.

¹⁶ See Bartolucci et al. (2013) and Zucchini and MacDonald (2009) for more information on LMM and its time series counterpart Hidden Markov Model.

We make one further assumption that the parameters of the LMM are homogeneous across houses and time:

Assumption 3. $P(s_{t+1}^i | s_t^i) = P(s' | s)$; $P(Cr_t^i | s_t^i) = P(Cr | s)$

Under assumptions 1 through 3, we can summarize our model of crime by a K by 1 vector $\mathbb{P}(Cr)$ of the probability distribution of crimes in each state, a K by K transition matrix of states \mathbb{T} , and a K by 1 vector of initial probabilities τ_0 . The vector τ_0 is the unconditional probability a house is in each state at period 0. The first step to estimating the MWP to avoid crime will be to estimate vectors $\mathbb{P}(Cr)$ and τ_0 and matrix \mathbb{T} using a variant of the EM algorithm, as we explain in more detail at the end of this section.

2.4.2 Agents Expectations

To move from the LMM to actually calculating agent's beliefs about the probability of each state it is necessary to define what is the information set available to agents. We will assume here that agents know the LMM, and that they observe the sequence of monthly crimes for the house until the present. With this information at hand, they will form beliefs using Bayes rule, while also accounting for the fact that the state may change, that is, that they are learning about a moving target.

Figure 2.13 illustrates the sequence of events within each period. There are only two exogenous events in a period. First, total monthly crimes is realized. This is observed by all agents. Then, at the end of the period, states may transition or not. Although agents must account for these transitions, the outcome is hidden from them. To understand how agents address both these events when updating their beliefs it is useful to split the Bayesian updating in two stages, which is shown below (figure 2.13).

Agents start period t with a prior $\mathbb{P}(s_t^i | I_{t-1}^i)$, where I_{t-1}^i is the information set for house i at the end of period $t - 1$, and thus excludes total monthly crimes in t . Their

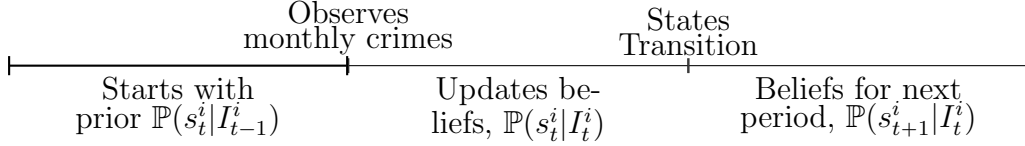


FIGURE 2.13: Timeline of events

beliefs are updated taking into account the new information, Cr_t^i , to obtain $\mathbb{P}(s_t^i | I_t^i)$. Note that $I_t^i \equiv I_{t-1}^i \cup Cr_t^i$. Equation 2.1 provides the rule for the Bayesian updating.

$$\mathbb{P}(s_t^i = k | I_t^i) = \frac{P(Cr_t^i | s_t^i = k) \cdot \mathbb{P}(s_t^i = k | I_{t-1}^i)}{\sum_{j=1}^K P(Cr_t^i | s_t^i = j) \cdot \mathbb{P}(s_t^i = j | I_{t-1}^i)} \quad (2.1)$$

The beliefs in equation 2.1 still do not account for any state transitions. Equation 2.2 shows how to use these beliefs to construct agent's prior at the beginning of $t + 1$. The equation is very straightforward. It says that the prior of being in state k in $t + 1$ will depend on the posterior for each state j in t and the probability that each state will transition to k , given by the element $\mathbb{T}_{j,k}$ from the transition matrix.

$$\mathbb{P}(s_{t+1}^i = k | I_t^i) = \sum_{j=1}^K \mathbb{P}(s_t^i = j | I_t^i) \cdot \mathbb{T}_{j,k} \quad (2.2)$$

We can use equations 2.1 and 2.2 to write the one-step updating rule from the start of period t to the start of period $t + 1$:

$$\mathbb{P}(s_{t+1}^i = k | I_t^i) = \frac{\sum_{j=1}^K P(Cr_t^i | s_t^i = j) \cdot \mathbb{P}(s_t^i = j | I_{t-1}^i) \cdot \mathbb{T}_{j,k}}{\sum_{j=1}^K P(Cr_t^i | s_t^i = j) \cdot \mathbb{P}(s_t^i = j | I_{t-1}^i)} \quad (2.3)$$

Finally, we show in equation 2.4 how agents form beliefs about the probability of being in any state in period $t + \Delta$, conditional on their beliefs in t . Given the Markov structure of the problem, it is very easy to calculate such beliefs. The beliefs for next period are just the beliefs today multiplied Δ times by the transposed transition matrix. The equation below is in matrix form:

$$\mathbb{P}(s_{t+\Delta}^i | I_t^i) = (\mathbb{T}^T)^\Delta \cdot \mathbb{P}(s_t^i | I_t^i) \quad (2.4)$$

2.4.3 Price Equation

Finally, we show how expectations enter the price function. We make two main assumptions. Firstly, agents have homogeneous preferences over crime. Secondly, after an agent buys a house, he cannot sell it. House prices are an equilibrium outcome. Prices will adjust so that agents are indifferent between all houses on sale.

For simplicity, we ignore other amenities or house characteristics here and focus only on crime.¹⁷ Each period, starting when the house is bought, the agent derives dis-utility $u(Cr)$ from observing Cr crimes in its neighborhood. Of course, the agent does not know what the number of crimes will be in the future. Thus the value of house i in t will depend on agents' expectations over the future realizations of crime.

Let $\nu(s)$ be the conditional value function of owning a house in state s today. We can write the recursive value functions for a house, conditional on the current state:

$$\begin{aligned} \nu &= \mu_{Cr} + \beta \cdot \mathbb{T} \cdot \nu \\ \nu &= (I - \beta \cdot \mathbb{T})^{-1} \cdot \mu_{Cr} \end{aligned} \quad (2.5)$$

where

$$\nu = \begin{bmatrix} \nu(1) \\ \vdots \\ \nu(K) \end{bmatrix}; \quad \mu_{Cr} = \begin{bmatrix} \mathbb{E}(u(Cr) | s = 1) \\ \vdots \\ \mathbb{E}(u(Cr) | s = K) \end{bmatrix}.$$

As a remainder, \mathbb{T} is the transition matrix from the LMM, β is the intertemporal discount rate, μ_{Cr} is vector of average expected utilities conditional on state, and ν is the vector of conditional value functions. Thus, the value of a house, conditional

¹⁷ This is the same as assuming that other amenities and house characteristics are additively separable from crime in the utility function.

on its current state, is the present discounted flow of expected utilities. Of course, neither us nor the agents know in what state a house really is, so that the value agents place in the house is some weighted average of the conditional value functions in which the weights are the believed probability of each state. That is, the value of a house conditional on information at the beginning of period t (that is, before crime in t is observed) is

$$\begin{aligned}
V(\mathbb{P}(s_t^i|I_{t-1}^i)) &= \mathbb{E}[\nu(s)|I_{t-1}^i] \\
&= \mathbb{P}(s_t^i|I_{t-1}^i) \cdot \nu \\
V(\mathbb{P}(s_t^i|I_{t-1}^i)) &= \mathbb{P}(s_t^i|I_{t-1}^i) \cdot (I - \beta \cdot \mathbb{T})^{-1} \cdot \mu_{Cr} \tag{2.6}
\end{aligned}$$

According to equation 2.6, houses' values will differ only in so far as the prior probability of each state conditional on information at the beginning of the period differs.

We assume that agents subtract the dis-utility of log house prices, p_t^i , linearly from the utility function. Prices are pinned down by a non-arbitrage condition. Since agents are homogeneous, the system is in equilibrium if and only if agents are indifferent between all houses. This condition implies that for any house

$$\begin{aligned}
V(\mathbb{P}(s_t^i|I_{t-1}^i)) - p_t^i &= C \\
p_t^i &= C + \mathbb{P}(s_t^i|I_{t-1}^i) \cdot (I - \beta \cdot \mathbb{T})^{-1} \cdot \mu_{Cr} \tag{2.7}
\end{aligned}$$

where C is a constant. That is, prices must decrease to compensate for houses which are more likely to be in more dangerous areas in the present ($\mathbb{P}(s_t^i|I_{t-1}^i)$), but still taking into account the full sequence of expected future crimes, including that states may change ($(I - \beta \cdot \mathbb{T})^{-1}$).

2.4.4 Estimation Procedure

The estimation is done in three steps. First, for a given crime type, buffer, and number of states we estimate the associated LMM. Second, we use the sequence of crimes for each house and the LMM estimates to form agents' beliefs over each house. At this point it is also necessary to determine which months agents observe when the house is bought. We should note that neither steps at this point depend on the price equation, and thus only need to be ran once for given crime, buffer, and number of states. Finally, with the beliefs in hand, we can estimate the price equation. We now explain each of these steps in detail.

When estimating the LMM, we make the simplifying assumption that in each state crime is distributed according to a Poisson with rate λ_j . Although it is theoretically possible to estimate a non-parametric discrete distribution of monthly crimes for each state, assuming a Poisson distribution vastly decreases the number of parameters that need to be estimated. This is specially true for property crimes, in which the range of observed crimes is considerably large.

Our problem in this first step then breaks down to estimating the vector of initial probabilities, τ_0 , the transition matrix, \mathbb{T} , and the vector of Poisson rates, Λ , with each element λ_j being the average monthly crimes in state j . The likelihood for this problem is

$$\ell(\Theta; \{Cr_t^i\}_{i,t}) = \sum_{i=1}^N \log \left(\sum_{\{s_t^i\}_{t=1}^T} P(\{Cr_t^i\}_{t=1}^T | \{s_t^i\}_{t=1}^T, \Theta) \cdot P(\{s_t^i\}_{t=1}^T | \Theta) \right)$$

where $\Theta = \{\tau_0, \mathbb{T}, \Lambda\}$, and the summation inside the log term is over all possible histories for house i , $\{s_1^i, \dots, s_T^i\}$.

Two main issues come up when maximizing this log-likelihood. First, the summation inside the log term makes the problem extremely non-linear, complicating the

maximization. In addition, the number of terms in the summation is J^T , and thus increases exponentially with time. We will employ here an usual method for estimating LMM, the Baum-Welch algorithm, which is a special case of the Expectation-Maximization (EM) algorithm (see Baum et al. (1970) and Dempster et al. (1977a)). The EM algorithm allows for the maximization of the likelihood with an iterative two-step procedure that is repeated until convergence, and is specially suited for problems with missing data.

We will estimate the model using the crime sequences at a 0.4 mile around each house. Each sequence has 29 observations, starting at 25 months before the sale up to 4 months after. When estimating the LMM, we will make use of the assumption that the model remains constant through time. This reduces the number of parameter to be estimated to $k \cdot (k - 1) + (k - 1) + k$ for the transition matrix, initial probability distribution, and Poisson rates, respectively. Also, time in each crime sequence can be described entirely in relation to the moment of the sale, that is, $t \in \{-25, \dots, 4\}$. The model will be estimated separately for violent crimes and property crimes, assuming 2, 3, and 4 states.

Finally, the EM algorithm does not guarantee convergence to the global maximum. As a precaution we estimate each LMM from 20 different, randomly picked starting points, and choose the estimates with the higher likelihood.

Now we can use the estimates from the LMM and crime sequences to form agents' beliefs over each house. Starting with an initial set of beliefs, we will use equation 2.3 and observed crimes to update them. There are two arbitrary decisions to be made, though.

As we saw in section 2.3, there is a lag until crimes start influencing house prices. In general, we don't see any effects for violent crimes less than 60 days from the sale. As we argued then, this lag most likely stems from the usual waiting time between deciding to buy a house and having the transaction officially recorded. Now when

forming beliefs, we must decide until when agents observe crimes as well. We choose to show estimates for the MWP assuming agents observe crimes up to 30, 60, and 90 days before the sale.

The second decision is regarding what prior to use. We should be careful so that the prior does not drive the results, specially since in the real word people are likely to hold very different priors depending on the characteristics of the house and the neighborhood. Instead of using different priors for different houses, we chose to use τ_0 as the prior for all houses and assume agents observe from 22 to 24 months of crime. The main idea here is that with such a long sequence of crimes to learn from, agents will not place almost any weight in the initial prior at the time of the sale. Sure enough, we see that for our sample after 22 months beliefs are extremely heterogeneous across houses.

We use a price regression similar to equation 2.7 to estimate the MWP to avoid crime. We use a utility function that is linear in crimes.

$$p_t^i = \alpha \cdot \mathbb{P}(s_t^i | I_{t-1}^i) \cdot (I - \beta \cdot \mathbb{T})^{-1} \cdot \Lambda + X_i \beta + \delta_{h(i)} + u_t^i \quad (2.8)$$

As in section 2.3, we control for house characteristics X_i and a census block group by year by quarter fixed effect $\delta_{i,t}$, which should filter out any endogeneity. We set $\beta = 0.995$ for the monthly intertemporal discount rate, which is roughly equal to a discount of 5% per year.

Because crimes enter the utility function linearly, α should be interpreted as the MWP to avoid one crime today. The term multiplying α is the expected present discounted flow of crimes associated with house i . Prices can change in response to changes in beliefs, in the transition matrix, or in the distribution of crimes in a state, but since the transition matrix and the states are assumed to be the same for all houses and across time, α is identified only from the variation in beliefs for each

Table 2.6: Transition Matrix for Violent Crimes, 0.4 mile, 2 states

$t \setminus t+1$	1	2
1	0.988	0.012
2	0.023	0.977

Table 2.7: Initial Distributions for Violent Crimes, 0.4 mile, 2 states

States	1	2
Prob.	0.627	0.373

house. This does not hinder our ability to study counter-factuals based on changes in \mathbb{T} or Λ .

2.5 Results

We begin by showing estimation results for the LMM for violent crimes 0.4 mile around a house in tables 2.6 through 2.14, assuming 2, 3 and 4 states. We focus on the results for 4 states, which is our preferred specification.

Table 2.14 shows the estimated Λ vector, the expected number of crimes conditional on each state. As expected, the values for each state vary considerably, ranging from 0.329 crimes in a month to 4.218. The initial distribution (table 2.13), on the other hand, is relatively flat, with state 2 being slightly more common and state 4 slightly less.

The transition matrix is presented in table 2.12, with rows being current states

Table 2.8: Lambda for Violent Crimes, 0.4 mile, 2 states

States	1	2
Lambda	0.735	3.052

Table 2.9: Transition Matrix for Violent Crimes, 0.4 mile, 3 states

$t \setminus t+1$	1	2	3
1	0.992	0.008	0.000
2	0.009	0.984	0.007
3	0.000	0.018	0.982

Table 2.10: Initial Distributions for Violent Crimes, 0.4 mile, 3 states

States	1	2	3
Prob.	0.377	0.41	0.213

Table 2.11: Lambda for Violent Crimes, 0.4 mile, 3 states

States	1	2	3
Lambda	0.46	1.553	3.767

Table 2.12: Transition Matrix for Violent Crimes, 0.4 mile, 4 states

$t \setminus t+1$	1	2	3	4
1	0.993	0.007	0.000	0.000
2	0.006	0.989	0.005	0.000
3	0.000	0.010	0.983	0.008
4	0.000	0.000	0.019	0.981

Table 2.13: Initial Distributions for Violent Crimes, 0.4 mile, 4 states

State	1	2	3	4
Prob.	0.248	0.354	0.256	0.142

Table 2.14: Lambda for Violent Crimes, 0.4 mile, 4 states

State	1	2	3	4
Lambda	0.329	1.046	2.198	4.218

Table 2.15: Transition Matrix, 12 months, for Violent Crimes, 0.4 mile, 4 states

$t \backslash t + 1$	1	2	3	4
1	0.923	0.075	0.002	0.000
2	0.067	0.879	0.052	0.002
3	0.003	0.099	0.821	0.077
4	0.000	0.010	0.188	0.801

and columns future states. This matrix is particularly interesting to us, since one of our main points is that agents should account for the chance that a neighborhood may change in the future. We find that the probability a state will change ranges from 0.7% to 1.9% in a month, and it is increasing with the expected number of crimes. These numbers may seem low, but we must keep in mind that these are monthly probabilities. Alternatively, we calculate the transition matrix over a 12 months period in table 2.15. The small monthly probabilities add up to create a probability of leaving the current state between 7.5% and 19.9%. These are very large numbers, that homeowners should account for when valuating a house.

The same estimates for 2, 3 and 4 states are presented in tables 2.16 through 2.24 for property crimes. Again we discuss only the estimates for 4 states. Unlike violent crimes, which are relatively uncommon in most of the city, property crimes are frequent everywhere. This explains the estimates for Λ in table 2.24. The numbers are not only much higher than for violent crimes, but also more disperse: the safest state faces 11.6 property crimes on average in a month, and the least safe, 173.708.

According to the initial distribution, table 2.23, state 4 is relatively rare, with

Table 2.16: Transition Matrix for Property Crimes, 0.4 mile, 2 states

$t \setminus t+1$	1	2
1	0.967	0.033
2	0.136	0.864

Table 2.17: Initial Distributions for Property Crimes, 0.4 mile, 2 states

States	1	2
Prob.	0.782	0.218

Table 2.18: Lambda for Property Crimes, 0.4 mile, 2 states

States	1	2
Lambda	27.038	91.906

Table 2.19: Transition Matrix for Property Crimes, 0.4 mile, 3 states

$t \setminus t+1$	1	2	3
1	0.928	0.072	0.000
2	0.062	0.924	0.014
3	0.000	0.111	0.889

Table 2.20: Initial Distributions for Property Crimes, 0.4 mile, 3 states

States	1	2	3
Prob.	0.399	0.523	0.078

Table 2.21: Lambda for Property Crimes, 0.4 mile, 3 states

States	1	2	3
Lambda	16.193	45.694	140.251

Table 2.22: Transition Matrix for Property Crimes, 0.4 mile, 4 states

$t \setminus t + 1$	1	2	3	4
1	0.935	0.065	0.000	0.000
2	0.038	0.910	0.052	0.000
3	0.000	0.127	0.859	0.014
4	0.000	0.000	0.077	0.923

Table 2.23: Initial Distribution for Property Crimes, 0.4 mile, 4 states

State	1	2	3	4
Prob.	0.253	0.48	0.224	0.043

Table 2.24: Lambda for Property Crimes, 0.4 mile, 4 states

State	1	2	3	4
Lambda	11.606	33.287	66.597	173.708

Table 2.25: Transition Matrix, 12 months, for Property Crimes, 0.4 mile, 4 states

$t \setminus t + 1$	1	2	3	4
1	0.530	0.379	0.086	0.004
2	0.223	0.554	0.206	0.018
3	0.126	0.507	0.312	0.056
4	0.038	0.242	0.311	0.409

only 4.3% of the houses. That means state 4 is necessary to accommodate a small portion of the houses at each point in time. But it is not the case that this small portion of houses remains the same over time. The transition matrix in table 2.22 shows that about 7.7% of houses in state 4 transition to state 3 every month. Thus there is actually quite a bit of change even among the least safe houses.

In general the distribution of property crimes seems to change at a faster pace than violent crimes. This is evident in table 2.25, where we calculate the transition matrix for the 12 month period. For such a long period, change is more the rule than the exception.¹⁸

We use the estimated LMM to build expectations of the present discounted flow of crimes for each house, and use this to estimate the MWP to avoid crime from equation 2.8.¹⁹ Results for violent and property crimes are in tables 2.26 and 2.27, respectively. In the previous section we said it was necessary to make a choice regarding the agent's information set, that is, what crimes can he observe before buying a house. The regressions in section 2.3 provide a compelling argument for conditioning information to 60 days before the sale. We use this as our preferred specification through out the rest of the section, but still show estimates of the MWP for different assumptions.

Few estimates of the MWP to avoid property crimes are significant even at 10%. It seems that Chicagoans have little preferences over property crimes. The estimates of the MWP to avoid violent crime, on the contrary, are all significant at a 1% level. The numbers seem really low at first, specially if compared to the coefficients we observed in section 2.3: for our preferred specification, with 4 states and conditional on 60

¹⁸ One possibility is that these changes actually reflect seasonality in crimes, which we are not accounting for at this point. Alternatively, it could be that the number of states is not enough to get a reasonable model. In this case, if a considerable part of the sample has their "true" λ between two of the estimated values, it could seem that states are changing when it is really that none of the states fit the true distribution of this subsample well enough.

¹⁹ We assume a monthly discount rate of 0.995.

Table 2.26: MWP to avoid a violent crime (x100), coef and t-stat

# States	30 days	60 days	90 days
2	-0.0503	-0.0492	-0.0407
	-3.2769	-3.2030	-2.6438
3	-0.0478	-0.0487	-0.0442
	-6.5710	-6.6960	-6.0697
4	-0.0301	-0.0318	-0.0270
	-5.5159	-5.8374	-4.9619

Table 2.27: MWP to avoid a property crime (x100), coef and t-stat

# States	30 days	60 days	90 days
2	-0.0042	0.0006	-0.0018
	-1.4778	0.2296	-0.6177
3	-0.0023	-0.0015	-0.0032
	-1.2618	-0.7919	-1.6960
4	-0.0025	0.0012	-0.0020
	-1.8345	0.8859	-1.4391

days prior to sale, if the agent knew there would be 1 additional crime tomorrow, he would like to pay -0.032% less for the house.

The estimates from this section and section 2.3 cannot be compared though, since they refer to very different channels. The coefficients in section 2.3 reflect solely how much agents learn about a house. Since those crimes already happened by the time the house is sold, they do not enter agents' preferences directly. All the effect is through the change in beliefs. The MWP, on the contrary, is the effect of having one additional crime tomorrow for sure, given that this crime provides no additional information about crimes in the following periods. In this context, the small MWP sounds more reasonable: this one additional crime is happening at a 0.4 mile radius from the agent's house, and thus he is very unlikely to be the victim, and it does not give him any additional information about the future.

A simple exercise can show how the small MWP translates into a much higher

Table 2.28: Contribution to price when certain to be in state (%)

States	1	2	3	4
%	-0.053	-0.0765	-0.1151	-0.1469

effect of past crimes on house prices. Assume an agent has a flat prior over a house in the beginning of period 0, that is, he gives the same probability for the house being in each state, 25%. There are two scenarios. In the first, he observes 0 crimes around the house, updates his beliefs, and decides what is the value of the house at the beginning of period 1. In the second scenario, he observes 1 crime, updates his beliefs, and decides again the house's value.²⁰ For a MWP of -0.032%, the second scenario leads to a price 1.9 percentage points lower than the first. The effect is so large because this additional crime in the past tells something not only about crime tomorrow, but also every period in the future. It is also much larger than the coefficients in section 2.3 because in this exercise the agent starts from a flat prior, that is, he has no information about the house. In the real data, agents usually have more precise beliefs when they observe the crime, and as a consequence do not update their beliefs as much.

Another way to evaluate our estimates of the MWP is to compare the prices of houses that differ solely in the beliefs over them. Table 2.28 does this by comparing the prices of a house when agents are certain it is in each of the four states. The values are relative to a house that faces no crimes in the future. The difference in prices is striking: a house believed to be in state 4 must sell at a discount of almost 10 percentage points compared to a house in state 1. Although the effect of one single crime is small, when we look at the present discounted value of crimes it adds up to a considerable size.

²⁰ In both scenarios he must account for possible changes in the state after he observes the number of crimes.

Table 2.29: MWP to avoid violent crime today (x100), assuming no state transitions, coef and t-stat

# States	30 days	60 days	90 days
2	-0.0042	-0.0038	-0.0039
	-2.3255	-2.0694	-2.1401
3	-0.0140	-0.0141	-0.0135
	-7.7120	-7.7477	-7.4250
4	-0.0085	-0.0088	-0.0088
	-4.5531	-4.6684	-4.6986

We check the importance of the assumption that states may change by estimating the MWP to avoid violent crimes using an identity matrix instead of the estimated transition matrix (we use the same initial distributions and Poisson rates as in table 2.26). The results are presented in table 2.29. The estimates are between three and ten times lower, depending on the number of states used. This striking difference is a direct consequence of assuming houses are stuck in one state. Under this assumption, houses in dangerous areas are expected to have high crimes forever, making the expected discounted flow of crimes for them larger than it would be if states were allowed to change, while the opposite is true for houses in safer areas. Thus, the same amount of variation in house prices is being explained by a larger variation in expected present discounted flow of crimes.

2.5.1 Comparing structural and reduced form estimates

In section 2.3 we showed how the effect of crime on house prices fades with time. This pattern helped motivate our model. If the distribution of crime in a neighborhood is fixed, observed crimes contain the same amount of information independent of when they are observed. As a consequence, we should see house prices reacting to past crimes with the same strength, for all lags. Moreover, as the sequence of observed crimes increases, the effect of each single crime on house prices should converge to zero. By allowing the distribution of crime to change over time we change this

result. In our model, as time passes it becomes increasingly more likely that the distribution of crime has changed. For this reason, the information contained in each signal depreciates with time.

We finish this section by testing if the speed with which information depreciates in our model is comparable to what we observe in the data. We do so by creating simulated crime data and house prices, and comparing the coefficients when we regress house prices on monthly crimes in this dataset and in our actual data. Since our estimates of MWP for property crimes are not significant, we do this exercise only for violent crimes.

We start by using our estimated LMM to simulate a sequence of 29 states for 200000 houses. We use our preferred specification for this exercise, so the initial states are randomly drawn from the initial distribution in table 2.13, and the remaining states are drawn from the transition matrix in table 2.12, conditional on the previous state. We then draw a random sequence of monthly crimes for each house, in each period, using the vector Λ in table 2.14.

We now have a simulated crime data that is structured in the same way as our data. The next step is to forecast house prices. Using the steps described in section 2.4 we calculate the expected present discounted number of crimes and multiply it by the MWP to avoid crime to get prices for all houses. We use a MWP of -0.000318, our estimate for violent crimes, assuming four states and that information is observed until 60 days before the sale.

We finish by regressing simulated monthly crimes on simulated house prices. Figure 2.14 superposes these estimates (red line) to the coefficients from the actual data (bar plots). Note that we include crimes up to 24 months before the sale in these regressions, instead of 12 months.

The size of coefficients from the simulated and true data seem to decrease at a very similar rate. This is not a trivial result. The original coefficients were estimated

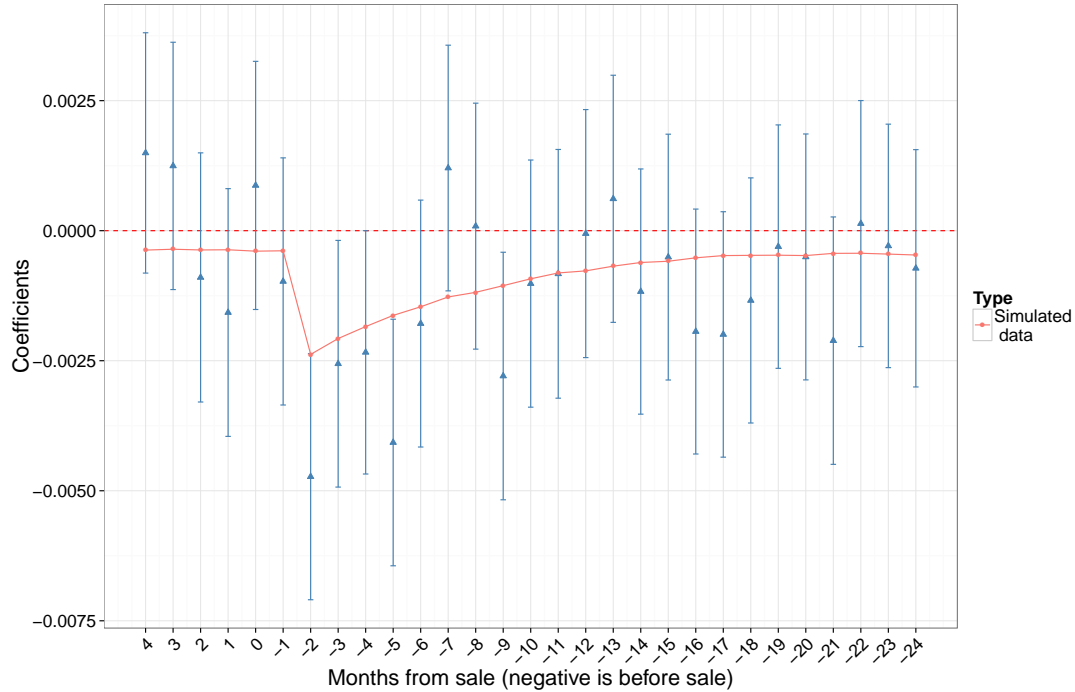


FIGURE 2.14: Violent crimes at a 0.4 mile radius, Census Block Group by quarter FE. Dependent variable is log-price. Each dot is the percentage effect of an additional violent crime. Negative months indicate crimes before the sale.

straight from the data, without any reference to the LMM. The relative size of the coefficients from the simulated data, on the other hand, is completely determined by the LMM. The rate at which they decrease is almost entirely a product of the estimated transition matrix, which regulate how past crimes enter current beliefs. That is, even though according to the transition matrix the probability of changing state at any month is not very high, it is enough to reproduce the pattern we see in the data.

Finally, the model seems to understate the initial impact of crimes on house prices, at $t = -2$. The estimate for the simulated data is at the upper bound of the 95% confidence interval, so the difference is not statistically significant. Still, it suggests agents have a stronger reaction to very recent crimes than the model predicts.

2.6 Conclusion

In this paper we estimated the effect of past crimes on house prices and the MWP to avoid crime for the city of Chicago. We used very rich data on reported crimes and house transactions, which allowed us to explore very detailed fixed effects (census block group by year by quarter) to control for any omitted variables.

In the first part of the paper, we investigated different aspects of the relation between past crimes and house prices that the earlier literature could not approach. Most of the papers so far had to resign themselves to estimate the effect of crimes aggregated in large, pre-specified geographical areas. We added to this literature by showing how the impact of a crime on house prices diminishes with the distance to the house, and that for the city of Chicago houses more than 0.4 mile away from a violent crime are not affected by it. Property crimes do not have a statistically significant impact at any distances we could test.

We were also the first paper to look at the effect of past crimes with higher than annual frequency. We found that one additional violent crime will reduce house prices by 0.5% two months before the sale. The effect is smaller for later months, and becomes insignificant after five months. This effect is heterogenous across types of crime. Homicides have by far the largest impact, with one crime reducing house prices by 1.5% two months before the sale; it remains the largest effect even if we consider an increase of one standard deviation instead, which is considerably lower for homicides than other crimes.

In the second part of the paper we explained why the coefficients of past crimes on house prices do not tell us much about the MWP to avoid crime. We went on to develop a model to estimate the MWP. The model is grounded on two basic assumptions: agents are forward looking, and for this reason try to learn about the probability distribution of crime for each house, and the distribution of crime may

change over time. We model the distribution of crimes as a Latent Markov Model. This is a flexible framework that can be easily adapted to different models of the home buying problem.

Our estimate of the MWP to avoid crime is small but significant. This result is consistent with large price differences due to expected crime: a house that agents are certain to be in the most dangerous state will sell for 10 percentage points less than that agents are certain to be in the safest state. Moreover, we showed that the model estimates fit the patterns of past crimes and house prices fairly well.

All in the Mix: Spillovers and the Agglomeration of Neighborhood Retail

3.1 Introduction

Agglomeration is the most visible feature of the retail landscape in cities. Upscale apparel shops commonly concentrate in streets such as Fifth Avenue in New York, Bond Street in London, or Rue de Rivoli in Paris. Furniture stores provide another example of shops that tend to cluster. The examples of agglomeration are not restricted to similar businesses: shopping malls and main streets thrive on a diverse mix of retail stores and restaurants that both attract and are attracted by a constant stream of foot traffic.

Businesses concentrate because of positive spillovers that arise from agglomeration. For example, retail stores in busy streets may profit from shared foot traffic: customers are attracted to the diversity of options and often make unplanned visits to multiple stores.¹ Agglomeration also reduces consumer's search costs in the pres-

¹ Gould et al. (2005) shows that shopping mall managers explicitly account for shared foot traffic. Large retail chains attract the most customers, and thus get discounts on rents and other favorable contract terms compared to small stores.

ence of imperfect information (Wolinsky, 1983; Dudey, 1990). Apparel and furniture stores are examples of businesses that can benefit from reduced search costs. Of course, not all spillovers are positive. Competition is an example of a spillover that discourages business agglomeration.²

Spillovers are of particular interest to policy makers seeking to revitalize struggling neighborhoods. Revitalization plans often cite an increased supply of retail stores and restaurants as necessary to attract more white collar jobs and residents to those neighborhoods.³ In essence, these plans see struggling neighborhoods as being trapped: businesses are not attracted to them because there is not enough demand, while shoppers avoid them because there are not enough businesses. Revitalization plans attempt to solve this coordination problem partly by incentivizing the entry of new retail. Their assumption is that a critical mass of retail stores and restaurants will make the neighborhood desirable to residents and workers once again. An important question then is if spillovers are strong enough to sustain this new equilibrium.

In this paper, I develop a dynamic entry model to estimate the spillovers between retail and service establishments. Businesses' profits depend directly on the number of stores of different types in their immediate neighborhood. I allow spillovers to vary across business types, which accommodates the different mechanisms behind retail agglomeration.

A key feature of my model is that it explicitly incorporates landlords. A business

² Seim (2006) demonstrates that competition leads to considerable spatial differentiation in the video rental industry and Bresnahan and Reiss (1991) shows that competition curbs the number of establishments in a county for different retail and service industries.

³ The city of Buffalo makes this point in its downtown revitalization plan: "It will be a great Downtown when [...] residents, workers and visitors are able to purchase clothing, gift items, groceries, and other merchandise days, nights and weekends; it is more exciting because ground floor retail space is fully occupied and thriving; retail mix is competitive with other options in the region; retail adds economic value and tenant amenities to the building stock" (City of Buffalo, 2003)

must negotiate a lease with his landlord before it can open. The landlord internalizes the spillovers that the business would generate to his other storefronts, and accounts for those spillovers during the negotiation. As a consequence, landlords have an incentive to bring businesses that will contribute to the neighborhood.

An important challenge to estimate spillovers is the presence of unobserved local attributes that shift either demand or costs for certain businesses independently of spillovers. These local attributes act as natural (dis)advantages to certain types of businesses. For instance, a touristic attraction creates demand for certain types of retail stores and restaurants, and will support a concentration of stores that is very different from what is found in a residential area, both in terms of the quantity and the types of stores. If the estimation procedure ignores the role of unobserved attributes, it will mistakenly attribute to spillovers the effect on agglomeration due to unobserved attributes.

I explore a new source of variation in the data to separate spillovers from unobserved attributes. I argue that comparing the behavior of two different of landlords in the same location, one who owns multiple storefronts and one who does not, uncovers information on spillovers, while keeping unobserved attributes constant. Since landlords internalize entrant's spillovers to their other storefronts, landlords who own multiple storefronts will encourage the entry of businesses that will improve the value of all their storefronts. The same is not true for single storefront owners. I find that the data corroborates this claim. For example, apparel stores, which tend to agglomerate, are even more likely to do so in jointly owned storefronts. I discuss the intuition for the identification strategy at length in section 3.2.

Another challenge in estimating my model is the size of the state space. The value functions of businesses and landlords depend on a large number of state variables: (unknown) cluster type, the ownership structure of storefronts, the number of storefronts surrounding it, the businesses occupying them, the number of storefronts

there are surrounding each neighbor, and so on. Thus, it is not feasible to compute the value function using the standard methods. I overcome this challenge by applying the Sieve Value Function Iteration method to approximate the value function over a subset of state variables and functions of those variables (Arcidiacono et al., 2013).

I estimate my model using data on openings of new retail businesses in San Diego, CA, from 2002 to 2011. I obtained the data from business license records at the Office of the City Treasurer of San Diego. The data lists businesses that held a Business Tax Certificate in San Diego during the period. I aggregate businesses in four categories: apparel shops; restaurants, bars and coffee shops; nail salons and barber shops; and other retail, a category to control for the remaining types of retail. The data on storefront owners comes from the San Diego Assessor's Office, and I match businesses to landlords by merging businesses to the closest parcel.

I find that spillovers are extremely important to explain business agglomeration. If businesses' profits were stripped of spillovers, the share of vacant storefronts would increase by about 17 percentage points. Spillovers are the strongest between businesses of the same type, specifically for apparel stores and other retail, which indicates that search costs may be the main driver behind agglomeration of local businesses. The exception is personal care establishments, which do not seem to collocate. Unobservable characteristics of the neighborhood play an important role in shaping the composition of businesses in a neighborhood as well. I show that explicitly controlling for unobserved attributes has a strong effect on significance of the estimates of spillovers across business types, but it does not have a large impact on spillovers between similar stores.

I use my estimates of spillovers to evaluate two counterfactual experiments. First, since businesses do not internalize the spillovers they create, the business composition in a cluster is often sub-optimal. A landlord who owns a large share of storefronts,

and thus internalize some of the spillovers, can solve the coordination problem and move the cluster to a better equilibrium. I test this hypothesis by looking at how the expected profits in a cluster change as the share of jointly owned storefronts increases. In my simulations, expected profits increases between 2% and 30% when all storefronts are owned by the same landlord, compared to the case when all storefronts are independently owned.⁴ The results confirm the importance of storefront ownership to cluster outcomes. The increase in expected profits is due to the better match between neighboring stores, a direct consequence of landlords internalizing spillovers.

Second, I evaluate the government’s ability to increase business density in a neighborhood. Many authors have discussed at length why urban density is a desirable feature for cities (Jacobs, 1961; Glaeser, 2011). New businesses create jobs, provide amenities to residents and increase the number of eyes on the street, which may reduce crime. The question I investigate is whether spillovers can create enough demand for retail and restaurants following the subsidies to business entry that the high occupancy equilibrium remains profitable even after the subsidies are withdrawn. This question is related to a vast a literature investigating if a “big push” through large, place-based policies can be successful in moving regions or cities to a different equilibrium (Glaeser and Gottlieb, 2008; Kline, 2010; Kline and Moretti, 2014).

I test this by simulating a five year large subsidy to entry costs (between 65% and 100% of entry costs) to a single neighborhood in San Diego. I find that subsidized neighborhoods are not better off than non-subsidized ones in the long run. Interestingly, I show that the reason for this is not that spillovers are too weak, but that entry costs are too low. As a consequence, there are always enough stores willing to

⁴ As I show in section 3.7, the wide range in the increase of average profits is a result of other neighborhood characteristics.

open and wait to see if other stores will come, and there are no coordination failures for the subsidy to solve.

3.1.1 Literature Review

This paper is part of a body of work that has looked at spillovers in retail and service stores. Most papers in this literature have focused on spillovers created by large retail stores. Vitorino (2012) estimates a structural entry model for large chain stores in shopping malls, and finds evidence of positive spillovers between midscale chain stores. Benmelech et al. (2014) and Shoag and Veuger (2014) both use stores closing due to national chain bankruptcies to estimate their spillovers on surrounding businesses. They find that the number of stores and people employed decreases significantly following store closings. Unfortunately, their identification strategy relies on establishments belonging to a national chain and thus is not applicable to study small establishments.

Cosman (2014) is the paper closest in spirit to this one. It uses a continuous time entry model to estimate spillovers between nightlife venues in Chicago. The paper finds large positive effects, but fails to account for unobserved attributes. My paper contributes to this literature by providing estimates of spillovers between small businesses of different types, while offering an identification strategy that is suitable to the analysis of small establishments and takes into account unobserved local attributes.

This paper also contributes to the literature that studies the consequences of land concentration for different neighborhood outcomes (Brooks and Strange, 2011). To the best of my knowledge, this is the first paper to investigate the role of storefront ownership on business composition, and to show how storefront owners influence the business composition in equilibrium.

Finally, this paper is also part of a long literature in Industrial Organization

that uses information on firm entry to recover parameters of the profit function (Bresnahan and Reiss, 1991; Seim, 2006; Davis, 2006; Jia, 2008; Dunne et al., 2013; Arcidiacono et al., 2012). The main intuition behind these models is that firm’s decisions of when and where to enter reveal information about their preferences over observable market characteristics, even in the absence of information on prices and revenue. In particular, I base my model on Arcidiacono et al. (2012), with entries and exits taking place in continuous time.

The paper is organized as follows. In section 3.2 I present a simple example that illustrates how landlords influence the probability that certain businesses will enter. Section 3.3 explains how several different data sources were combined and presents reduced form evidence that storefront ownership is relevant to explain business entries. I introduce the full model in section 3.4 and discuss the estimation strategy in section 3.5. In section 3.6 I discuss the results from the structural estimation, and in sections 3.7 and 3.8 I present present the two counterfactuals. Section 3.9 concludes.

3.2 A Static Example of Entry

A key feature of my model is that landlords internalize the spillovers from their tenants. In this section, I show how this feature affects landlords’ preferences over the entry of different business types, and argue that comparing the behavior of landlords who own multiple multiple storefronts and those who do not allows me to separate spillovers from unobserved attributes. In the spirit of my model, I lay out a static example in which a landlord is matched to a business owner and must decide to lease him a storefront or not. This section imposes some additional assumptions that are relaxed in section 3.4, but the intuition carries through.

There are two types of players. Businesses are either restaurants or barber shops, identified by the subscripts r and b respectively, and must rent a storefront to earn profits. Landlords own one or more storefronts and extract all the business profits

as rent.⁵ Thus, their goal is to maximize the total sum of profits in their storefronts. Profit functions are:

$$\begin{aligned}\Pi_r(I_r, I_b, \xi_r) &= 1 \cdot I_r + 0 \cdot I_b + \xi_r + \varepsilon \\ \Pi_b(I_r, I_b, \xi_b) &= 2 \cdot I_r - 0.5 \cdot I_b + \xi_b + \varepsilon\end{aligned}$$

where I_r and I_b are the number of surrounding restaurants and barber shops, ξ_r and ξ_b are the effects of unobserved local attributes on profits, which affect all businesses in the cluster, and $\varepsilon \stackrel{iid}{\sim} \mathbb{N}(0, \sigma)$ are idiosyncratic profit shocks. The coefficients imply that restaurants produce positive spillovers to both business types. Barber shops, on the other hand, do not produce any spillovers for restaurants, and steal profits from other barber shops. Vacant storefronts do not earn any returns to the landlord, nor create spillovers to businesses.

An important challenge in estimating spillovers is the presence of ξ 's. Since ξ varies between clusters, restaurants and barber shops will be relatively more common in clusters where ξ_r or ξ_b are larger. As a consequence, ignoring ξ 's may lead to biases in the estimates of spillovers between businesses of the same type.⁶

In this section, I focus on situations in which the landlord must decide to rent a single vacant storefront to a specific business type, taking the current mix of businesses in the cluster as given. This set-up is close to the sequential game played in the continuous time model in section 3.4.⁷ If the landlord leases the storefront, the business will generate profits, which will be payed to the landlord as rent. Since the

⁵ This is obviously an unrealistic assumption. Its goal is to align landlords' incentives with profit maximization in a straightforward way. In section 3.4 I show that landlords have similar incentives if the rent is decided by Nash bargaining, when they will be maximizing the aggregate value of all his storefronts.

⁶ If ξ_r and ξ_b are correlated, unobserved attributes will also induce biases in the estimates of cross-spillovers. A positive correlation would lead to upward biases both in the spillovers from restaurants to barbers shops and from barber shops to restaurants. A negative correlation would generate a downward bias instead.

⁷ In section 3.4, landlords and businesses foresee that contracts will be renegotiated.

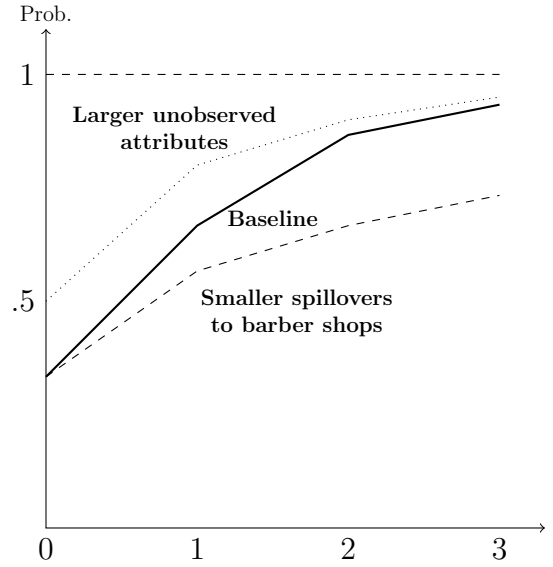
game is static, if the storefront remains vacant, neither the landlord nor the potential entrant receive anything.

The probability that a storefront is leased depends on the expected profit by the entrant, and how its entry will affect the profits of the other businesses leasing from the landlord. Thus, the probability depends on the mix of businesses in the cluster in two related ways. Firstly, since incumbents produce spillovers, they affect the potential entrant's expected profit. Secondly, since incumbents will be affected by the entrant's spillovers, in so far as any of those incumbents are also leasing from the entrant's landlord, changes to their profits must be accounted for.

For instance, assume that the landlord has the option to lease a storefront to a restaurant. There are three other storefronts nearby, all occupied by barber shops. The restaurant's profit is $\Pi_r = \xi_r + \varepsilon$, and the probability that the landlord will accept the offer when he does not own any of the remaining storefronts is $P(\xi_r + \varepsilon > 0)$. But the new restaurant will generate positive spillovers to all the barber shops, increasing their profit by 2. If the landlord owns one of the other storefronts, the probability that the restaurant enters increases to $P(\xi_r + \varepsilon + 2 > 0)$.

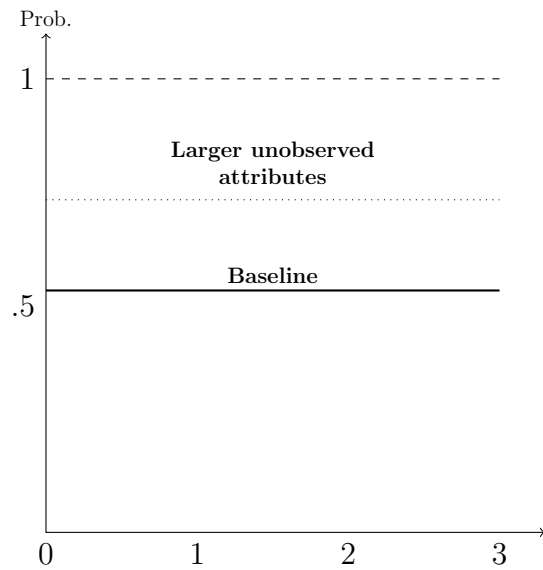
Picture 3.1(a) illustrates how the probability changes as the landlord progressively owns more storefronts. The solid and dashed lines present the probabilities for the same unobserved attributes, but for different amounts of spillovers. The solid line uses the parametrization from the equation above, while the dashed line assumes smaller spillovers from restaurants to barber shops. Stronger spillovers produced by the entrant translate to steeper changes in probability for each additional storefront owned by the landlord.

The dotted line uses the same spillovers as the solid one, but has larger unobserved attributes ξ_r . Entry probabilities increase for all quantities of jointly owned storefronts. As we can see, comparing how much probabilities vary between different clusters holding the number of jointly owned storefronts constant reveals information



Surrounding storefronts owned by central landlord (k)

(a) restaurant surrounded by 3 barber shops



Surrounding storefronts owned by central landlord (k)

(b) barber shop surrounded by 3 restaurants

FIGURE 3.1: Entry probabilities as a function of storefront ownership

about the unobserved local attributes ξ .

It is also useful to examine how the probability curve is shaped if the potential entrant is a barber shop and the incumbents three restaurants. This case is presented in picture 3.1(b). Since barber shops do not create any spillovers to restaurants, the

curve is flat. Changes in storefront concentration do not affect the entry probability.⁸ In summary, these two examples clearly show how comparing entry probabilities for different landlords in the same cluster reveal information on the spillovers created by the entrant. Comparisons between clusters while holding storefront ownership and the composition of surrounding businesses constant reveal information about unobserved attributes.

3.3 Data

Section 3.2 illustrated how landlords influence business entry. This section provides empirical evidence to support this claim. I start by describing the different data sources. Then, I move on to some empirical evidence on clustering and mixing within San Diego and how it is affected by joint storefront ownership.

3.3.1 Description

I obtained business data from business license records at the Office of the City Treasurer of San Diego. It lists businesses that held a Business Tax Certificate in San Diego starting in 1986, but to match the timespan of the other datasets I restrict it to businesses that were active at some moment between January 2002 and December 2011. The dataset has information on business and owner names, ownership type, the dates it started and stopped paying taxes in San Diego, its NAICS code, business phone number, and address.⁹

From this data I extract a subset based on NAICS codes and regroup them to form 4 broad groups: (1) apparel shops, (2) restaurants, bars and coffee shops, (3)

⁸ This is true because the model it is a single-period model. If the landlord had the outside option of searching for a different tenant, then each additional storefront owned would lead to a decrease in the entry probability for barber shops, since it would become increasingly appealing to wait for a restaurant.

⁹ Unfortunately the exit date available from the City Treasurer is not the actual exit date, but the date its last license expired. Since licenses must be renewed every twelve months, this date provides a 12-month interval during which the business closed.

personal care services (beauty parlors, nail salons and barber shops), and (4) other retail stores (furniture stores, electronic stores, bookstores, pharmacies and beauty shops, food and beverage stores).¹⁰

I use data on parcel ownership to match businesses to landlords. This data comes from SanGIS, an authority connected to the City of San Diego and the County of San Diego which is responsible for keeping a data warehouse with several geographic files for the area.¹¹ The map is from March 2014 and contains all parcels in the city of San Diego, with some basic information from the Assessor's Office. Most importantly, it contains the name of the parcel owner for each parcel.¹²

Buildings may have multiple owners if different people own a share in it, including apartments or floors. It is impossible to know who has the right to rent space to commerce in the building. I circumvent this problem with an algorithm to choose the "building owner" that maximizes joint ownership.¹³

To match businesses to landlords, I first geocode the business data using Google's service. The success rate for the geocoding is close to 100% (see table 3.2). Then, I merge the business and the parcel datasets by matching each business to its closest parcel. Since Google has rooftop accuracy for most addresses, 88% of them fall within a parcel, and 95.7% are within 30 feet. Table 3.1 breaks down the number of

¹⁰ Workers in the personal care services often register a company in their name and rent a booth at their workplace. This creates an extra layer of complexity in identifying such services. First, some workers register their home address instead of their work address in their companies. Second, several businesses will be registered at the same addresses at any point in time. As I explain later, I only use businesses in clusters with at least 5 storefronts, thus the first issue should not be a big concern. As for the second point, I assume that all personal care services registered to the same address at the same point in time are a single establishment.

¹¹ <http://www.sangis.org/>

¹² At this point, I do not use a panel data on parcel owners, so I match all businesses to the 2014 parcel owners.

¹³ The algorithm involves a set of criteria for narrowing down the pool of owners in a propriety, listed here in the order they are used. First, keep only non-government owners if any. Then, check if there are one or more names who own stakes in at least one other parcel. If there is still more than own name, see if any are a company, and keep them. Finally, keep the name who owns the most parcels in the city of San Diego.

Table 3.1: Number of business-parcel matches by distance (feet)

	Total	Percentage
0	13,558	88
(0,30]	1,206	7.8
(30,60]	576	3.7
(60,90]	29	0.19
(90,135]	15	0.097
Not matched	4	0.026

Table 3.2: Total business observations before and after geocoding, 2002-2011.

	Initial Total	Success rate	Final Total
Apparel Stores	2,307	100	2,298
Other retail	4,316	99	4,293
Personal Care	2,363	100	2,354
Restaurants	5,336	100	5,327

businesses by match distance.

Unfortunately, there is no information on the number of storefronts per address. I solve this problem by using the maximum number of businesses I observe at the same time in each address. Because I see only an approximated exit date, I work with a lower bound for the number of storefronts in each parcel, calculated by assuming that the exit date is the maximum between the entry date plus one day and the exit date minus 365 days, since licenses should be renewed yearly.

For the demand side, I use the LEHD Origin-Destination Employment Statistics (LODES) dataset to collect the number of working residents within 800 ft and 1600 ft of each storefront for the years of 2002 through 2011.¹⁴ LODES is built from

¹⁴ LODES uses synthetic data methods to protect confidential information about workers. The dataset should have similar properties to the original one, but some patterns may not hold at very small scales. I test this by calculating the correlation of total working residents in the 800 ft buffers from the 2010 LODES and total population in the same 800 ft buffers from the 2010 Census and find a correlation of around 0.9.

the Longitudinal Employer-Household Dynamics (LEHD), a comprehensive national panel of employers and employees maintained by the US Census (Abowd et al., 2004), and includes the number of workers and working residents for every Census block in the United States, as well as the number of commuters between each pair of Census blocks in the same state. LODES also classifies workers by broad industry, age, and income categories. In particular, I divide working residents between those who get paid more and less than \$ 3,330.00 a month, the highest income group available at LODES.

3.3.2 Business Clusters

I group business addresses into clusters. As in section 3.2, the assumption for the remainder of the paper is that unobserved attributes are constant within the cluster. As a consequence, it is desirable that clusters be small, and neighborhood characteristics within each homogenous. On the other hand, clusters must be large enough so that there is enough power left to estimate spillovers.

I define clusters so that any two businesses within 300 feet of each other are in the same cluster. I choose 300 feet because it is the distance used for the spillovers in the structural estimation. As an illustration, picture 3.2 presents three clusters in my data, each identified by dots of a different color. Storefronts outside the three clusters are colored black. These three clusters were chosen because of their number of storefronts, 16, which is similar to the median cluster in my sample. Even though these clusters are very close to each other, they show very different mixes of businesses. The lower left, green, cluster has mostly apparel and other retail stores, the upper left, blue, cluster has several personal care establishments, such as nail and hair saloons, some retail and restaurants, while the red cluster on the right has a larger concentration of restaurants. These differences support the idea that spillovers between small businesses are very local.

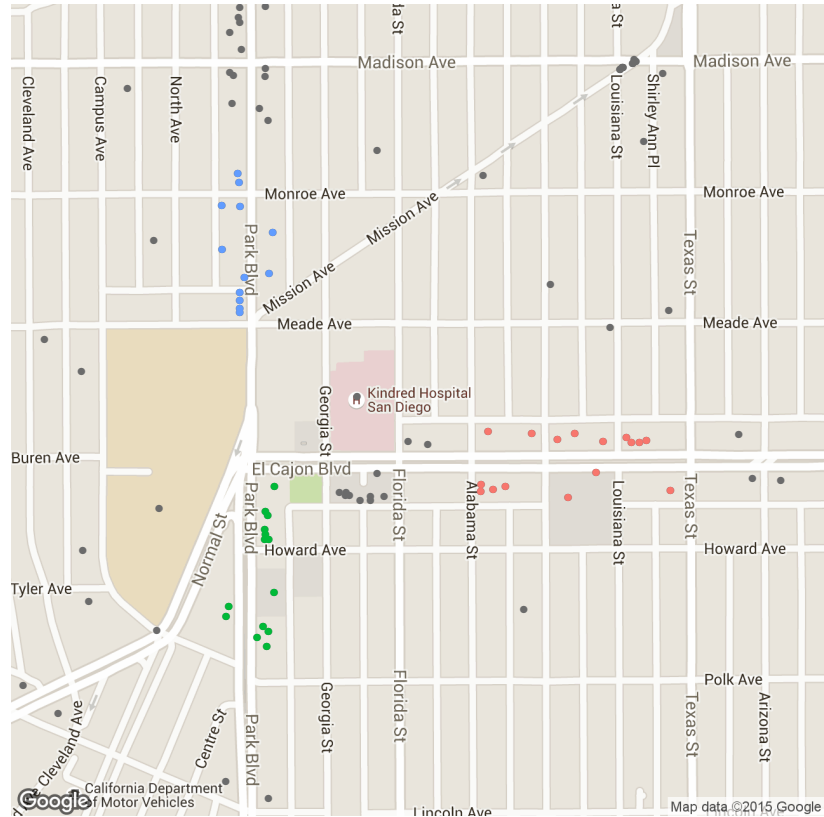


FIGURE 3.2: Three business clusters in San Diego. Red, green and blue dots represent storefronts in each of the three clusters. All other storefronts in the area are marked as dark dots.

The main advantage of this cluster definition, which I refer to as the main definition, over using pre-determined geographies such as Census Tracts, is that the main definition produces more compact clusters than Census Tracts or traditional neighborhood definitions would. This is clear in picture 3.2: the three clusters are small, occupying at most a few blocks, mostly in the same street. In general, I find that the median cluster with the main definition in San Diego has less than half the area and two thirds of the numbers of storefronts of the median cluster using Census tracts borders (see table 3.3).¹⁵

¹⁵ I define the area of a cluster as the area of the convex hull that covers all of its storefronts.

Table 3.3: Cluster definitions, total storefronts and area (sqft). To ensure that both cluster definitions are comparable, only Census tracts with more than 10 storefronts were used in this table. Area is the area of the convex hull that covers all storefronts in a cluster. 5 clusters in the main definition and 14 Census tracts have only 1 or 2 storefronts. Those clusters have an “area” of 0.

	Total Storefronts		Area (sqft)	
	Cluster	Census Tract	Cluster	Census Tract
Min.	11	11	0	0
1st Qu.	12	17	65,150	72,760
Median	17	24	177,200	390,800
Mean	31	37	539,000	2,323,000
3rd Qu.	26	37	370,400	2,204,000
Max.	485	320	13,420,000	44,530,000

3.3.3 Descriptive Statistics

In the remainder of the paper, I exclude from my sample any clusters with less than 5 storefronts. I also exclude any entries into parcels with more than 25 storefronts, as well as any entries within 300 feet of those storefronts. The goal of excluding large parcels and their neighbors is to avoid using shopping malls in my estimation, since spillovers may affect stores in shopping malls and their neighbors very differently than they would if those stores were on the street.

Table 3.4 presents descriptive statistics for the number of surrounding businesses by type for the entry spells in the sample. For convenience, the numbers were computed at the end of each spell.¹⁶ The table is split in two groups of columns with mean and standard deviation for different distance ranges: a 300 ft radius circle and a ring between 300 ft and 500 ft around the storefront. There are two rows for each business type. The row labelled “All” includes all businesses within range, while the second row, labelled “Same landlord”, counts only those businesses with the same landlord as the entrant.

¹⁶ Table 3.6 presents descriptive statistics for the remaining control variables.

Table 3.4: Number of surrounding businesses at the time of entry

	300 ft circle		300 ft - 500 ft ring	
	Mean	Std. Dev.	Mean	Std. Dev.
Total				
All	4.6	5.1	4.2	6.2
Same landlord	0.82	1.4	0.072	0.51
Apparel				
All	0.62	1.5	0.69	2
Same landlord	0.093	0.44	0.011	0.25
Other retail				
All	1.1	1.4	1.1	1.6
Same landlord	0.22	0.59	0.016	0.14
Personal care				
All	1.1	1.4	0.95	1.6
Same landlord	0.16	0.39	0.018	0.15
Restaurants				
All	1.8	2.4	1.5	2.6
Same landlord	0.34	0.77	0.026	0.21

Table 3.5: Description of control variables

Variable	Description
Historical	1 if part of historical district
Coastal	1 if within 800 ft of the sea
log(Area)	Log of the parcel's area
Parking lots, 300ft	1 if any parcels within 300 ft with parking lot land use code
3-4 floors	1 if parcel's use code is for 3 or 4 floors
Shopping center	1 if parcel's use code is for shopping centers
log(1 + Residents), 800ft	Log of 1 plus total residents within 800 ft
log(1 + Residents), 1600ft	Log of 1 plus total residents between 800 ft and 1600 ft
High income share, 800ft	Share within 800 ft earning more than \$ 3,333.00 per month
High income share, 1600ft	Share between 800 ft and 1600 ft earning more than \$ 3,333.00 per month
No residents, 800ft	1 if there are no residents within 800 ft
No residents, 1600ft	1 if there are no residents between 800 ft and 1600 ft

Table 3.6: Control variables at time of entry

	Mean	Std. Dev.
Historical	0.14	0.35
Coastal	0.08	0.27
log(Area)	10	1.5
Parking lots, 300ft	0.46	0.5
3-4 floors	0.58	0.49
Shopping center	0.17	0.38
log(1 + Residents), 800ft	4.7	2.1
log(1 + Residents), 1600ft	6.1	1.9
High income share, 800ft	0.34	0.2
High income share, 1600ft	0.37	0.17
No residents, 800ft	0.13	0.34
No residents, 1600ft	0.061	0.24

The share of businesses with same landlord within 300 ft is slightly under 50%, and decreases considerably in the outer ring. This shows that landlords are more likely to own storefronts that are close together, and is partly explained by buildings with multiple storefronts. A feature of the data is that while the average number of neighbors of each type is low, their standard deviation is pretty large, which demonstrates that the distribution of the number of surrounding businesses has a long right tail, an evidence of agglomeration.

Table 3.7: Time until a vacant storefront is leased, by entrant type (in years). Right censored spells are those storefronts that remained vacant by the end of the sample period.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Total spells
Total	0.0027	0.42	2.1	3.3	5.4	12	8,821
Apparel	0.0027	0.27	1.6	2.7	4.3	12	680
Other retail	0.0027	0.15	1.4	2.6	3.9	12	1,312
Personal care	0.0027	0.84	2.5	3.7	5.6	12	664
Restaurants	0.0027	0.025	1	2.2	3	12	3,139
Right censored	0.027	1.5	3.9	4.8	7.6	12	3,026

Table 3.7 shows the total number of vacant storefront spells and their duration by the entrant's type, including right censored spells. Restaurants are by far the most common business, comprising more than half of non-censored spells. Restaurants enter faster on average, with the average entry happening within 2.2 years, at least half a year shorter than for other types. Finally, right censored spells are vacant storefront spells that were still active in December, 2011. These spells are very common and have longer durations on average. This is at least partly explained by how the total number of storefronts and vacancies are inferred from the data. For instance, storefronts leased to businesses not in my sample, such as dental clinics or pet shops, will show up as vacant.¹⁷

3.3.4 Reduced form evidence

Storefronts are extremely concentrated in San Diego. The 5% of Census tracts with the most storefronts in my sample have 27% of all storefronts in San Diego (table 3.8). These numbers were calculated using only storefronts that overlay with one of my clusters, which ensures that zoning restrictions do not play a large role in explaining the agglomeration. Neither is the agglomeration explained by the number of working residents or the land area of those Census tracts. The 5% most populous and the 5% largest Census tracts in this sample account for only 6% and 10% of all storefronts, respectively. Thus, there seems to be considerable room for spillovers to explain the agglomeration.

Agglomeration also varies considerably across business types. I use all entries in my data to regress the number of businesses of each type within 300 feet of the entrant on the entrant's type. The coefficient estimates are presented in table 3.9. Rows refer to entrant's type, and columns are the different types of surrounding

¹⁷ Also, since the number of storefronts is not observed directly, it is possible that some vacant storefronts are really storefronts that ceased to exist at some point.

Table 3.8: Share of storefronts in Census Tracts ranked by different criteria. Each row ranks Census Tracts by a different criterium. Columns represent the share of storefronts in San Diego located in Census Tracts in the top 20th, 10th and 5th percentiles according to each criterium. I use only Census tracts that overlay with the clusters in my sample, to avoid the concentration due to zoning restrictions.

	20 %	10 %	5 %
Most storefronts	0.55	0.39	0.27
Most residents	0.18	0.11	0.06
Most land area	0.22	0.14	0.10

businesses. The excluded type is all regressions is Apparel stores, so the intercept is the average total businesses of a certain type within 300 ft of an apparel store the moment it opened. The remaining coefficients are the difference between the average for Apparel stores and the other types.

Table 3.9: Regression of number of businesses within 300 ft on entrant type. Columns refer to type of surrounding businesses. Rows are the entrant's type, and the reference type is apparel stores. The intercept is the mean number of businesses of the column's type surrounding opening apparel stores. * 5%, ** 1%, and *** 0.1%.

	Apparel	Other retail	Personal care	Restaurants
(Intercept)	1.22***	1.275***	1.34***	2.15***
Other retail	-0.70***	-0.059	-0.40***	-0.70***
Personal care	-0.52***	-0.263***	-0.36***	-0.55***
Restaurants	-0.55***	-0.035	-0.20**	-0.17

The results show that different types favor locations with very different business compositions. Apparel stores seem to favor agglomeration more than any other type does. In particular, the average apparel store opens next to double the number of other apparel stores than do other entering businesses. There are also other interesting findings, such as personal care establishments avoiding other retail, and both other retail and personal care avoiding restaurants. While these patterns could

be driven either by unobserved local attributes or spillovers, they strongly suggest that spillovers vary according to business types.

Finally, I use some of the insights from section 3.2 to present empirical evidence that landlords internalize spillovers. I estimate a Cox proportional hazards model in which vacant storefronts face competing risks that a business of each type may enter. The hazard for each entrant type is affected both by the total number of different businesses and how many of those share the entrant's landlord.¹⁸

The idea that the hazard rate of entry can be affected both the total number of businesses and the number of businesses with the same landlord as the entrant is analogous to the discussed in section 3.2. Assuming that spillovers exist, the potential entrant only cares about the composition of businesses surrounding the storefront. It is the landlord who considers the effects of the entry on his other storefronts. Thus, to the extent that the business composition is being controlled for, any additional effects of businesses with the same landlord on the entry hazard suggests that landlords play a role in the entry decision.

Table 3.10 shows the exponential of the coefficient estimates for businesses within 300 ft. The remaining control variables are in table 3.11. Each column shows the effects to a particular risk (entrant type). Due to the functional form of the hazard rate, the numbers can be directly interpreted as the percentage increase (or decrease) in the hazard rate of a specific business type from having one additional business within 300 ft. For instance, an apparel store increases the hazard rate that another apparel store will enter by 14%. If the two stores have the same landlord, there is an additional effect of 49%, leading to a total increase of 70% of the baseline hazard rate.¹⁹

¹⁸ A full description of control variables is available on table 3.5.

¹⁹ Alternatively, the inverse of the exponential can be interpreted as the percentage increase in the expected waiting time, conditional on no businesses of different types entering. This interpretation assumes that the arrival of different business types are independent processes.

Table 3.10: Cox Proportional Hazard model, 300 ft, exponential of coefficients. Columns refer to entrant’s types. Rows “All” include all surrounding businesses, while “Same landlord” count only businesses leasing from the storefront owner. Coefficients of control variables in table 3.11. * 5%, ** 1%, and *** 0.1%.

	Apparel	Other retail	Personal care	Restaurants
Vacant				
All, 300ft	0.94**	0.97	1.05	0.99
Same landlord, 300ft	1.02	0.83***	0.87	0.82***
Apparel				
All, 300ft	1.14***	1.02	1.02	1.03
Same landlord, 300ft	1.49***	0.91	1.18	0.89*
Other retail				
All, 300ft	0.9	1.05	0.88	1.01
Same landlord, 300ft	1.32**	1.15*	1.16	1.07
Personal care				
All, 300ft	1.13*	0.99	0.9	1.02
Same landlord, 300ft	1.7***	1.3*	0.27***	1.33***
Restaurants				
All, 300ft	1.09**	1	0.99	1
Same landlord, 300ft	0.72***	0.9	1.15	1.26***

Table 3.11: Cox Proportional Hazard control variables, 300 ft

	Apparel	Other retail	Personal care	Restaurants
Other controls				
log(1 + Residents), 800ft	1.05	1.13**	0.98	0.98
High income share, 800ft	1.54	0.37**	0.82	0.83
log(1 + Residents), 1600ft	1	0.95	1.12	1.03
High income share, 1600ft	0.22**	0.4*	1.8	0.9
log(Area)	1.08	1.11**	0.95	1.03
Parking lots, 300ft	1.4**	0.81*	0.8	1.1
3-4 floors	1.83***	1.92***	1.5*	1.05
Shopping center	1.66*	2.34***	1.38	1.12

The estimates suggest strong positive spillovers between businesses of a same type, except for personal care. The effect is particularly strong for apparel shops,

although it is important to note that these numbers may be partly driven by unobserved neighborhood characteristics. The results are less clear when it comes to spillovers across different types. It seems that apparel shops benefit from more other retail stores and personal care establishments, specially if they have the same landlord, but other businesses do not seem to benefit from spillovers as much. Still, the estimates support the assumption that landlords internalize spillovers.

Vacant storefronts have a considerable negative impact on the hazard rates, in particular for other retail stores and restaurants, with a decrease of around 20% in the hazard that one of those businesses will enter when the vacant storefront belongs to the entrant's landlord. This points to two possible conclusions. First, unobservable characteristics make certain areas inhospitable for most businesses, and this is captured in the survival model through the number of vacant storefronts. Second, it is the absence of other businesses by itself that make an area bad to enter due to lack of spillovers. That is, there is a coordination problem, and no business wants to be the first to enter.

Finally, I have expanded the Cox regression to include businesses between 300 ft and 500 ft (tables 3.12 and 3.13). The estimates for businesses within 300 feet are not changed, but I find no evidence that businesses farther than 300 feet influence entry. Thus, I focus my structural estimation on businesses within 300 feet only.

3.4 A Model of Business Location

In this section I lay out a continuous time model of business entry and exit that disentangles two different forces influencing the mix of businesses: spillovers and unobserved neighborhood attributes. Spillovers are assumed to be a very local phenomenon, and its strength will vary with the mix of surrounding businesses, while unobserved attributes are constant within the cluster and affect all businesses of the same type identically.

Table 3.12: Cox Proportional Hazard model, 300 ft and 500 ft, exponential of coefficients. Columns refer to entrant’s types. Rows labeled “All” include all surrounding businesses, while those labeled “Same landlord” count only businesses leasing from the storefront owner. * 5%, ** 1%, and *** 0.1%.

	Apparel	Other retail	Personal care	Restaurants
Vacant				
All, 300ft	0.94*	0.98	1.05	0.98
Same landlord, 300ft	1.03	0.82***	0.87	0.82***
All, 500ft	1	0.99	1	1.01
Same landlord, 500ft	1	0.71	1.2	1.07
Apparel				
All, 300ft	1.12**	1.02	1.03	1.04
Same landlord, 300ft	1.48***	0.85	1.19	0.88*
All, 500ft	1.05*	1.05*	0.99	1.02
Same landlord, 500ft	1.11	1.17	0	0.79
Other retail				
All, 300ft	0.91	1.06	0.89	1.01
Same landlord, 300ft	1.32**	1.15*	1.14	1.07
All, 500ft	1.02	1.01	0.99	1
Same landlord, 500ft	0.85	0.31	1.36	0.79
Personal care				
All, 300ft	1.09	1	0.9	1.02
Same landlord, 300ft	1.7***	1.33*	0.27***	1.33***
All, 500ft	1.06	0.91*	0.96	0.98
Same landlord, 500ft	0.54	1.47	1.76	1.11
Restaurants				
All, 300ft	1.1**	1	1	1
Same landlord, 300ft	0.71***	0.9	1.15	1.26***
All, 500ft	0.96	1.02	0.97	1
Same landlord, 500ft	1.45	0.92	1.73*	0.99

Decisions regarding entry and exit are made jointly by landlords, who own storefronts, and businesses. Landlords can own one or more storefronts, and receive rent for the leased ones. I assume that the set of landlords is given, and that storefront ownership is constant over time, that is, landlords cannot buy or sell properties.

Businesses belong to one of J types, such as restaurants or apparel shops. To

Table 3.13: Cox Proportional Hazard control variables, 300 ft and 500 ft

	Apparel	Other retail	Personal care	Restaurants
Other controls				
log(1 + Residents), 800ft	1.06	1.12**	0.97	0.98
High income share, 800ft	1.47	0.37*	0.86	0.83
log(1 + Residents), 1600ft	1	0.96	1.12	1.03
High income share, 1600ft	0.22**	0.39*	1.74	0.88
log(Area)	1.09	1.12**	0.93	1.03
Parking lots, 300ft	1.44**	0.81*	0.81	1.09
3-4 floors	1.89***	1.93***	1.48*	1.05
Shopping center	1.76**	2.3***	1.29	1.12

earn profits, a business must rent a storefront. Profits are a function of the mix of surrounding businesses, other control variables and the unobserved attribute. The parameters of the profit function may vary according to business type.

A vacant storefront is matched at a rate δ to some type- j business drawn from a pool with infinitely many potential entrants. The business in turn draws a productivity shock ε from a standard normal distribution. If the match surplus is positive, the storefront is leased at a rent r determined through Nash-Bargaining. If the match surplus is negative, the landlord will continue searching, while the business receives a pay-off of zero. The match surplus includes an entry cost e_j which I assume is paid by the business, without loss of generality.

An existing lease is renegotiated at a rate λ . At every renegotiation the business draws a new productivity shock ε' , which substitutes the older draw. If the new match surplus is positive, the lease is renegotiated at some new rent r' . If the surplus is negative, the lease is terminated, the landlord searches for a new tenant and the business pays the exit cost c_j and disappears.

In the remainder of this section I first describe the business profit function. Then, I present the landlord's and business' decision problems. I finish the section explain-

ing how the rent is determined, and introducing the entry and exit probabilities. For the rest of the paper I adopt the convention of referring to the storefront in question as the central storefront, to distinguish it from its surrounding ones. Similarly, I will use central business when referring to the business leasing that storefront.

3.4.1 The Profit Function

I model profit as a function of the number of businesses by type and distance, the number of residents by income group and distance, observable neighborhood characteristics, and two unobservable components, the neighborhood attribute and the productivity shock. This reduced form profit function is very common in the entry literature (Dunne et al., 2013; Jia, 2008; Seim, 2006).

The profit function for a type j business located in storefront l is

$$\Pi_j(l, \mathcal{C}, \xi, \varepsilon) = \sum_{m=1}^{Dist} \sum_{j'=1}^J \alpha_{j,j'}^m I_{j'}^m + \kappa_j X_l + \xi_{c(l),j} + \varepsilon$$

where $I_{j'}^m$ is the number of type j' businesses in ring m , $Dist$ is the number of distance rings used, and X_l are observed characteristics which include both demographic, neighborhood, and storefront characteristics. $\xi_{c(l),j}$ is the unobserved attribute for a type j business in l 's cluster, $c(l)$, and ε is the random productivity shock.

The coefficients $\alpha_{j,j'}^m$ are a flexible way to capture spillovers, since business types are allowed to influence each other with differing strengths. The downside of this specification is that competition is lumped together with spillovers for businesses of the same type. As a consequence, $\alpha_{j,j'}^m$ should be interpreted as the spillovers produced by a j' business on a type j incumbent, net of losses accruing from increased competition.²⁰

²⁰ Ideally we would like to separate competition from spillovers, but without information on sales and prices or number of shoppers, it is not possible to separate both without strong assumptions on the form of spillovers

In addition to observable variables, profits are also affected by the unobserved attribute in l 's cluster, $\xi_{c(l),j}$. I assume clusters belong to one of K types, defined by different vectors $\xi = \{\xi_j\}_{j=1}^J$. This setup accounts for the fact that some clusters may have unobservable attributes, such as a pleasant view, nice architecture, or even historical reasons, that make them disproportionately appealing to specific types.

3.4.2 Landlord's Value Function

I model the landlord's value functions at the storefront level, and the full value function for a landlord is the sum of the value functions for all his storefronts. As a consequence, any decisions made by the landlord must account for the effects on all of his storefronts.

The equation below describes v_j^d , the value function of a storefront belonging to landlord d and leased to a type- j business. The value of a leased storefront is a function of the rent payed by the tenant, r , and a series of state variables that describe the distance of the other storefronts in the cluster, the businesses occupying each storefront, their landlords, and observable variables. These state variables are summarized by $\mathcal{C} = \{\mathcal{G}, \{L^d\}_{d \in D}, \mathcal{X}, \mathcal{J}\}$, where \mathcal{G} is the location of each storefront, \mathcal{J} is the type of business in each storefront, $\{L^d\}_{d \in D}$ is the set of storefronts belonging to developer d , and \mathcal{X} are the observable variables. Finally, l is the location of storefront v_j^d .

For convenience, I define the function $s(\cdot, \mathcal{C})$ as the function that updates \mathcal{C} to reflect a business entry or exit. For example, $s(j, l, \mathcal{C})$ adds to state \mathcal{C} a type j store in location l , and $j = 0$ adds a vacant storefront instead. Addition, I use L_o^d and L_v^d as the sets of storefronts belonging to d that are leased and vacant, respectively.

The flow value of a leased storefront is equal to the rent plus three terms. First, there is the chance of a nearby vacant storefront being leased. This event would affect the value function because it changes the business profit, and as a consequence the

expected rent when renegotiating the lease. A type j' business matched to storefront l' owned by landlord d' will lease it with probability $\mathbb{P}(\text{entry}; j', d', l', \mathcal{C})$.²¹

$$\begin{aligned}
\rho v_j^d(l, \mathcal{C}, r) = & r + \sum_{d' \in D} \sum_{l' \in L_{\mathcal{C}}^{d'}} \sum_{j'=1}^J \delta \cdot \mathbb{P}(\text{entry}; j', d', l', \mathcal{C}) \cdot \underbrace{[v_j^d(l, s(j', l', \mathcal{C}), r) - v_j^d(l, \mathcal{C}, r)]}_{\text{Business opens in } l'} \\
& + \sum_{d' \in D} \sum_{l' \in L_{\mathcal{C}}^{d'}} \lambda \cdot \mathbb{P}(\text{exit}; d', l', \mathcal{C}) \cdot \underbrace{[v_j^d(l, s(0, l', \mathcal{C}), r) - v_j^d(l, \mathcal{C}, r)]}_{\text{Business closes in } l'} \\
& + \lambda \cdot \mathbb{E}_{\varepsilon} \underbrace{\max_{0,j} \{v_0^d(l, s(0, l, \mathcal{C})) - v_j^d(l, \mathcal{C}, r), v_j^d(l, \mathcal{C}, r') - v_j^d(l, \mathcal{C}, r)\}}_{\text{Landlord renegotiates the lease in } l}
\end{aligned}$$

The second term refers to nearby businesses closing. The probability that a business in storefront l' with landlord d' exits is $\mathbb{P}(\text{exit}; d', l', \mathcal{C})$. Because business and landlord can renegotiate the rent before renewing the lease, the current rent does not influence the exit probability. By the same reasoning, rents paid by surrounding businesses do not affect their probability of exiting and thus the value of the central storefront does not change if the renegotiation is successful.

Finally, in the last term the lease is renegotiated. The lease is renewed at a new rent r' if the surplus is positive. Otherwise, the the lease is terminated, the storefront is vacated and the landlord must look for a new renter.

The flow value for vacant storefronts is pretty similar. The first two terms in the right hand side of the equation denote entries and exits in surrounding storefronts. These affect the value function through the expected profit of a potential entrant. In the last term storefront l is matched to a type j business. In case the match surplus is positive, the storefront is leased at a rent r' . Again, this term is integrated over the distribution of profit's random component ε .

²¹ The negotiation between business and landlord and the solutions for the rent and the probabilities of entry and exit are explained at length in subsection 3.4.5

$$\begin{aligned}
\rho v_0^d(l, \mathcal{C}) &= \sum_{d' \in D} \sum_{l' \in L_o^{d'}} \sum_{j'=1}^J \delta \cdot \mathbb{P}(\text{entry}; j', d', l', \mathcal{C}) \cdot \underbrace{[v_0^d(l, s(j', l', \mathcal{C})) - v_0^d(l, \mathcal{C})]}_{\text{Business opens in } l'} \\
&+ \sum_{d \in D} \sum_{l' \in L_o^d} \lambda \cdot \mathbb{P}(\text{exit}; d, l', \mathcal{C}) \cdot \underbrace{[v_0^d(l, s(0, l', \mathcal{C})) - v_0^d(l, \mathcal{C})]}_{\text{Business closes in } l'} \\
&+ \sum_{j=1}^J \delta \cdot \underbrace{\mathbb{E}_\varepsilon \max_{0,j} \{0, v_j^d(l, s(j, l, \mathcal{C}), r') - v_0^d(l, \mathcal{C})\}}_{\text{Landlord negotiates new lease for } l}
\end{aligned}$$

3.4.3 Business' Value Function

Businesses differ on their type, location, and productivity shock ε , which is re-drawn every negotiation. I assume that when a business closes (or does not enter) it cannot search for a new location, and thus receives a return of zero indefinitely. This is equivalent to assuming that the business returns to a pool with infinitely many potential entrants.

The value function below is for a type j business located in l , which is owned by landlord d and has a profit draw ε :

$$\begin{aligned}
\rho B_j(l, \mathcal{C}, r, \xi, \varepsilon) &= \Pi_j(l, \mathcal{C}, \xi, \varepsilon) - r \\
&+ \sum_{d' \in D} \sum_{l' \in L_o^{d'}} \sum_{j'=1}^J \delta \cdot \mathbb{P}(\text{entry}; j', d', l', \mathcal{C}) \cdot \underbrace{[B_j(l, s(j', l', \mathcal{C}), r, \xi) - B_j(l, \mathcal{C}, r, \xi)]}_{\text{Business opens in } l'} \\
&+ \sum_{d' \in D} \sum_{l' \in L_o^{d'}} \lambda \cdot \mathbb{P}(\text{exit}; d', l', \mathcal{C}) \cdot \underbrace{[B_j(l, s(0, l', \mathcal{C}), r, \xi) - B_j(l, \mathcal{C}, r, \xi)]}_{\text{Business closes in } l'} \\
&+ \lambda \underbrace{\mathbb{E}_\xi \max_{0,j} (-c_j, B_j(l, \mathcal{C}, r', \xi')) - B_j(l, \mathcal{C}, r, \xi)}_{\text{Business owner renegotiates lease with landlord}}
\end{aligned}$$

The first term is the flow profit net of rent. A business' profit depends on the

number of other firms and it is through profits that the entry and exit of other businesses affect the value function. The next two terms account for the entries and exits in surrounding storefronts and the last term refers to the renegotiation. As before, if the match surplus remains positive, the lease is renewed at some rent r' . Otherwise, the business leaves after paying a cost c_j .

3.4.4 Match surplus

Negotiations follow a Nash-bargaining setup. The business will lease a storefront if and only if the match has a positive surplus, and the rent is chosen so that the business keeps a fraction $1 - \beta$ of the surplus. The match surplus is the sum of the landlord and the business surplus, and for a type j entrant it is given by the equation below. The variables $j(l')$ and $r(l')$ denote the business type and rent in lot l' and are equal to 0 if the lot is vacant.

$$\begin{aligned}
S_j^d(l, \mathcal{C}, \tau) &= \underbrace{v_j^d(l, s(j, l, \mathcal{C}), r) - v_0^d(l, \mathcal{C})}_{\text{Landlord's direct surplus}} \\
&+ \underbrace{\sum_{\substack{l' \neq l \\ l' \in L^d}} [v_{j(l')}^d(l', s(j, l, \mathcal{C}), r(l')) - v_{j(l')}^d(l', \mathcal{C}, r(l'))]}_{\text{Landlord's indirect surplus}} \\
&+ \underbrace{B_j(l, s(j, l, \mathcal{C}), r, \xi, \varepsilon) - 0 - e_j}_{\text{Business' surplus}}
\end{aligned}$$

A new lease affects the landlord through two different channels. First, the landlord will receive a new flow rent r in exchange for leasing the space. I call this the entry's direct effect on the landlord's surplus. Second, the entry will affect the value of the landlord's other storefronts through spillovers. I call this the entry's indirect effect on the landlord's surplus, and it is captured in the equation by the difference in the value of his other storefronts before and after the entry.

Although the new lease does not affect the current rents paid by the other tenants, it will influence the rent on any new (or renewed) leases. As a consequence, the more storefronts a landlord owns, the more he internalizes the spillovers created by the new entrant. On the other hand, a landlord who owns a single storefront will not benefit directly from the spillovers generated by his renter, and thus will have little incentive to increase them.

The business receives zero value if it does not lease the storefront. Thus, its surplus is simply the value of renting storefront l at rent r , $B_j(l, s(j, l, \mathcal{C}), r, \xi, \varepsilon)$, minus the entry cost e_j .

The surplus of a renegotiation is only slightly different. While a new entrant must pay an entry cost, an incumbent must pay the exit cost c_j if he decides to leave. Thus, the surplus of a renegotiation is

$$\begin{aligned}
S_j^d(l, \mathcal{C}, \tau) &= v_j^d(l, \mathcal{C}, r) - v_0^d(l, s(0, l, \mathcal{C})) \\
&\quad + \sum_{\substack{l' \neq l \\ l' \in L^d}} [v_{j(l')}^d(l', \mathcal{C}, r(l')) - v_{j(l')}^d(l', s(0, l, \mathcal{C}), r(l'))] \\
&\quad + B_j(l, s(j, l, \mathcal{C}), r, \xi) + c_j
\end{aligned}$$

An important aspect of the surplus function is that it is not a function of the rent, even though the value functions are. This follows because rent does not add or subtract to the match's value. It simply redistributes surplus from the business to the landlord.

There are two main assumptions that support this result. First, both the business and the landlord have risk neutral preferences. Second, agents get to choose a new rent in every renegotiation. As a consequence, the exit probability does not depend on the current rent and the value functions of both business and landlord are linear on rents. Moreover, rent can be interpreted as a fixed rate bond with a discount

equal to the inter-temporal discount rate ρ and the arrival rate of renegotiation, λ . That is:

$$B_j(l, \mathcal{C}, r, \xi, \varepsilon) = B_j(l, \mathcal{C}, 0, \xi, \varepsilon) - \frac{r}{\rho + \lambda}$$

$$v_j^d(l, \mathcal{C}, r) = v_j^d(l, \mathcal{C}, 0) + \frac{r}{\rho + \lambda}$$

I substitute the equations above to get a surplus function without any reference to rents. Because the current rents in the landlord's other storefronts are not affected by the entry in l , they too cancel out.

$$S_j^d(l, \mathcal{C}, \xi) = v_j^d(l, s(j, l, \mathcal{C}), 0) - v_0^d(l, s(0, l, \mathcal{C}))$$

$$+ \sum_{l' \in L^d}^{l' \neq l} v_{j(l')}^d(l', s(j, l, \mathcal{C}), 0) - v_{j(l')}^d(l', s(0, l, \mathcal{C}), 0)$$

$$+ B_j(l, s(j, l, \mathcal{C}), 0, \xi) - e_j \cdot \mathbb{1}[\mathcal{C} \neq s(j, l, \mathcal{C})] + c_j \cdot \mathbb{1}[\mathcal{C} = s(j, l, \mathcal{C})]$$

3.4.5 Rent

The rent is decided by Nash-bargaining, with the landlord taking a fraction β of the total surplus. It is calculated by solving the following maximization problem of the Nash product, in which I make use of the fact that the value functions are linear on rent.

$$NB = \max_r \left[\Delta V_j^d(l, \mathcal{C}) + \frac{r}{\rho + \lambda} \right]^\beta \cdot \left[\Delta B_j(l, \mathcal{C}, \xi, \varepsilon) - \frac{r}{\rho + \lambda} \right]^{1-\beta}$$

where:

$$\begin{aligned}\Delta V_j^d(l, \mathcal{C}) &= \sum_{l' \in L^d}^{l' \neq l} v_{j(l')}^d(l', s(j, l, \mathcal{C}), 0) + v_j^d(l, s(j, l, \mathcal{C}), 0) \\ &\quad - \sum_{l' \in L^d}^{l' \neq l} v_{j(l')}^d(l', s(0, l, \mathcal{C}), 0) - v_0^d(l, s(0, l, \mathcal{C}))\end{aligned}$$

$$\Delta B_j(l, \mathcal{C}, \xi, \varepsilon) = B_j(l, s(j, l, \mathcal{C}), 0, \xi, \varepsilon) + c_j \cdot \mathbb{1}[\mathcal{C} = s(j, l, \mathcal{C})] - e_j \cdot \mathbb{1}[\mathcal{C} \neq s(j, l, \mathcal{C})]$$

I take the derivative of the log of NB and solve for r :

$$\begin{aligned}\frac{\partial \log(NB)}{\partial r} &= \frac{\beta}{\rho + \lambda} \left[\Delta V^d(l, \mathcal{C}) + \frac{r}{\rho + \lambda} \right]^{-1} \\ &\quad - \frac{1 - \beta}{\rho + \lambda} \left[\Delta B_j(l, \mathcal{C}, \xi, \varepsilon) - \frac{r}{\rho + \lambda} \right]^{-1} = 0 \\ r &= (\rho + \lambda) \{ \beta [\Delta B_j(l, \mathcal{C}, \xi, \varepsilon)] - (1 - \beta) [\Delta V_j^d(l, \mathcal{C})] \}\end{aligned}$$

The rent r is redistributing surplus from the business to the landlord. As a consequence, the rent increases if the business's surplus increases, but decreases if the landlord's surplus increases. That is, if the landlord owns more than one storefront, then the more spillovers a business produces, the lower the rent it will pay, all else equal, and the more likely it will be for the business to enter.

3.4.6 Entry and exit probabilities

I impose further structure on ε to get formulas for the entry and exit probabilities. I assume that ε is independently drawn from a standard normal distribution. The assumptions on the mean and standard deviation of ε are innocuous, since neither parameter is separately identified from other model parameters, as in most discrete choice models. The mean is not separately identified from the vectors ξ and the standard deviation is not separately identified from the profit parameters, entry and exit costs.

A business enters (or remains open) as long as the match surplus is positive. Given the distributions of ε , we have

$$\begin{aligned}\mathbb{P}(\text{entry}; j, d, l, \mathcal{C}, \xi) &= \mathbb{P}(S_j^d(l, \mathcal{C}, \xi, \varepsilon) \geq 0; j, d, l, \mathcal{C}, \xi) \\ &= \mathbb{P}(\bar{S}_j^d(l, \mathcal{C}, \xi) \geq -(\rho + \lambda)^{-1}\varepsilon; j, d, l, \mathcal{C}, \xi) \\ \mathbb{P}(\text{entry}; j, d, l, \mathcal{C}, \xi) &= \Phi((\rho + \lambda) \cdot \bar{S}_j^d(l, \mathcal{C}, \xi))\end{aligned}$$

where $\bar{S}_j^d(l, \mathcal{C}, \xi)$ is the surplus' deterministic part. ε is divided by $(\rho + \lambda)$ because the random productivity pays off until the lease is renegotiated, and the shocks re-drawn.

The exit probability is calculated in a similar fashion. Leases are terminated if, given a chance to renegotiate, landlord and business find that their match surplus is negative. Thus, the probability of exit is

$$\begin{aligned}\mathbb{P}(\text{exit}; j, d, l, \mathcal{C}, \xi) &= \mathbb{P}(S_j^d(l, \mathcal{C}, \xi, \varepsilon) < 0; j, d, l, \mathcal{C}, \xi) \\ &= \mathbb{P}(\bar{S}_j^d(l, \mathcal{C}, \xi) < -(\rho + \lambda)^{-1}\varepsilon; j, d, l, \mathcal{C}, \xi) \\ \mathbb{P}(\text{exit}; j, d, l, \mathcal{C}, \xi) &= 1 - \Phi((\rho + \lambda) \cdot \bar{S}_j^d(l, \mathcal{C}, \xi))\end{aligned}$$

where again $\bar{S}_j^d(l, \mathcal{C}, \xi)$ is the surplus' deterministic part. Note that even though the formulas are similar, the entry probability is not one minus the exit probability due to entry and exit costs.

3.4.7 Identification

In the absence of unobserved attributes, spillovers would be identified from the relationship between entry rates and the mix of surrounding businesses. If entry rates of type j businesses are higher in the presence of j' businesses then type j' must generate positive spillovers to type j .

The presence of unobserved attributes complicates the problem, and indeed poses the main challenge in identifying spillovers. I deal with this identification challenge

by using variation in spell durations due to storefronts ownership to help separate spillovers from unobserved attributes.

The key insight is that unobserved attributes shift spell durations in the same way for all storefronts, conditional on the mix of businesses surrounding it. Spillovers, on the other hand, will have different effects on duration for the same mix of businesses, depending on which of the other storefronts are owned by the landlord.

The intuition was already laid out in section 3.2, but it is worth revisiting it having in mind the model from section 3.4. If apparel shops produce positive spillovers among themselves, a vacant storefront will be more likely to be leased to an apparel shop if it is surround by other similar shops than if not, and the lease will last longer on average. If in addition those storefronts are owned by the same landlord, the effects of spillovers are amplified, since the landlord understands that each shop contributes to the other's profits, and thus to how much rent he can extract from them.²² The same mechanism applies when spillovers are negative.

Thus, while the difference in average spell durations across clusters (conditional on observables) helps identify cluster unobserved attributes, the difference in spell duration due to the mix of businesses located in jointly owned storefronts, conditional on the mix of all businesses, helps identify spillovers.

An important assumption is that unobserved attributes are homogenous within a cluster. If those attributes were more local, for instance with a similar range as the spillovers, then the estimation procedure will capture only an average of the unobserved attributes covering a cluster, and spillover estimates will still be biased. It is helpful here that clusters are relatively small (the median cluster has an area of 154,300 square feet).

²² In section 3.2 rents adjusted instantaneously in response to changes in the mix of businesses. In the model presented in section 3.4 rents only adjust following a successful renegotiation, but the expected present value of all storefronts still adjusts in response to any entries or exits, and the intuition carries through.

The remaining parameters in the model are the entry and exit costs and arrival rates. The later are analogous to setting the number of potential entrants in discrete time. Blevins (2014) argues that under a certain set of assumptions the arrival rates should be identified. Even if that is theoretically so, attempts with simulated data have proven it to be extremely difficult to estimate the correct arrival rates. Instead, I choose to calibrate λ and δ . I set $\lambda = 1$, so that leases get renegotiated on average once a year, and $\delta = 3$, which with four business types means that an empty storefront gets matched a potential entrant once a month on average.

Entry costs are identified from the average durations of vacant spells. They are not separably identified from the unobserved attributes though. For this reason, I set the unobserved attributes for the first cluster type to 0 for all business types. Finally, exit costs are not separately identified without exit spells, so they are set to 0.1, which causes businesses to stay open for 3 years on average.²³

3.5 Estimation

The model is estimated by maximizing the likelihood of the competing risk entry model, where entry by each business type denotes a different risk. I assume that each cluster belongs to one of a fixed set of cluster types, and each cluster type has a distinct vector of business type specific unobserved attributes. The unobserved attributes are intercepts in the business profit function which account for any natural local advantages. I treat the cluster's type as unknown and estimate both the unobserved attributes and the probabilities that a cluster has each cluster type. The likelihood is given by:

²³ I tried different values for the entry cost, and while the estimates for entry costs are very sensitive, the patterns observed in the spillovers are not affected.

$$L(\Theta; \{\mathcal{C}_t\}) = \prod_{i=1}^{n_E} \sum_{k=1}^K \left\{ P(\text{Cluster}_i = k) \cdot \exp \left[-\delta \sum_{t=1}^{T_i} \sum_{j=1}^J \tau_t \mathbb{P}(\text{entry}; j, d_i, l_i, \mathcal{C}_t, \xi_c) \right] \right. \\ \left. \cdot \prod_{j=1}^J [\delta \mathbb{P}(\text{entry}; j, d_i, l_i, \mathcal{C}_{T_i}, \xi_c)]^{\iota_{E,i}} \right\}$$

where n_E is the number of spells, \mathcal{C}_t are the observables in time t , K is the number of cluster types, and $\iota_{E,i}$ is 1 if the vacant storefront is rented to a type j business. The exponential term is the survival function of a vacant storefront and the next term is the product of hazard functions for each risk.

The probabilities change if \mathcal{C} changes. As a consequence, expected durations vary if surrounding businesses enter or exit. The spells are split to account for those changes, so that \mathcal{C} is constant within each sub-spell. T_i is the number of sub-spells for spell i and τ_t is the duration of sub-spell t .

There are two main challenges in estimating this likelihood. First, since it is necessary to integrate each spell over the cluster types, the survival and hazard functions are not separable in the log-likelihood anymore, making a direct optimization much more cumbersome. I use the Expectation-Maximization (EM) algorithm (Dempster et al., 1977b) to overcome this challenge. The EM algorithm is a very general tool for dealing with missing data problems. In this case, the algorithm uses the fact that, at the true Θ , the first order conditions of the original log-likelihood is identical to the first order conditions of the following function:

$$l^{EM}(\Theta; \{\mathcal{C}_t\}) = \sum_{i=1}^{n_E} \sum_{k=1}^K P(\text{Cluster}_i = k) \left\{ -\delta \sum_{t=1}^{T_i} \sum_{j=1}^J \tau_t \mathbb{P}(\text{entry}; j, d_i, l_i, \mathcal{C}_t, \xi_c) \right. \\ \left. + \sum_{j=1}^J \iota_{E,i} (\log(\delta) + \log(\mathbb{P}(\text{entry}; j, d_i, l_i, \mathcal{C}_{T_i}, \xi_c))) \right\}$$

The estimation is done iteratively, in two steps: the expectation step and the maximization step. In the expectation step, the vector $P_{m+1}(\text{Cluster}_i = k)$ is updated using Bayes' theorem, the previous probabilities $P_m(\text{Cluster}_i = k)$, and the likelihood for the current Θ_m . In the maximization step, the new Θ_{m+1} are estimated by maximizing the log-likelihood above, taking $P_{m+1}(\text{Cluster}_i = k)$ as given. The steps are repeated until convergence.

The second challenge involves computing the value functions. For each guess of Θ , a new set of value functions must be computed for businesses and landlords to calculate entry and exit probabilities. For a cluster with n storefronts, the number of potential business configurations is 5^n . Since value functions must be computed from the perspective of each storefront, and for both businesses (one value function for each business type) and landlords (one for each business type and one for vacant storefronts), the state space translates into a total of $9n \cdot 5^n$ value functions. Finally, value functions depend on the relative location of each pair of storefronts in the cluster (\mathcal{G}) and the ownership structure of storefronts ($\{\mathcal{L}\}_d$), and as a consequence they must be computed for each cluster in the sample. For a cluster with 100 storefronts, it means more than $7.0 \cdot 10^{72}$ possible states.

Instead of computing the exact value functions, I approximate them using the Sieve Value Function Iteration (SVFI) method (Arcidiacono et al., 2013).²⁴ SVFI

²⁴ See Arcidiacono et al. (2013) for a more complete treatment of the SVFI method, including proofs of convergence and convergence speed.

proposes to approximate the value function with a flexible function of the state variables. In practice, I use the first, second and third degree polynomials of the profit variables, their interactions, and dummies for cluster type, which are treated as known in the approximation procedure. That is, the value function is approximated by the function $\tilde{V} = X\Psi$, where X are the polynomial terms of the state variables and Ψ is the vector of sieve parameters.²⁵

Since the value function is not known in advance, the approximation is done iteratively. I start by choosing N points in the state space around which the value function will be approximated. For each of those points, let x_i be the vector of polynomials that will be used to approximate the value function at i , and $\{\tilde{x}_i\}$ be the set of vectors that will approximate the value functions at the right hand side of the value function at i . Finally, define X as the matrix of stacked x_i and let Ψ_k be the vector of sieve parameters at the k^{th} iteration.

After constructing the vectors x_i and $\{\tilde{x}_i\}$ and setting a starting value for Ψ_0 , the approximation repeats the following two steps until convergence. For the k^{th} iteration, in the first step I use $\{\tilde{x}_i\}$ and Ψ_k to compute the right hand side of the Bellman equation at i . This is $\tilde{V}_{k,i}$, the current guess for V_i . In the second step, I regress the vector of stacked $\tilde{V}_{k,i}$, \tilde{V}_k , on X to find the new vector Ψ_{k+1} that minimizes the distance between $X\Psi$ and \tilde{V}_k . I repeat the two steps until convergence. The SVFI returns a total of 9 vectors Ψ , for the 4 business value functions and 5 landlord value functions.

3.6 Results

In this section I present the structural estimates. I start by presenting the estimates for the model selected by the Bayesian Information Criteria, with 5 unobserved

²⁵ X is an extremely large vector, with variables that are often highly correlated. To avoid any multicollinearity problems during the approximation, I convert X into \tilde{X} , the principal components of X responsible for 99.9% of the variance in X .

cluster types. Then, I compare this model to the estimates without unobserved attributes. For convenience, I will refer to the specification with five unobserved cluster types as my main specification. The spillover estimates for both models are presented in table 3.14, and the intercepts and shares of each cluster type are given on table 3.15.

The results make it clear that the strongest spillovers happen between businesses of the same type. This is particularly true for apparel shops and other retail, which corroborates the hypothesis that search costs are important drivers of business agglomeration. Search costs are important for goods that are not purchased regularly, and this is what I find here.²⁶ Additionally, personal care is the only business type that does not create positive spillovers to similar establishments, which is also in line with the search cost theory.

Spillovers across types are important for some pairs of businesses, and can be negative. This is true both for the effect of other retail on personal care and of personal care on retail. One possible interpretation is that other retail and personal care attract types of foot traffic that are not a good match to the other. On the other hand, apparel stores and restaurants do receive positive spillovers from personal care and other retail, respectively. In particular, an additional retail store increases the profits of a restaurant by the same amount an additional restaurant would.

The estimates of spillovers and entry costs show how heterogenous are the profit function for each business type. On one extreme, there are apparel shops, which have the largest entry costs and spillovers, and so select carefully where to enter. On the other, restaurant's profits vary much less with spillovers, but restaurants face much lower entry costs. These results reflect the fact that restaurants are much more common than the other types, and seem to be found almost everywhere.

²⁶ It is worth mentioning that because other retail aggregates so many different categories, it is possible that these estimates understate the importance of spillovers for the different business types included under the umbrella of other retail.

Table 3.14: Estimates of structural parameters. The table presents estimates for the case without unobserved attributes and the main specification with five types of clusters.

	Apparel		Other retail	
	No unobs.	Main	No unobs.	Main
Apparel, 300ft	0.053 ***	0.064 ***	-0.006 ***	-0.015
Other retail, 300ft	-0.038 ***	-0.008	0.030 ***	0.049 ***
Personal care, 300ft	0.005 ***	0.037 **	-0.028 ***	-0.029 *
Restaurants, 300ft	-0.016 ***	0.002	-0.008 ***	-0.017 **
Entry cost	0.425 ***	0.475 ***	0.384 ***	0.377 ***
	Personal care		Restaurants	
	No unobs.	Main	No unobs.	Main
Apparel, 300ft	0.002	-0.011	-0.032 ***	-0.019 **
Other retail, 300ft	-0.035 ***	-0.035 ***	0.012 ***	0.024 ***
Personal care, 300ft	0.009 **	0.006	-0.013 ***	-0.020 *
Restaurants, 300ft	0.017 ***	0.004	0.021 ***	0.024 ***
Entry cost	0.431 ***	0.396 ***	0.326 ***	0.348 ***

The inclusion of unobserved attributes has a drastic impact to some coefficient estimates, but not all. Estimates of apparel's profit function are the most affected, in particular the cross-spillovers. The absence of unobserved attributes seems to lead to a downward bias in all its spillover coefficients. For restaurants, unobserved attributes seem to explain quite a bit of the previously large negative spillovers from apparel shops, while doubling spillovers from other retail.

Table 3.15: Intercepts and shares of cluster types

	Apparel	Other retail	Personal care	Restaurants	Shares
Cluster 1	0	0	0	0	0.31
Cluster 2	0.44	0.14	-0.29	-0.093	0.08
Cluster 3	-0.19	0.12	-0.23	0.36	0.4
Cluster 4	-0.065	-0.21	-0.15	0.035	0.21

Even if the unobserved attributes did not have any effects on spillover estimates, they would still be essential for any model trying to fit entry behavior of small businesses. Both Akaike Information Criteria and Bayesian Information Criteria choose five types of cluster over the model with one cluster type, that is, without unobserved attributes. Furthermore, table 3.15 shows how the estimates of cluster intercepts can be very large compared to spillovers. An apparel store in a cluster of type 1 would need seven other apparel stores within 300 feet to have the same expected profit it would in a cluster of type 2. It is not only that, but also there seems to be quite a bit of specialization across the cluster types. For example, type 2 clusters are extremely favorable to apparel stores, but very unfavorable to restaurants, while the opposite is true for type 3 clusters. As I show in section 3.7, sometimes a cluster's type will be enough to determine the business composition in the cluster.

I investigate the model's fit indirectly by first simulating ten years of matches for San Diego, and then estimating a Cox proportional hazard model on this data. The estimates from the Cox model can then be compared to the estimates from section 3.3.

I use as the starting point for the simulation the city of San Diego as of January 1st, 2002. I draw a type for each cluster using the cluster-specific probabilities calculated during the Expectation step of the EM algorithm. The estimates for the Cox proportional hazard models are presented in table 3.16.

As the results in table 3.16 show, the simulation reproduces fairly well the patterns in the reduced form estimates from table 3.10. More specifically, the estimates for businesses of the same type are very close to the estimates in the data. The exception is personal care, which the simulation has a hard time replicating.

Table 3.16: Cox Proportional Hazard model, 300 ft, simulated data. regressions on a 10 year simulation using the main specification. The numbers presented are exponential of coefficients.

	Apparel	Other retail	Personal care	Restaurants
Vacant				
All, 300 ft	1	0.97*	0.98	0.98**
Same landlord, 300 ft	1.02	1.07.	1.1.	1.09**
Apparel				
All, 300 ft	1.05***	1.04***	1.03.	1
Same landlord, 300 ft	1.55***	0.87*	0.88	0.81***
Other retail				
All, 300 ft	0.96	0.92.	1.03	1.05.
Same landlord, 300 ft	1.19*	1.58***	0.52***	1.1
Personal care				
All, 300 ft	1.01	0.89.	0.79**	0.94
Same landlord, 300 ft	1.41*	0.59*	1.45.	0.77.
Restaurants				
All, 300 ft	1.04	0.98	1.01	1.03.
Same landlord, 300 ft	1.03	1.06	1.03	1.22***

3.6.1 Spillovers versus unobservables

I assess the relative importance of unobserved attributes and spillovers by comparing three different simulations: the baseline, a case without spillovers, and a case without unobserved attributes. Again, for all simulations I start from San Diego, with the businesses open in January 1st, 2002, and the cluster types are drawn from the EM probabilities.²⁷ I simulate matches for the equivalent of 20 years to guarantee that the simulations have reached a steady state.

The first two simulations are straight forward. The baseline simulation will use the estimates from the main specification, while the case with no spillovers sets the

²⁷ I also use the same seed for each simulation, which insures that the paths will begin with the same draws. The paths will diverge though as different cases lead players to make different choices, changing the possible draws in future periods.

spillover parameters from those estimates to zero. The case without unobservables is more complicated because the model was estimated under the assumption that the unobserved attributes for the first cluster type are zero. Thus, a simulation with zero unobserved attributes is really assuming that all clusters are of type 1.

Instead, I set up this case by using the unconditional shares of cluster types, also a byproduct of the EM algorithm, to calculate the weighted average of unobserved attributes for each cluster type. Then, I take the simple average across business types of those weighted averages and entry costs, and use those values for all business and cluster types. This case should be interpreted as shutting off any competitive advantages across businesses that are not explained by observables, such as unobserved attributes and entry costs.

Table 3.17 shows the median Herfindahl-Hirschman Index (HHI) for business types, calculated at the cluster level. The HHI is a common measure of market concentration used in Industrial Organization. It varies between $1/K$ (least concentrated) and 1 (monopoly), where K is the number of firms in the market. Here, I use it to look at the concentration of business types instead. Since there are four business types, the minimum value for the HHI is 0.25.²⁸ I also add HHI estimates for a simulation without either spillovers or unobserved attributes as a benchmark of what the distribution of HHI would look like if businesses were randomly distributed across the city.

Spillovers and natural advantages are equally important in explaining the concentration of retail and service stores in San Diego. Thus, even if the inclusion of cluster types does not change spillover estimates dramatically, they are crucial to explain concentration in the city.

The similar levels of concentration generated by those two simulations hide other important dimensions in which they differ greatly. Table 3.18 presents the differences

²⁸ Vacant storefronts are excluded from the index.

Table 3.17: Business type concentration within clusters (HHI). HHI are calculated for each cluster, in four different situations: baseline, no spillovers, no unobserved attributes, and neither.

	Median
Baseline	0.57
No unobservables	0.50
No spillovers	0.52
Neither	0.44

in shares of business types and vacant storefronts for case, relative to the baseline simulation. The shares are calculated for the whole city.

Table 3.18: Difference in share of businesses across simulations, relative to baseline (p.p.)

	Vacant	Apparel	Other retail	Personal care	Restaurants
No unobservables	2.73	6.62	0.83	1.86	-12.04
No spillovers	17.54	-9.73	-1.90	0.85	-6.76

The main takeaway from this table is that spillovers are crucial to explain retail density: the absence of spillovers increases vacancy rates by 17 percentage points. Apparel’s profits are extremely reliant on spillovers, and as a consequence take the largest burden from assuming spillovers away. Restaurants also suffer some losses. On the other hand, unobserved attributes can create concentration across types, but not density. Excluding natural advantages impacts the composition of retail stores in the city, mostly by increasing the share of apparel and greatly decreasing the share of restaurants, but has little influence in the occupancy rates.

I finish the discussion of results looking at how unobserved attributes and spillovers contribute so that businesses are surrounded by neighbors that produce positive spillovers to them. Table 3.19 presents those numbers for each business type and

each simulation.

Table 3.19: Neighbors creating positive spillovers to business (%). Share of neighbors creating spillovers larger than 0.01. The denominator excludes vacant storefronts.

	Baseline	No spillovers	No unobservables
Apparel	87.7	69.7	84.9
Other	37.4	27.8	46.0
Personal care	66.7	57.1	49.1
Restaurant	82.1	76.7	73.5

As expected, the absence of spillovers leads to a strong decline in the share of positive spillover neighbors. More interestingly, three patterns emerge when comparing shares across the three simulations. For apparel stores, natural advantages have almost no impact on the share of neighbors that produce positive spillovers. The concentration is all driven by spillovers themselves. Other retail stores, on the other hand, are negatively impacted by natural advantages. Equalizing those advantages increases the share of well-matched neighbors by 10 percentage points. Finally, personal care and restaurants are actually more affected by the natural advantages than by spillovers. The equalization of natural advantages decreases the share of well-matched neighbors to those stores by more than the absence of spillovers does. Again, this is evidence that concentration of service establishments is driven more by unobserved attributes, while concentration of retail is driven by spillovers.

3.7 The effect of joint ownership on the mix of businesses

Joint ownership can play a key role in setting the mix of businesses in a cluster. This is because large storefront owners will push for entrants that maximize the value of all their storefronts, and as a consequence will produce better matches between neighboring stores. Profits then rise in response to larger spillovers.

To illustrate this idea, I chose a single cluster with 51 storefronts and simulate 150 different paths for different levels of ownership concentration. The simulation is set up so that only one landlord in the cluster owns more than one storefront. For a given concentration level, all paths have the same storefront ownership structure. I decide which storefronts are jointly owned according to the following algorithm. First, I pick the most connected storefront, that is, the storefront with the most neighboring storefronts within 300 feet. Then, I sort his neighbors according to their number of connections, and add them in sequence to the set of jointly owned storefronts until the target number of storefronts is reached or all neighbors have been selected. If the later happens, I repeat the process with their neighbors (excluding those already selected) until the target has been reached. This selection rule has the benefit that the set of jointly owned storefronts is always connected, in line with what is observed in the data. For most of this section I will present results using as unobserved attributes a weighted average of the estimates for unobserved attributes, with weights given by the share of each cluster type in the city.

Table 3.20 illustrates how joint ownership influences the average expected profits of stores in the cluster. Each rows presents the average profits and rents for a given level of joint ownership, normalized to profit and rent when there is no joint ownership. In this case, joint ownership causes a steep increase of 47% in average profits. Rents increase at a smaller pace, to a total of 11%. This result is driven by the assumptions in the model: since rents are set through Nash-bargaining, splitting the surplus in half, increases in expected profit always translate into a positive, but smaller, increase in rents.

An interesting result from table 3.20 is that most of the gains to average profits happen when joint ownership increases from 0% to 50% of the storefronts. After this point, increasing joint ownership leads to only minor increases in profits. This result

helps explain why ownership of storefronts is fragmented in the data.²⁹

Table 3.20: Average expected profits and rents for different concentrations of storefront ownership. Rows refer to shares of jointly owned storefronts. Both profits and rents are relative to their values with no jointly owned storefronts.

	Avg. Profit	Avg. Rent
0%	1	1
20%	1.087	1.012
40%	1.242	1.052
60%	1.378	1.088
80%	1.407	1.097
100%	1.471	1.112

The increase in average profits is direct result of the better match between neighboring stores. Figure 3.3 makes exactly this point. It shows how the share of neighbors that creates positive spillovers to the average store increases with joint ownership for all types of stores, except non-apparel retail, which is the rarest type of store in these simulations. The change in density of stores, on the other hand, is negligible: a decrease of 4% in the vacancy rate (2 out of 51 storefronts on average).

Finally, it is important to note that the importance of joint ownership will depend on other neighborhood characteristics. For example, some neighborhoods may be particularly well placed for certain types of retail or service establishments. For those neighborhoods, internalizing spillovers may do little to change the final composition of stores. Table 3.21 exemplifies this point by comparing changes in profits in a type 4 cluster to the baseline simulation. Joint ownership leads to a much more modest increase in average profits in a type 3 cluster, on the other of 2%. To understand why this is so, it is useful to go back to table 3.15, which shows the estimates for unobserved attributes for different cluster and business types. Clusters of type 3 are

²⁹ Indeed, 50% of storefronts in San Diego are managed independently. The other 50% are jointly owned, and split between several landlords.

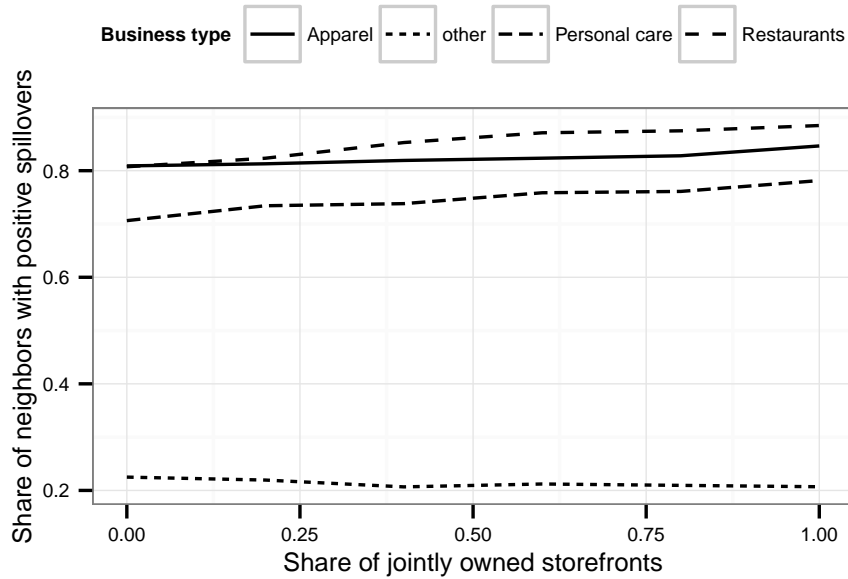


FIGURE 3.3: Shares of neighbors with positive spillovers

particularly advantageous to restaurants, which leads to a large agglomeration of restaurants, independent of storefront ownership. That is, in this case, the natural advantages create a strong focal point for the mix of retail in the neighborhood, and joint ownership has little to add to it.

Table 3.21: Average expected profits for baseline and type 3 cluster. Rows refer to shares of jointly owned storefronts, and profits are normalized by their values with no jointly owned storefronts. Type 3 cluster has strong natural advantages for restaurants.

	Baseline	Type 3
0%	1	1
20%	1.087	1.000
40%	1.242	1.004
60%	1.378	1.012
80%	1.407	1.014
100%	1.471	1.020

3.8 Subsidies and business density

It is often of the city's interest to increase business density in poorer areas. More businesses mean more jobs in the neighborhood, better amenities to residents and could translate into higher house prices. But it is not obvious that policies focused on increasing business density can have long-run implications after the policies are withdrawn. To make this point more clear, let's consider a case in which there are no spillovers, and the city implements a temporary subsidy to business entry. As entry costs go down, entry rates go up for all business types and as a consequence the equilibrium occupancy rate increases. When subsidy stops, entry rates go back to their earlier level. Exit rates are still constant, but since occupancy is much higher now than before the subsidy, total exits will exceed total entries until occupancy is back to the pre-subsidy levels.

At the same time, spillovers may create enough incentives to prevent the cluster from returning to the earlier equilibrium. As a consequence of spillovers, the same cluster may have different stable configurations, which will vary both on the mix of businesses and the occupancy rate. If the subsidy creates enough momentum in the cluster, occupancy rate may remain high even after the subsidy is over.

I study this question by simulating the effect of a 5 year subsidy to entry costs applied to a single neighborhood of San Diego. I set the subsidy to 0.3 for all businesses, which is between 60% and 100% of entry costs, depending on the business type. For simplicity, I treat the subsidy's introduction and discontinuance as unanticipated events. For comparison, I also simulate the same sequence of events, under the assumption that there is no subsidy. This second simulation will be used as a control group for the counterfactual. I repeat each simulation 150 times.³⁰

As figure 3.4 shows, the subsidies are extremely successful in increasing occupancy

³⁰ I have repeated this exercise assuming each of the five cluster types. The results are qualitatively identical, so I present the simulations with cluster type 4.

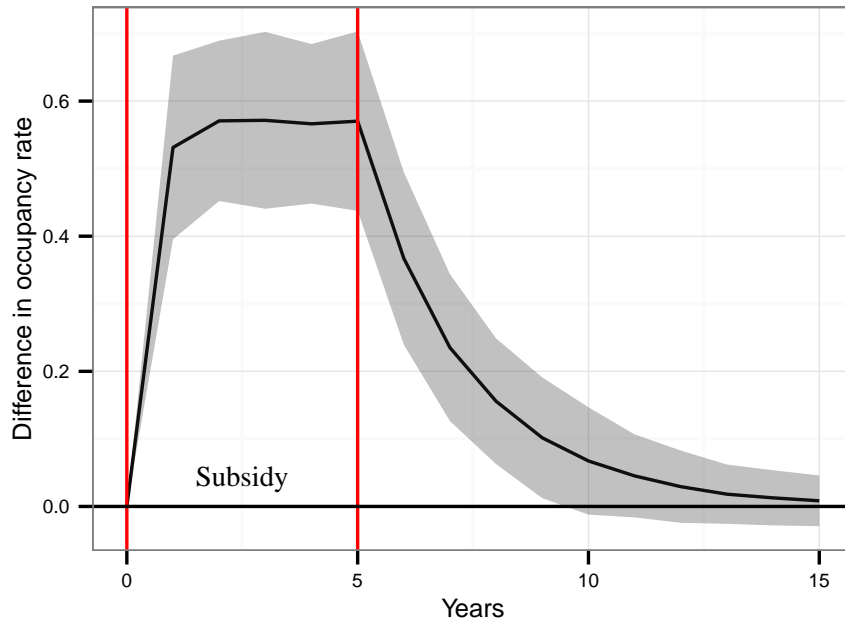


FIGURE 3.4: Average difference in occupancy rate relative to control simulation. The vertical axis is the average difference in occupancy rate between treatment and control simulations for 150 simulations.

rates at first. In two years, the subsidy increases occupancy rates on average by 60 percentage points compared to the control simulations. The initial success does not translate into long-run changes though: in 10 years all the gains in occupancy have dissipated. The result is robust to the size and length of the subsidy. Thus, it seems that subsidies to business entry by themselves are not enough to move a neighborhood to a higher occupancy equilibrium.

An advantage of a structural model is that it allows us to study how the policy responses would change with different parameters. I study two different scenarios. In the first scenario, I repeat the 150 simulations above, progressively increasing spillovers from one to four times the estimates. In the second scenario, I increase entry costs and unobserved attributes from one time to twice the estimates.

Figure 3.5 presents occupancy rates at the start of the subsidies, and 30 years

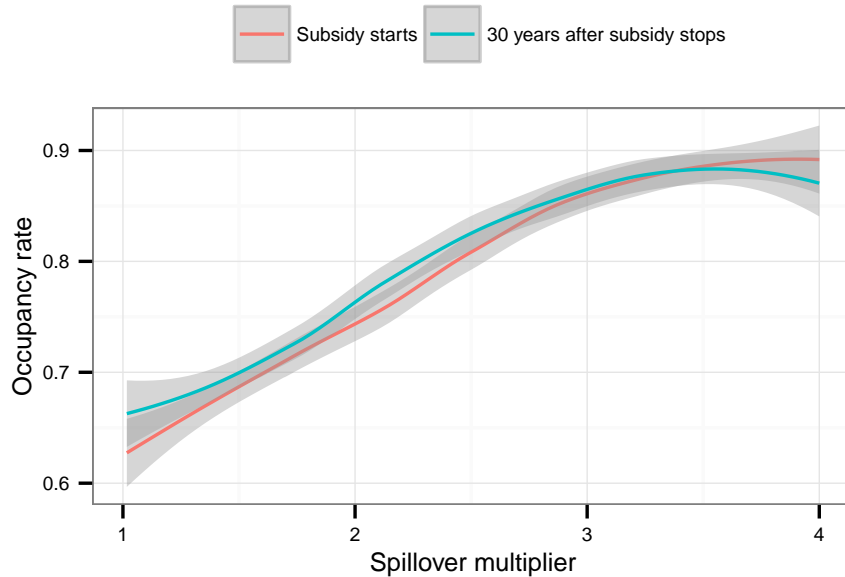


FIGURE 3.5: Occupancy rate for increasing spillovers, before and after subsidies. The vertical axis is occupancy rate and the horizontal axis is the spillover multiplier. The left picture is at the start of subsidies and the right picture is 30 years after stopping the subsidies.

after it stopped. Interestingly, increasing spillovers does not improve the long-run impact of subsidies. It does increase the equilibrium occupancy rate, but will not sustain higher rates after the subsidy STOPS. Thus, the original (lack of) results cannot be explained by weak spillovers.

Figure 3.6 provides an explanation of why previously subsidized neighborhoods do not fare better than neighborhoods that were not subsidized. Figure 3.6 shows occupancy rates on the vertical axis and the entry cost multiplier in the horizontal axis, for the moment when the subsidy started, and 30 years after it stopped. It shows that higher entry costs create the circumstances for subsidies to affect the long-run state of a neighborhood. In particular, an increase of as little as 30% in entry costs is enough for subsidies to have an impact.

The explanation for the absence of long-run effects from subsidies is not that

spillovers are not large enough, but that the entry costs estimated for San Diego are too low. As a consequence, the cost of opening a business and waiting for more stores to open is pretty low, and so there are no coordination failures for the city to solve. An increase in entry costs, on the other hand, makes it too expensive for businesses to open when there are no spillovers to support them. Thus, a situation emerges in which if there are few businesses open, no new businesses will want to come, but if there is a large enough mass of businesses, then the neighborhood will thrive at a higher occupancy rate.

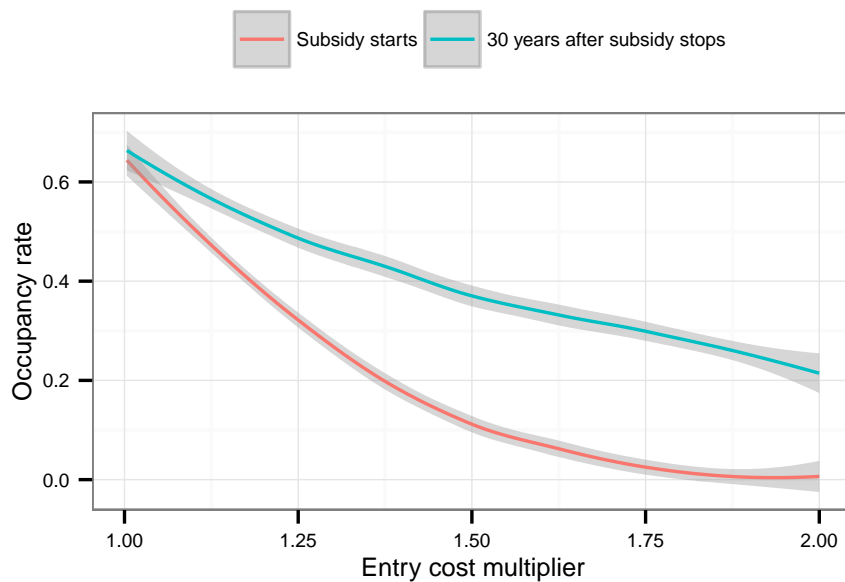


FIGURE 3.6: Occupancy rate for increasing entry costs, before and after subsidies. The vertical axis is occupancy rate and the horizontal axis is the entry cost multiplier. The left picture is at the start of subsidies and the right picture is 30 years after stopping the subsidies.

3.9 Conclusions

This paper estimates spillovers between retail and service stores and shows that they are essential to the agglomeration of small stores in San Diego. An important

challenge in estimating spillovers are unobserved local attributes. I overcome this problem by treating unobserved attributes as missing variables and estimating them. I show that landlords who own multiple storefronts internalize some of the spillovers, and so comparing entry into jointly and independently owned storefronts provide additional variation to help separate spillovers and unobserved attributes.

The existence of spillovers imply the possibility of path dependence in the mix and density of retail in a cluster. I explore this issue in two counterfactual exercises. First, I investigate the role of concentration of storefront ownership on the concentration of stores in the neighborhood. I show that expected profits increase with concentration of ownership because a large owner internalizes spillovers and as a consequence has incentives to lease to stores that maximize the total value of all storefronts. The effect is mostly driven by stores being better matched to their neighbors.

Second, I test if temporary subsidies to entry costs can cause areas with low storefront occupancy rates to switch to a high occupancy state. I find subsidized neighborhoods are not better off in the long run than they would have been without the subsidies. I show that the main reason behind this result is not the strength of spillovers, but the size of entry costs. Since estimated entry costs are low, there are enough stores willing to open for idiosyncratic reasons and wait to see if more store will follow. As a consequence, there are no coordination failures that the city can solve.

Future work on this topic could focus on two distinct directions. First, there are important questions regarding competition between neighborhoods that are not addressed in this paper. In particular, emergent neighborhoods achieve success in part at the cost of demand for retail and service stores in other neighborhoods. This general equilibrium effect should have important consequences to the cost-benefit analysis of locally targeted policies.

In addition, there is the question of the symbiotic relationship between changes

to retail and service stores in the neighborhood and changes to neighborhood demographics. Retail stores are an important part of amenities in the neighborhood, and changes to retail configuration are believed to play an important part in neighborhood gentrification, an important topic to American cities today. Better understanding the role of retail and service stores as an amenity is key to advancing our knowledge of the dynamics behind gentrification. The model developed here provides a good starting point to answer both of those questions.

A Dynamic Model of Neighborhood Segregation

4.1 Introduction

Schelling (1971) used a framework with two types of agents and simple behavioral algorithms to illustrate how even small preferences for neighborhood composition can lead to stark segregation in the city. The paper started a large literature that relies on simulations to study sorting and segregation,¹ but most of the work has focused on testing the robustness of Schelling's results to different utility functions (Pancs and Vriend, 2007; Zhang, 2004b; Laurie and Jaggi, 2003). In particular, to the best of our knowledge no work has been done to understand how agents' expectations may affect group segregation in the city. This, in spite of evidence that agents are forward looking. For example, Casey (2014) shows that the probability of a white family moving into a city block falls drastically after the first african american family moves in. Forward looking households are likely to anticipate changes to the neighborhood, and make decisions accordingly. As a consequence, under forward looking expectations neighborhoods may tip faster, as observed in Casey (2014), and

¹ For examples, see Laurie and Jaggi (2003); Wasserman and Yohe (2001); Zhang (2004a,b, 2011); Pancs and Vriend (2007).

overall segregation in the city will also be affected.

In this paper, we expand on Schelling (1971) by modeling households as forward looking agents. That is, each household will form beliefs about how his neighborhood and other neighborhoods of the city may evolve over time. As a consequence, households will have distinct preferences over houses with identical neighborhood composition if one of those neighborhoods is more likely to change than the other.

The main challenge in working with such models is that the Bellman equations describing household's preferences do not have a closed form solution. Solving for the Bellman equations numerically is also out of question due to the curse of dimensionality: a city with 10,000 houses and 4 types of households would require solving a problem with $4 \cdot 10^4 \cdot 5^{1e4}$ unknowns. Instead, we show how the Sieve Value Function Iteration method (Arcidiacono et al., 2013) can be used to approximate the Bellman equations.

We use our model to simulate how expectations affect segregation. We use 4 types of agents in our simulations, $\{+1, +2, -1, -2\}$, defined by their group, $+$ or $-$, and their preferences, which can be segregationist (1) or integrationist (2). We find that segregation increases when households have forward looking expectations, compared to the myopic case. Those households with segregationist preferences have on average 2.5 percentage points more neighbors of their own group, and are better off for that. On the other hand, households with integrationist preferences are worse off than in the myopic world, since they must live with less mixing than they would like.

Forward looking expectations are likely to affect how households respond to different trends observed in American cities in the recent past, such as gentrification spurred by the movement of young highly educated whites back to the city (Couture and Handbury, 2016), the large migration of hispanic families to certain cities and neighborhoods (Rubn Hernandez-Len, 2000), or the out-migration of - mostly high

earners - from cities like Detroit (Hartley, 2013).

Our model provides a good starting point to model those questions within the Schelling (1971) framework. For now, we provide in section 4.4 suggestions on how our model can be modified to address those different questions. Future work should focus on understanding how households react to those trends, and what is the role played by expectations in shaping their response.

The paper is structured as follows. We present the model in section 4.2, explain how we use the Sieve Value Function Iteration method to approximate the Bellman equations, and discuss how forward looking expectations may affect household choices. The simulations are presented in section 4.3. We present three different extensions of our model, all with practical implications, in section 4.4. Finally, we conclude with section 4.5.

4.2 Model

Our model builds on the basic framework of Schelling (1971), and developed by Laurie and Jaggi (2003); Wasserman and Yohe (2001); Zhang (2004a,b, 2011); Páncs and Vriend (2007). The city is populated by agents belonging to different groups, that have preferences over their neighborhood's composition. We refer to those groups generically as $+$ and $-$, and in different contexts they may be interpreted as the agents' skin color, wealth, or ethnicity. Each agent occupies a different house in the city, with some of the houses left vacant.

We follow the literature and treat the houses as being squared cells, organized as a checkerboard. We assume that the cells at the city's southern edge border those at its northern edge, and the ones at its eastern edge border those at its western edge, forming a donut shaped geometric figure known as a torus. Thus, each cell has a total of 8 immediate neighbors. The picture below exemplifies a city with a random allocation of agents:

This city has a total of 10,000 houses, distributed across 100 rows and 100 columns. The five different colors mark the cell's occupant. White cells are vacant, blue cells are occupied by $-$ households, and red cells by $+$ households. Households have different types of utility over their neighborhood composition, thus the five colors. In particular, some households will have strict preference for segregation (lighter cells), while others will have strict preference for integration (darker cells). We refer to those households by the numbers 1 and 2, respectively. The preferences combine with the households' original groups to form 4 household types, $+1$, $+2$, -1 , and -2 , with utility functions

$$\begin{aligned}
u^{+1}(\mathbf{N}^l) &= \delta_1 \cdot \frac{\sum_{n_i^l \in \mathbf{N}^l} \mathbf{1}[n_i^l > 0] - \mathbf{1}[n_i^l < 0]}{8} \\
u^{-1}(\mathbf{N}^l) &= \delta_1 \cdot \frac{\sum_{n_i^l \in \mathbf{N}^l} \mathbf{1}[n_i^l < 0] - \mathbf{1}[n_i^l > 0]}{8} \\
u^{+2}(\mathbf{N}^l) = u^{-2}(\mathbf{N}^l) &= \delta_2 \cdot \left(1 - \frac{\left| \sum_{n_i^l \in \mathbf{N}^l} \mathbf{1}[n_i^l > 0] - \sum_{n_i^l \in \mathbf{N}^l} \mathbf{1}[n_i^l < 0] \right|}{8} \right)
\end{aligned}$$

where l is the household's location, $\mathbf{N}^l = \{n_1^l, \dots, n_8^l\}$ are the types of the 8 households surrounding l (vacant cells are marked as 0), and δ_1 and δ_2 are preference parameters.

Time is discreet. Every period, one household in the city is randomly matched to a vacant cell. The household will move to the new cell if the cell offers it a larger value than his current location. The value of a cell is determined by the current composition of its surrounding cells, plus the household's expectations over how the composition will change over time. That is, households account for the possible future paths the neighborhood may take.

This is a key difference between our approach and that of the previous literature. Traditionally, papers have modeled agents as myopic decision makers, their valuation

of a cell depending only on the current, immediate neighbors of that cell. This in spite of Schelling (1971) having already highlighted the importance of the forward looking aspect of decision making in the real estate market.² In particular, changes to neighborhood composition are often predictable, and signals indicating tipping or gentrification will help agents coordinate their actions.³⁴

The value of a cell is described by the following Bellman equation:

$$V_l^{k_l}(\mathbb{N}) = u^{k_l}(\mathbf{N}_l) + \beta \left\{ \sum_{v \in \mathbf{v}} \frac{1}{P \cdot V} \left[\mathbb{P}_{lv}^{k_l}(\mathbb{N}) \cdot V_v^{k_l}(s(l, v, \mathbb{N})) + (1 - \mathbb{P}_{lv}^{k_l}(\mathbb{N})) \cdot V_l^{k_l}(\mathbb{N}) \right] + \sum_{\substack{i \neq l \\ i \in \mathbf{l}}} \sum_{v \in \mathbf{v}} \frac{1}{P \cdot V} \left[\mathbb{P}_{iv}^{k_i}(\mathbb{N}) \cdot V_l^{k_i}(s(i, v, \mathbb{N})) + (1 - \mathbb{P}_{iv}^{k_i}(\mathbb{N})) \cdot V_l^{k_i}(\mathbb{N}) \right] \right\}$$

where P and V are the number of households and vacant cells in the city, and \mathbb{N} is a vector of length $P + V$ matching each cell in the city to its occupant's type, or 0 for vacant cells. The function $s(i, v, \mathbb{N})$ represents the new vector obtained from swapping entries i and v in the original vector \mathbb{N} . k_i is the type of the agent occupying cell i , \mathbf{l} and \mathbf{v} are vectors of entries in \mathbb{N} that are different from and equal to 0, that is, \llcorner and \simeq are the location of all occupied and vacant cells in the city, respectively.

We follow the discrete choice literature and assume that households receive a preference shock $\varepsilon \sim \mathbb{N}(0, \sigma)$ when deciding to move.⁵ As a consequence, the household

² See Bayer et al. (2011) for a comprehensive discussion of the importance of modeling forward looking agents in the housing market.

³ Casey (2014) shows that a single african american family moving into a block decreases the probability that a white family will move in in the future.

⁴ Wasserman and Yohe (2001); Laurie and Jaggi (2003) have explored the idea that agents have preferences over the composition of cells beyond their immediate surroundings and have shown that this assumption can lead to different dynamics of neighborhood segregation. Those papers interpret the preference over farther cells as being the agent's vision field, but it could be equally interpreted as a mechanism to capture forward looking expectations.

⁵ Zhang (2004a,b, 2011) also assume a random component to households' preferences.

will move from i to v with probability

$$\begin{aligned} \mathbb{P}_{iv}^{k_i}(\mathbb{N}) &= \mathbb{P}(V_v^{k_i}(s(i, v, \mathbb{N})) + \varepsilon > V_i^{k_i}(\mathbb{N})) \\ &= \Phi\left(\frac{V_v^{k_i}(s(i, v, \mathbb{N})) - V_i^{k_i}(\mathbb{N})}{\sigma}\right) \end{aligned}$$

4.2.1 Value function approximation

The main challenge posed by the forward looking model is that the value functions are not easily computed. The functions do not have a closed form solution. For a fixed city size (number of rows and columns), the function has to be evaluated at a finite number of points. Thus, at least in theory the function could be solved numerically. Unfortunately, the scale of the problem makes a numerical solution infeasible. For a city with 10,000 houses, the numerical computation of the value function for a specific cell would require solving a problem with $10^4 \cdot 5^{1e4}$ unknowns. The value function could be computed for a fixed share of household types and vacant houses, which would decrease the number of unknowns considerably, but still not enough to make it computationally feasible.

Alternatively, we propose approximating the value function using the Sieve Value Function Approximation (SVFI) method. The method is described at length in Arcidiacono et al. (2013), so we give only a brief introduction to the SVFI here. The key insight of the SVFI method is that a value function can be approximated by flexible functions of its state variables, similarly to how a function can be approximated by an Euler expansion.

Thus, for a previously chosen set of flexible functions of state variables $f(X)$, where X are the state variables, and $f(\cdot)$ is the vector returned by the flexible functions of X , the value function $V(x)$ will be approximated by $V(x) \approx \Theta f(x)$.

If $V(X)$ was known for enough values of X , the problem could be solved by a simple Ordinary Least Square regression. Unfortunately, the whole challenge is that $V(X)$ is unknown to us. Arcidiacono et al. (2013) propose to solve this problem recursively. Starting with an initial guess Θ_0 , use the right-hand side of the value function equation to construct a guess of $V(X)$ for multiple values of X , called $V_0(X)$. Then, regress $V_0(X)$ on $f(X)$ to obtain a new guess Θ_1 . The process is repeated until Θ_m and Θ_{m+1} are sufficiently close.

A key factor for the quality of the approximation is the choice of both the function $f(\cdot)$ and the state variables X . For example, we could approximate the value function using only the composition of the immediate neighborhood. The downside of this approximation is that it disregards any moves between cells that are not in within 1 cell of the household. For a household with 8 occupied cells in his neighborhood, this means ignoring the actions of 99.9% of the households in the city. Alternatively, we could choose a very complete set of state variables, but focus only on linear functions of those variables. This approximation would ignore non-linearities in the value function. For example, it would treat being 30 cells away from the edge of an undesirable cluster as being three times better than being 10 cells away, when it is possible that both locations had similarly stable neighborhoods.

Our choice of $f(X)$ were made with the goal to minimize those concerns. We approximate the value function of types +1 and -1 households with share of neighbors by type within the first and second rings around the cell, the distance to the closest + and the closest - clusters, and quadratic functions of those variables. The value function of +2 and -2 households is approximated by those same variables, plus the modulus of the difference in the shares of + and - households in each of the two closest rings.

Finally, while the SVFI method considerably decreases computational requirements, even its implementation is challenging, due to the large summation over the

possible choices for all other households in the city. This summation has a total of $(P - 1) \cdot V$ elements, and its computation is extremely cumbersome within the SVFI's iterative algorithm. Thus, we approximate the summation by drawing a total of 1,000 potential moves without replacement. Moves to and from the household's surroundings have a larger impact in the value function, and we account for this by oversampling those moves. In particular, we draw one move from each occupied cell within the two rings closest to the household, plus a move by an agent of each type to any vacant cells within those two rings (the weights are calculated accordingly).

4.2.2 The consequences of forward looking behavior

Forward looking agents make their decisions keeping in mind the future of their neighborhood and the whole city. Before moving to the simulations, we discuss intuitively here how the forward looking behavior will alter agent's evaluation of a cell and its neighborhood. We focus on agents of type +1 and -1, whose utility functions are closest to the literature.

First, it is important to note that agent's preferences are a function only of their immediate neighborhood. That is, on a daily basis, none of the agents care about the identity of neighbors two or more cells away. Segregation preferences are extremely local. Yet, forward looking behavior will ensure that agents look beyond their immediate neighborhood when making decisions. That is because while distant cells will not affect agents directly, they will help agents predict how their neighborhood will evolve over time. For example, for a +1 agent, a mixed neighborhood close to a cluster of + cells will be more valuable than a similar neighborhood by a cluster of - cells, since the chances that the mixed neighborhood will change to a + neighborhood in the future are much higher in the first scenario. A similar argument can be made for a cell well within a cluster of similar agents versus a cell closer to its edge.

The second cell is more at risk to see its neighbors' types changing in the future.⁶

Second, forward looking expectations will also influence households' evaluation of their own immediate neighborhood. This is because households know that their neighbors could move out any time, and be replaced by someone of a different type. Thus, while a household's utility on any single day remains the same whether the household is forward looking or not, the expected value of his immediate neighborhood 5 or 10 years from then will be very different for a forward looking household. For better or for worse, he knows many of his neighbors will move out eventually.

Finally, a forward looking agent also knows that he too will have opportunities to move. As a consequence, he may be more tolerant about the composition of his neighborhood right now, anticipating the possibility of moving to a more segregated neighborhood in the future.

In summary, forward looking households differ from myopic households because they value their non-immediate surroundings (first mechanism), and because they discount the importance of their present immediate neighborhood (second and third mechanisms). The three mechanisms will affect a household's tolerance for diversity in ways that are often contradictory. Thus, it is not obvious how the household's foresight will affect neighborhood segregation in the city.

4.3 Simulations

In this section, we present the results from the simulations. We start by showing how segregation emerges when there are forward looking expectations, and how different household types decide where to live. Then, we compare those results with the myopic case, and show that there is less segregation under myopic than under forward

⁶ A similar mechanism has been explored in Laurie and Jaggi (2003), which models households with preferences over both immediate and more distant neighbors. The authors interpret such preferences as the household's field of vision, but one can think of them also as an strategy to simulate a forward looking behavior directly in household's preferences.

looking expectations.

We set δ , the preference parameter, and σ , the standard deviation of the preference shock, both to one. We treat every period as a different day, and set the discount rate to 0.9999, which implies a yearly discount rate of 0.964. We use a city with 10,000 houses (cells), organized into 100 rows and 100 columns, and four household types. The initial distribution of residents and vacancies within the city is randomly generated by drawing for each cell, with replacement, a household type from the set $\{-2, -1, 0, 1, 2\}$, with probabilities $\{.1, .375, .05, .375, .1\}$. Figure 4.2 illustrates the starting point for the simulation. Finally, we simulate 100,000 iterations, which for a population of around 9,500 translates into an average of 11 chances to move per household. Henceforth, we refer to a sequence of 9,500 iterations as a cycle.

As we see in figure 4.3, after the first cycle there is already quite a bit of segregation, but at this point similar households are grouping into several, relatively small clusters. After 5 cycles, a clear pattern emerges. As shown in figure 4.4, the small clusters merged into larger ones. The cluster mergers intensify as the simulation continues. By the 10th cycle, clusters are large, and their borders are much smoother.

Interestingly, there are almost no integrated areas, even though 20% of households would prefer an integrated neighborhood. Instead, those households locate mostly at the borders of perfectly segregated clusters, while inner cells are occupied by types +1 and -1. It is possible that this pattern is due to the much larger share of type 1 households in the city (75% of the cells, versus 20% for type 2). As it is, type 2 households insulate + and - clusters from each other, but it is not clear if they increase the stability of the clusters by creating a buffer between them, or if they decrease it by having type 2 households work as trojan horses.

Finally, we would like to compare simulations with myopic and forward looking

expectations. Figures 4.6 and 4.7 show how the utility and the share of same-sign neighbors evolves for the average type 1 and type 2 households. We focus first on the comparisons for type 1 households. It is clear that as the city segregates, those households are made better off. But it seems they are able to segregate more, and faster, with forward looking expectations. It is an intuitive result: if households can predict how a neighborhood will change, they can take measures to protect themselves, and find a neighborhood that will give him a greater welfare in the future.

Type 2 households go through a more interesting dynamic. They do pretty well at the start of the simulation, since the random sampling ensures to many of them an integrated neighborhood. But as the other households start to segregate, they are made worse off. There are few vacant cells in integrated areas at first, and it takes them several cycles to find a house that provides them at least as much utility as the initial random sample. As we saw earlier, those houses are the ones between different clusters.

Unlike type 1 households, type 2 are better if all households are myopic. This is exactly because type 1 are better at segregating with forward looking expectations. Since 1 form a large majority of residents, they impose the segregation on those households that would prefer to live in integrated areas. As a consequence, while in the myopic case households of type 2 are eventually better off than they were when the simulation started, in the forward looking case they do no better than the random sampling.

Thus, the simulations show that adding forward looking expectations to Schelling's original model has the power of strengthening even more segregation, reinforcing the original findings in Schelling (1971).

4.4 Model extensions

The model introduced in section 4.2 is a good starting point to illustrate how forward looking expectations can be added to the setup of Schelling (1971), and section 4.3 showed that when households anticipate the future, they will be more successful in segregating themselves. In addition to these results, we believe that our model provides a good framework to address other questions that have been ignored by this literature so far. In particular, in this section we show how our model can be extended to address three different questions. First, how changes in the city's socio-economic profile affect its neighborhoods. Second, how households respond to a large decrease in population, best exemplified by Detroit. Third, how households respond to large, unanticipated changes in amenities in a neighborhood, such as a new magnet school or the demolition of a housing project.

4.4.1 Changes in demographics

It is not uncommon for cities to undergo rapid changes in the composition of their population. As examples, we can list the gentrification of inner cities following the movement of wealthy young residents from the suburbs to downtown (Couture and Handbury, 2016), or the fast increase in the share of hispanic population in several American cities Rubn Hernandez-Len (2000). Guerrieri et al. (2013) has discussed the consequences of such changes in a linear model with myopic agents. Our model allows us to study those changes in a bi-dimensional city with forward looking agents.

We include demographic changes to our model by assuming probabilities that a household will be substituted by a household of a different group. That is, a $-$ household will move out and a $+$ household will move in with probability q^+ . Alternatively, a $+$ household will switch with a $-$ household with probability q^- . Every period, either some household gets a chance to move as before, or a household

is forced to switch with a household of a different group. The revised value function is:

$$\begin{aligned}
V_l^{k_l}(\mathbb{N}) = & u^{k_l}(\mathbf{N}_l) + \beta \left\{ \left(1 - q^+ \sum_{i \in \mathbb{I}} \mathbb{1}[k_i < 0] + q^- \sum_{i \in \mathbb{I}} \mathbb{1}[k_i > 0] \right) \right. \\
& \left[\sum_{v \in \mathbf{v}} \frac{1}{P \cdot V} \left[\mathbb{P}_{lv}^{k_l}(\mathbb{N}) \cdot V_v^{k_l}(s(l, v, \mathbb{N})) + (1 - \mathbb{P}_{lv}^{k_l}(\mathbb{N})) \cdot V_l^{k_l}(\mathbb{N}) \right] \right. \\
& \left. + \sum_{\substack{i \neq l \\ i \in \mathbb{I}}} \sum_{v \in \mathbf{v}} \frac{1}{P \cdot V} \left[\mathbb{P}_{iv}^{k_i}(\mathbb{N}) \cdot V_l^{k_l}(s(i, v, \mathbb{N})) + (1 - \mathbb{P}_{iv}^{k_i}(\mathbb{N})) \cdot V_l^{k_l}(\mathbb{N}) \right] \right] \\
& \left. + q^+ \sum_{i \in \mathbb{I}} \mathbb{1}[k_i < 0] \cdot V_l^{k_l}(s(i, +, \mathbb{N})) + q^- \sum_{i \in \mathbb{I}} \mathbb{1}[k_i > 0] \cdot V_l^{k_l}(s(i, -, \mathbb{N})) \right\}
\end{aligned}$$

where the first parenthesis within the curly brackets is the probability that the city composition will remain the same, and $s(i, +, \mathbb{N})$ is a function that sets the group of the household in cell i to $+$.

To understand how this setup creates demographic change, first we should note that q^+ and q^- determine a steady state proportion of $+$ and $-$ groups. In particular, to have a constant ratio of $+$ to $-$ households in the city, it is required that the probability of some household from group $+$ leaving the city be the same as the probability for some household from group $-$ to leave it. That is, $q^- N^+ = q^+ N^-$, where N^+ and N^- are the number of households from groups $+$ and $-$ in steady state. If the actual number of $+$ residents in the city is larger than N^+ , the probability of one of them emigrating is larger than the probability of a new $+$ household moving to the city, and the number of $+$ households will be expected to decrease. Thus, by choosing the right values for the initial share of household types in the city and emigration probabilities q , we can simulate the effect of demographic change, whether it is caused by changes in the share of high earners in the city, or different ethnic or

racial groups.

4.4.2 Changes in population density

Several American cities have seen a rapid decrease in population density in the post-war, as jobs disappear and residents move elsewhere looking for better opportunities. This is true for many cities in the Rust belt, with Detroit being the most well known example (Hartley, 2013). Such changes in density can lead to large re-sorting within the city, as the housing price declines and vacancies increase.

Our model can accommodate the out-migration by treating it as an exogenous process. That is, we assume that the probability that a household will leave the city does not depend on how segregated the city or his neighborhood are. We follow a similar strategy as before, and denote by q^m the probability that a vacant cell will be filled by a migrant, increasing the city population by 1. For simplicity, we assume that the new resident's type is draw from a discrete uniform distribution. A current resident has a probability q^x of leaving the city each period. Without loss of generality, we normalize the utility from leaving the city to 0. The value function is now:

$$\begin{aligned}
V_l^{k_l}(\mathbb{N}) &= u^{k_l}(\mathbb{N}_l) + \beta \left\{ (1 - q^x \cdot P + q^e \cdot V) \cdot \right. \\
&\left[\sum_{v \in \mathbf{v}} \frac{1}{P \cdot V} \left[\mathbb{P}_{lv}^{k_l}(\mathbb{N}) \cdot V_v^{k_l}(s(l, v, \mathbb{N})) + (1 - \mathbb{P}_{lv}^{k_l}(\mathbb{N})) \cdot V_l^{k_l}(\mathbb{N}) \right] \right. \\
&\left. + \sum_{\substack{i \neq l \\ i \in \mathbf{1}}} \sum_{v \in \mathbf{v}} \frac{1}{P \cdot V} \left[\mathbb{P}_{iv}^{k_l}(\mathbb{N}) \cdot V_l^{k_l}(s(i, v, \mathbb{N})) + (1 - \mathbb{P}_{iv}^{k_l}(\mathbb{N})) \cdot V_l^{k_l}(\mathbb{N}) \right] \right] \\
&\left. + q^e \sum_{v \in \mathbf{v}} \sum_{k \in \{1, 2, -1, -2\}} V_l^{k_l}(s(v, k, \mathbb{N})) + q^x \sum_{i \in \mathbf{1}}^{i \neq l} V_l^{k_l}(s(i, 0, \mathbb{N})) \right\}
\end{aligned}$$

where again the first parenthesis within the curly brackets is the probability that

the city composition will remain the same, and $s(i, k, \mathbb{N})$ is a function that sets the resident of cell i to k , with 0 indicating a vacant cell.

This model creates changes to population density much in the same way that the earlier model created changes in demographics. For given q^e and q^x , there is a ratio of population to vacancy $\star P / \star V$ such that $\star P \cdot q^x = \star V \cdot q^e$. $\star P$ determines the long-run population density in the city (remember that the city size is fixed). Thus, if the simulation starts with a population $P > \star P$, then the population will gradually converge to $\star P$. The speed of converge is determined by the absolute values of q^e and q^x .

4.4.3 Changes in amenities

There are instances in which amenities in a neighborhood can change rapidly and dramatically: a large increase in policing, the opening of a new metro station or magnet school, the demolishing of housing projects, and so on. In those circumstances, the changes to the neighborhood may induce a change in the demand for houses in the area and shift the composition of households in the medium to long run in several city neighborhoods.

Our model can be adapted to account for such changes in two ways: by treating them as unanticipated, one-off events, or by treating them as random events with a known probability. In both instances, forward looking behavior by the households could lead to considerably different responses to changes, when compared to the myopic case. We start by illustrating the unanticipated case, and then build on it to create a model with anticipated changes to amenities.

We add amenities to our model by assuming that the city can be divided in areas with different amenities. For example, for a 100 by 100 city, we could model each section of 5 by 5 cells as having the same amenity. Thus, cells in rows 1 through 5 and columns 1 through 5 would have the same amenity, and cells in rows

6 through 10 and columns 1 through 5 would have a different amenity. Amenities affect households' utility directly, and different types of households may value the amenities differently. Let ω be a matrix matching each cell to its amenity values for type + and - households, respectively. That is, ω_{i1} is the utility from amenities for a household + in cell i . Then, the utility functions can be rewritten as:

$$\begin{aligned}\hat{u}^{+1}(\mathbf{N}^1, \omega) &= \omega_{i,1} + \delta_1 \cdot \frac{\sum_{n_i^l \in \ell \mathbf{N}^1} \mathbf{1}[n_i^l > 0] - \mathbf{1}[n_i^l < 0]}{8} \\ \hat{u}^{-1}(\mathbf{N}^1, \omega) &= \omega_{i,2} + \delta_1 \cdot \frac{\sum_{n_i^l \in \ell \mathbf{N}^1} \mathbf{1}[n_i^l < 0] - \mathbf{1}[n_i^l > 0]}{8} \\ \hat{u}^{+2}(\mathbf{N}^1, \omega) &= \omega_{i,1} + \delta_2 \cdot \left(1 - \frac{|\sum_{n_i^l \in \ell \mathbf{N}^1} \mathbf{1}[n_i^l > 0] - \sum_{n_i^l \in \ell \mathbf{N}^1} \mathbf{1}[n_i^l < 0]|}{8} \right) \\ \hat{u}^{-2}(\mathbf{N}^1, \omega) &= \omega_{i,2} + \delta_2 \cdot \left(1 - \frac{|\sum_{n_i^l \in \ell \mathbf{N}^1} \mathbf{1}[n_i^l > 0] - \sum_{n_i^l \in \ell \mathbf{N}^1} \mathbf{1}[n_i^l < 0]|}{8} \right)\end{aligned}$$

An unanticipated change in amenities can be simulated by changing the value of ω (under the assumption that the households observe the change). The value function itself remains unchanged, except for including the matrix ω as a state variable.

We add anticipated changes first by assuming that there is a limited number of types of amenities. For simplicity, we present the case where there are only two amenity types here, say a and b . Every period, a neighborhood may transition from a to b with probability w_{ab} , and from b to a with probability w_{ba} . Finally, we define W as the number of sections the city has been split in, and ω as a function matching each section in W to an amenity type a or b . Again, we use $s_2(i, a, \omega)$ as meaning that we should set the i^{th} area in ω to a . Then, the value function becomes:

$$\begin{aligned}
V_l^{k_l}(\mathbb{N}, \omega) &= u^{k_l}(\mathbf{N}_l, \omega) + \beta \left\{ \left(1 - w_{ab} \sum_{i \in W} \mathbb{1}[\omega(i) = a] + w_{ba} \sum_{i \in W} \mathbb{1}[\omega(i) = b] \right) \cdot \right. \\
&\quad \left[\sum_{v \in \mathbf{v}} \frac{1}{P \cdot V} \left[\mathbb{P}_{lv}^{k_l}(\mathbb{N}, \omega) \cdot V_v^{k_l}(s(l, v, \mathbb{N}), \omega) + (1 - \mathbb{P}_{lv}^{k_l}(\mathbb{N}, \omega)) \cdot V_l^{k_l}(\mathbb{N}, \omega) \right] \right. \\
&\quad \left. + \sum_{\substack{i \neq l \\ i \in \mathbf{1} v \in \mathbf{v}}} \frac{1}{P \cdot V} \left[\mathbb{P}_{iv}^{k_i}(\mathbb{N}, \omega) \cdot V_l^{k_l}(s(i, v, \mathbb{N}), \omega) + (1 - \mathbb{P}_{iv}^{k_i}(\mathbb{N}, \omega)) \cdot V_l^{k_l}(\mathbb{N}, \omega) \right] \right] \\
&\quad \left. + w_{ab} \sum_{i \in W} \mathbb{1}[\omega(i) = a] \cdot V_l^{k_l}(\mathbb{N}, s_2(i, b, \omega)) + w_{ba} \sum_{i \in W} \mathbb{1}[\omega(i) = b] \cdot V_l^{k_l}(\mathbb{N}, s_2(i, a, \omega)) \right\}
\end{aligned}$$

4.5 Conclusion

It has been 4 decades since Schelling's seminal paper first used simulations to draw insights on how households' preferences can lead to segregation within cities. His main finding, which has been confirmed by the subsequent literature, is that even small preferences for neighborhood composition can lead to large segregation in the city.

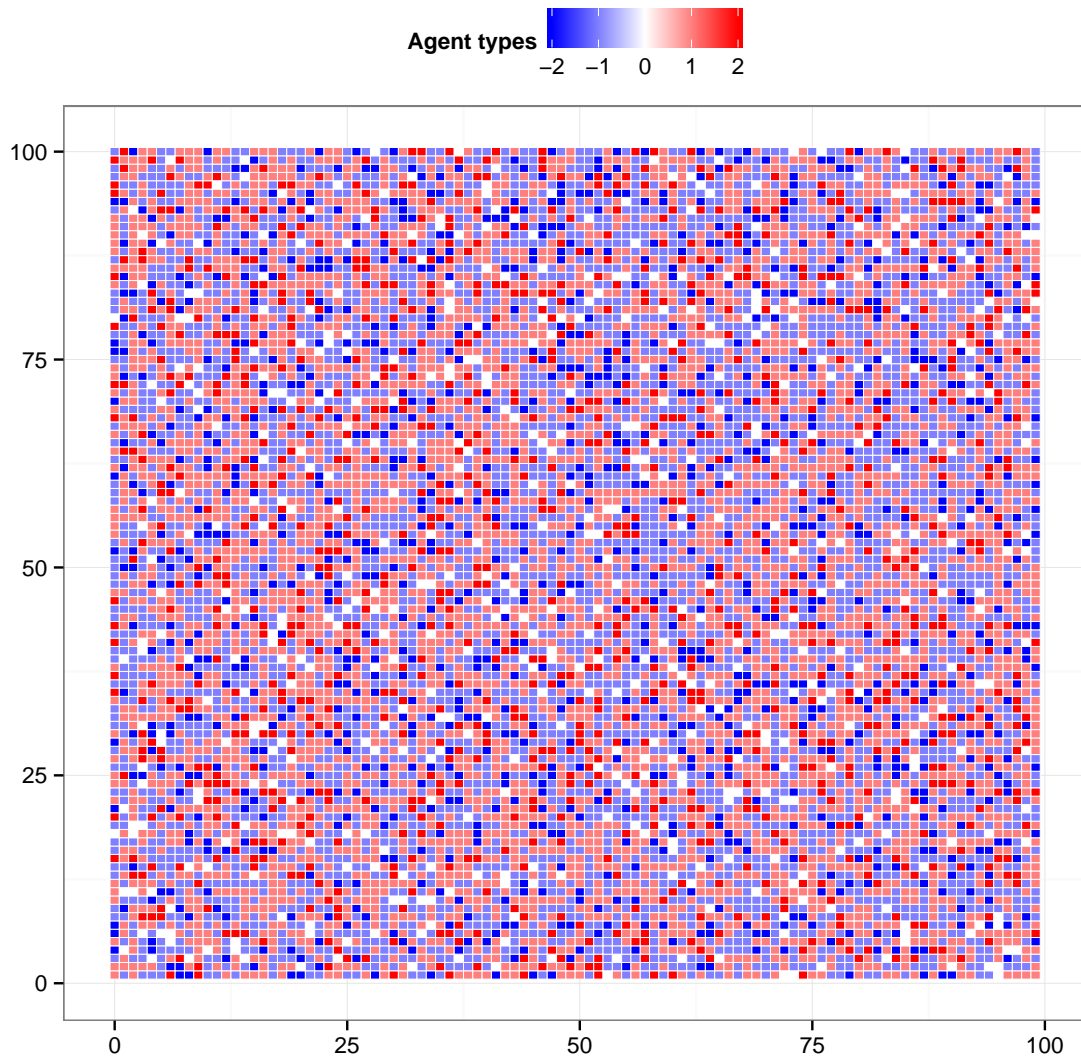
In this paper, we showed for the first time how to model households with forward looking expectations within Schelling's framework. Computing the exact Bellman equations is not feasible, so we approximated them using the Sieve Value Function Iteration method. Then, we used this framework to compare levels of segregation when using myopic and forward looking households.

We found that households are more able to segregate if they have forward looking expectations. As a consequence, households who prefer similar neighbors are better off, while households who prefer integrated neighborhoods would prefer a myopic world. Also, integrated neighborhoods are pretty rare in our simulation. Instead, households who prefer an integrated neighborhood seem to locate at the border of segregated neighborhoods, working as a buffer between the other households.

Future work should focus on testing the results' robustness to different initial conditions and parameters. It will be interesting to explore how the breakdown of residents across different preference types (segregation versus integration) affects segregation in the city.

We also showed how minor modifications can extend our model to address several policy relevant questions. In particular, our model can be used to study the effects of in-migration of different groups, whether they are wealthier tech workers in San Francisco, or hispanic immigrants in Texas, population decline in cities in the Rust belt, and large changes to amenities, such as the construction of a new magnet school or the demolition of a housing project. Future versions of this paper will look into those questions, and into understanding how those changes affect sorting within the city.

FIGURE 4.1: Example of a city with a random allocation of agents. White cells are vacant. Southern and northern borders are connected, as are eastern and western borders.



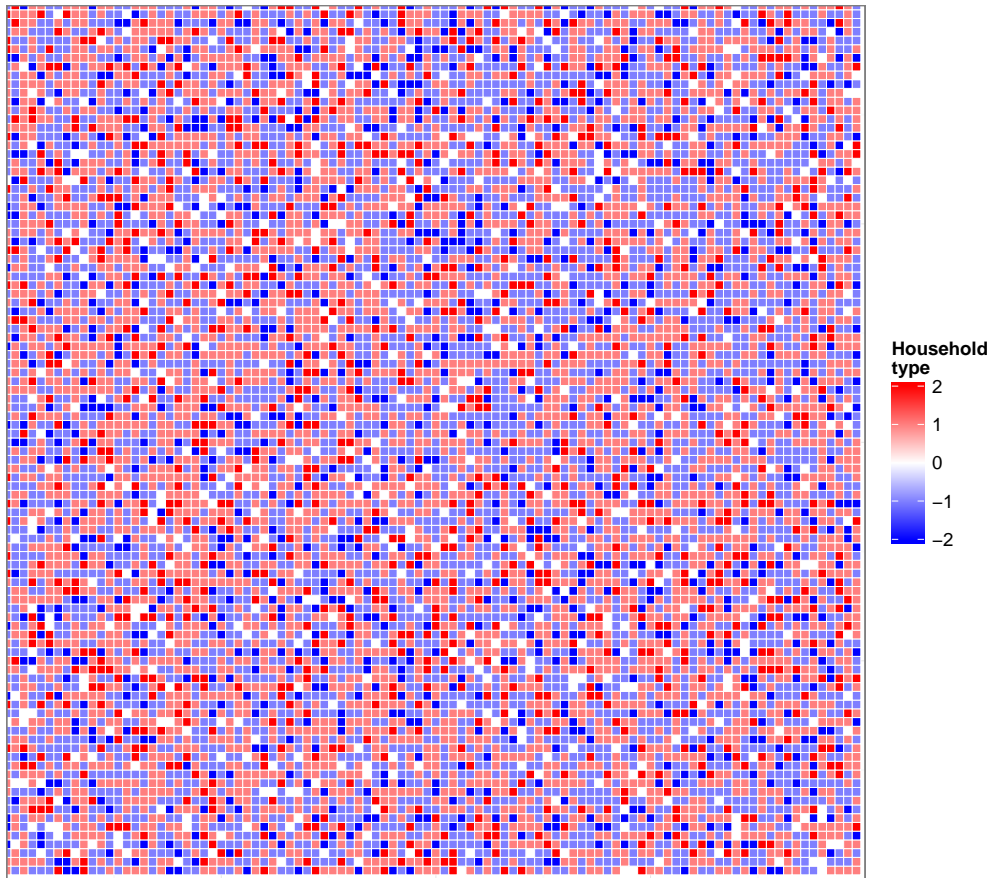


FIGURE 4.2: City at the start of the simulation

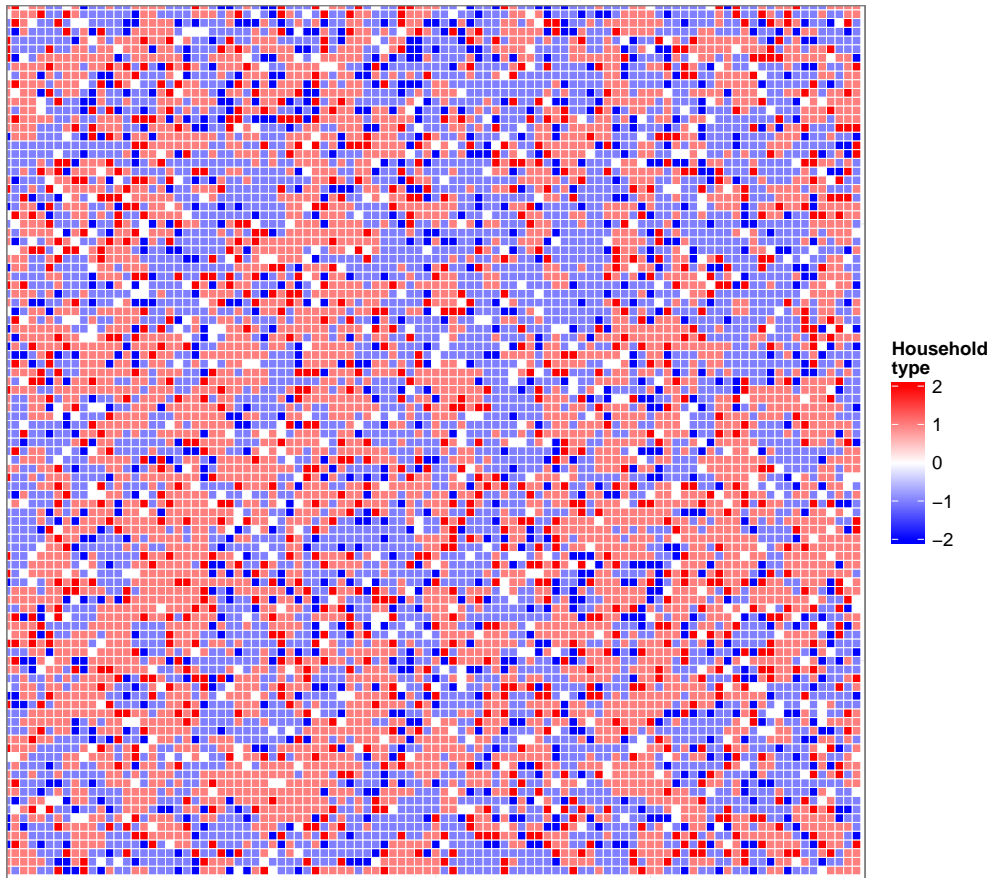


FIGURE 4.3: City after 1 cycle. A cycle is a sequence of 9,500 iterations.

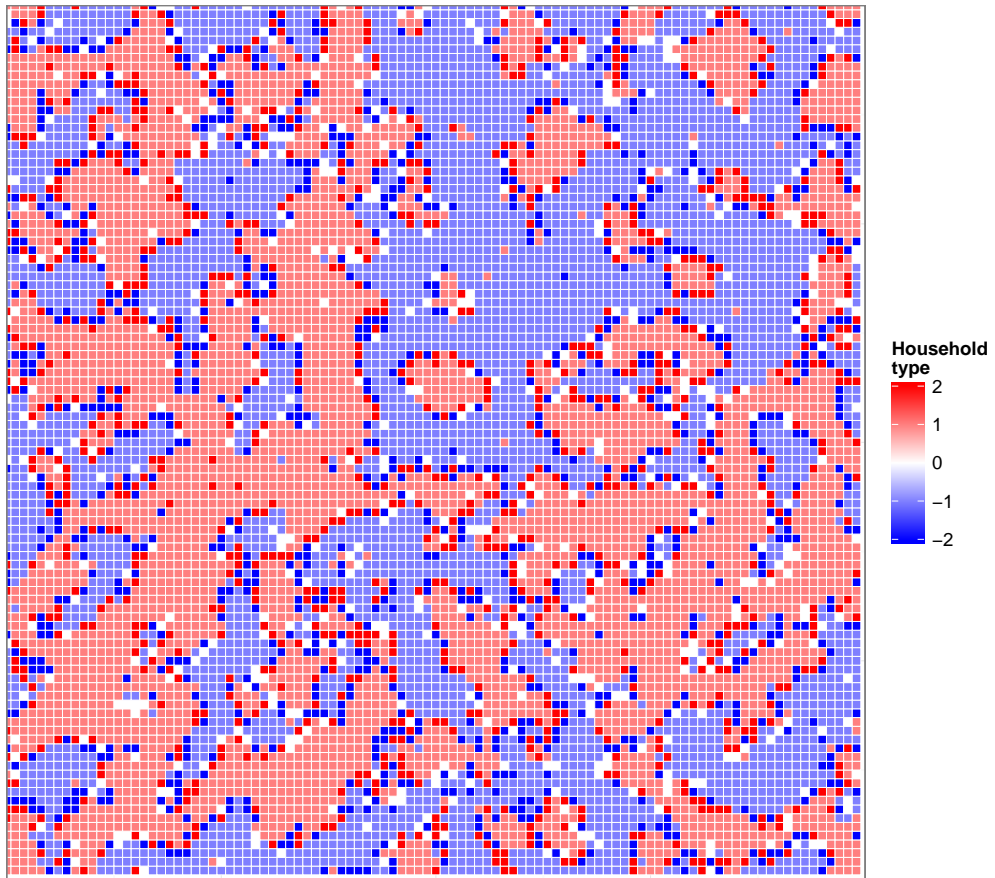


FIGURE 4.4: City after 5 cycles. A cycle is a sequence of 9,500 iterations.

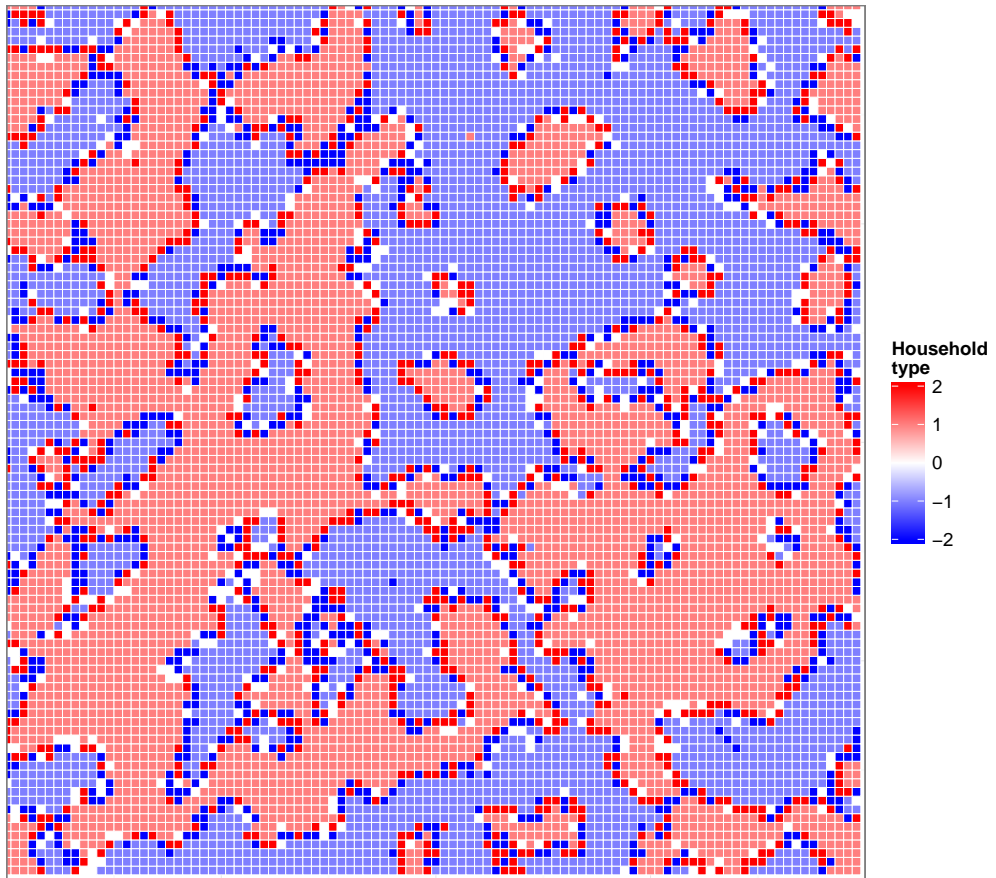
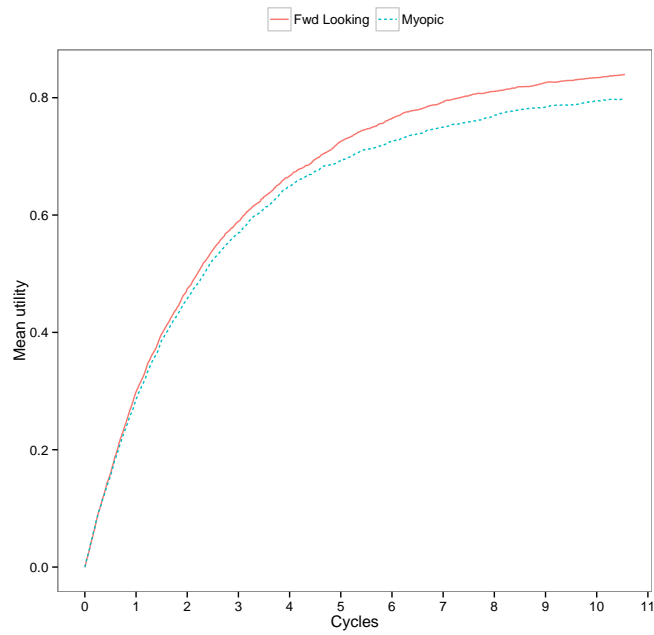
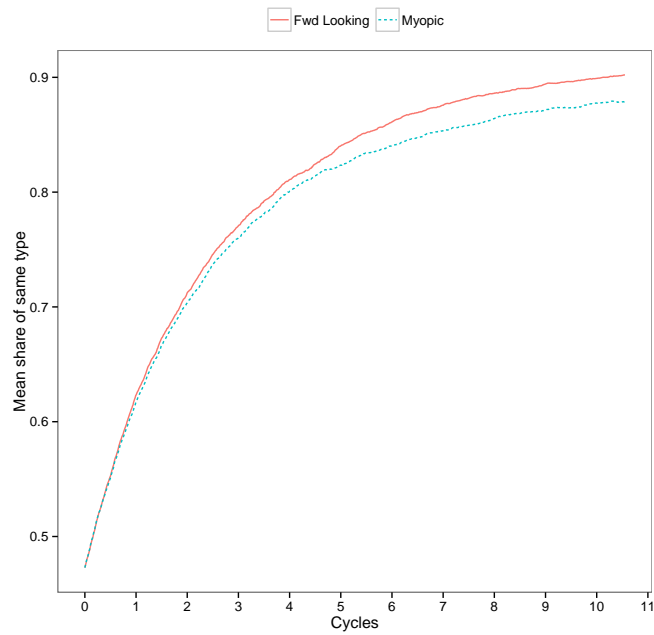


FIGURE 4.5: City after 10 cycles. A cycle is a sequence of 9,500 iterations.

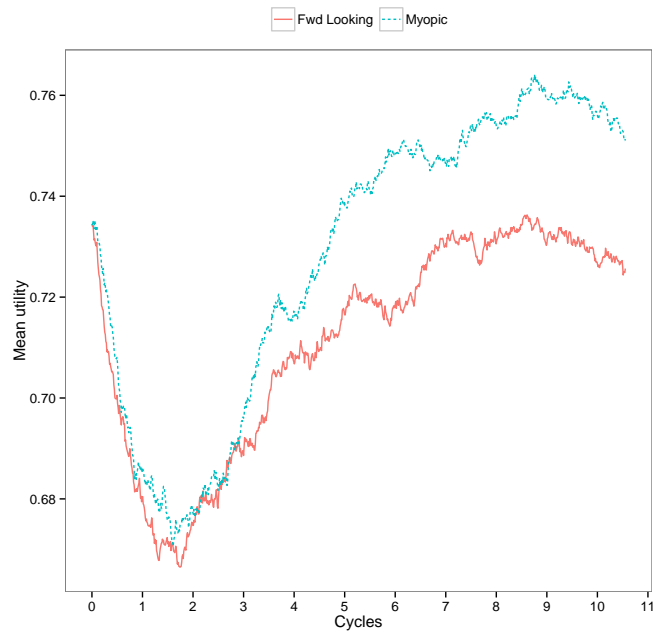


(a) Utility

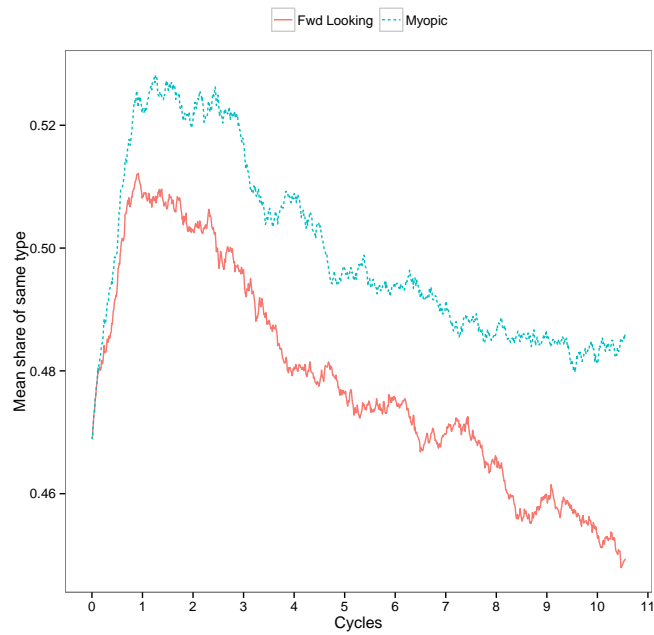


(b) Share of similar neighbors

FIGURE 4.6: Evolution of utility and share of same-sign neighbors for the average +1 and -1 households during simulation



(a) Utility



(b) Share of similar neighbors

FIGURE 4.7: Evolution of utility and share of same-sign neighbors for the average +2 and -2 households during simulation

Bibliography

- Abowd, J. M., Haltiwanger, J., and Lane, J. (2004), “Integrated longitudinal employer-employee data for the United States,” *American Economic Review*, pp. 224–229.
- Arbel, Y., Ben-Shahar, D., Gabriel, S., and Tobol, Y. (2010), “The local cost of terror: Effects of the second Palestinian Intifada on Jerusalem house prices,” *Regional Science and Urban Economics*, 40, 415–426.
- Arcidiacono, P., Bayer, P., Blevins, J. R., and Ellickson, P. B. (2012), “Estimation of dynamic discrete choice models in continuous time with an application to retail competition,” Tech. rep., National Bureau of Economic Research.
- Arcidiacono, P., Bayer, P., Bugni, F. A., and James, J. (2013), “Approximating High-dimensional Dynamic Models: Sieve Value Function Iteration,” in *Structural Econometric Models*, eds. E. Choo and M. Shum, chap. 7, pp. 45–95, Emerald Group Publishing Limited.
- Bartolucci, F., Farcomeni, A., and Pennoni, F. (2013), *Latent Markov models for longitudinal data*, Chapman & Hall/CRC Taylor and Francis Group.
- Baum, L. E., Petrie, T., Soules, G., and Weiss, N. (1970), “A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains,” *The annals of mathematical statistics*, 41, 164–171.
- Bayer, P., McMillan, R., Murphy, A., and Timmins, C. (2011), “A dynamic model of demand for houses and neighborhoods,” .
- Benmelech, E., Bergman, N., Milanez, A., and Mukharlyamov, V. (2014), “The Agglomeration of Bankruptcy,” Tech. rep., National Bureau of Economic Research.
- Bishop, K. C. and Murphy, A. D. (2011), “Estimating the willingness to pay to avoid violent crime: A dynamic approach,” *The American Economic Review*, 101, 625–629.
- Blevins, J. R. (2014), “Identification and estimation of continuous time dynamic games,” *The Ohio State University Working Paper*.

- Bresnahan, T. F. and Reiss, P. C. (1991), “Entry and competition in concentrated markets,” *Journal of Political Economy*, pp. 977–1009.
- Brooks, L. and Strange, W. C. (2011), “The micro-empirics of collective action: The case of business improvement districts,” *Journal of Public Economics*, 95, 1358–1372.
- Casey, M. (2014), “The Microdynamics of Neighborhood Transition,” .
- City of Buffalo, T. (2003), “Queen City Hub: A Regional Action Plan for Downtown Buffalo,” Tech. rep.
- Cosman, J. (2014), “Industry dynamics and the value of variety in nightlife: evidence from Chicago,” *Working Paper*.
- Couture, V. and Handbury, J. (2016), “Urban Revival in America, 2000 to 2010,” .
- Davis, P. (2006), “Spatial competition in retail markets: movie theaters,” *The RAND Journal of Economics*, 37, 964–982.
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977a), “Maximum likelihood from incomplete data via the EM algorithm,” *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 1–38.
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977b), “Maximum likelihood from incomplete data via the EM algorithm,” *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 1–38.
- Dudey, M. (1990), “Competition by Choice: The Effect of Consumer Search on Firm Location Decisions,” *The American Economic Review*, 80, pp. 1092–1104.
- Dunne, T., Klimek, S. D., Roberts, M. J., and Xu, D. Y. (2013), “Entry, exit, and the determinants of market structure,” *The RAND Journal of Economics*, 44, 462–487.
- Gibbons, S. (2004), “The Costs of Urban Property Crime*,” *The Economic Journal*, 114, F441–F463.
- Glaeser, E. (2011), *Triumph of the city: How our greatest invention makes US richer, smarter, greener, healthier and happier*, Pan Macmillan.
- Glaeser, E. L. and Gottlieb, J. D. (2008), “The economics of place-making policies,” Tech. rep., National Bureau of Economic Research.
- Gould, E. D., Pashigian, B. P., and Prendergast, C. J. (2005), “Contracts, externalities, and incentives in shopping malls,” *Review of Economics and Statistics*, 87, 411–422.

- Guerrieri, V., Hartley, D., and Hurst, E. (2013), “Endogenous gentrification and housing price dynamics,” *Journal of Public Economics*, 100, 45–60.
- Hartley, D. (2013), “Urban Decline in Rust-Belt Cities,” .
- Jacobs, J. (1961), *The death and life of great American cities*, Vintage.
- Jia, P. (2008), “What Happens When Wal-Mart Comes to Town: An Empirical Analysis of the Discount Retailing Industry,” *Econometrica*, 76, 1263–1316.
- Kline, P. (2010), “Place based policies, heterogeneity, and agglomeration,” *The American Economic Review: Papers and Proceedings*, pp. 383–387.
- Kline, P. and Moretti, E. (2014), “Local Economic Development, Agglomeration Economies, and the Big Push: 100 Years of Evidence from the Tennessee Valley Authority*.” *Quarterly Journal of Economics*, 129.
- Laurie, A. J. and Jaggi, N. K. (2003), “Role of ‘vision’ in neighbourhood racial segregation: a variant of the Schelling segregation model,” *Urban Studies*, 40, 2687–2704.
- Leggett, C. G. and Bockstael, N. E. (2000), “Evidence of the effects of water quality on residential land prices,” *Journal of Environmental Economics and Management*, 39, 121–144.
- Linden, L. and Rockoff, J. E. (2008), “Estimates of the Impact of Crime Risk on Property Values from Megan’s Laws,” *American Economic Review*, 98, 1103–27.
- Lynch, A. K. and Rasmussen, D. W. (2001), “Measuring the impact of crime on house prices,” *Applied Economics*, 33, 1981–1989.
- Pancs, R. and Vriend, N. J. (2007), “Schelling’s spatial proximity model of segregation revisited,” *Journal of Public Economics*, 91, 1–24.
- Pope, D. G. and Pope, J. C. (2012), “Crime and property values: Evidence from the 1990s crime drop,” *Regional Science and Urban Economics*, 42, 177–188.
- Pope, J. C. (2008), “Fear of crime and housing prices: Household reactions to sex offender registries,” *Journal of Urban Economics*, 64, 601 – 614.
- Rubn Hernandez-Len, V. Z. (2000), “‘Making Carpet by the Mile’: The Emergence of a Mexican Immigrant Community in an Industrial Region of the U.S. Historic South,” *Social Science Quarterly*, 81, 49–66.
- Schelling, T. C. (1971), “Dynamic models of segregation,” *Journal of mathematical sociology*, 1, 143–186.
- Seim, K. (2006), “An empirical model of firm entry with endogenous product-type choices,” *The RAND Journal of Economics*, 37, 619–640.

- Shoag, D. and Veuger, S. (2014), “Shops and the City: Evidence on Local Externalities and Local Government Policy from Big Box Bankruptcies,” *HKS Working Paper No. RWP14-019*.
- Skogan, W. G. (1999), “Measuring what matters: crime, disorder, and fear,” in *Measuring What Matters: Proceedings From the Policing Research Institute Meetings*, ed. R. H. Langworthy, pp. 37–54, U.S. Department of Justice, Washington, DC.
- Thaler, R. (1978), “A note on the value of crime control: Evidence from the property market,” *Journal of Urban Economics*, 5, 137 – 145.
- Tita, G., Petras, T., and Greenbaum, R. (2006), “Crime and Residential Choice: A Neighborhood Level Analysis of the Impact of Crime on Housing Prices,” *Journal of Quantitative Criminology*, 22, 299–317.
- Vitorino, M. A. (2012), “Empirical entry games with complementarities: An application to the shopping center industry,” *Journal of Marketing Research*, 49, 175–191.
- Wasserman, H. and Yohe, G. (2001), “Segregation and the provision of spatially defined local public goods,” *The American Economist*, pp. 13–24.
- Wolinsky, A. (1983), “Retail Trade Concentration due to Consumers’ Imperfect Information,” *The Bell Journal of Economics*, 14, pp. 275–282.
- Zhang, J. (2004a), “A dynamic model of residential segregation,” *Journal of Mathematical Sociology*, 28, 147–170.
- Zhang, J. (2004b), “Residential segregation in an all-integrationist world,” *Journal of Economic Behavior & Organization*, 54, 533–550.
- Zhang, J. (2011), “Tipping and Residential Segregation: A Unified Schelling Model*,” *Journal of Regional Science*, 51, 167–193.
- Zucchini, W. and MacDonald, I. L. (2009), *Hidden Markov models for time series: an introduction using R*, CRC Press.

Biography

Eduardo Ferreira Jardim was born on October 15, 1985, in Ribeirão Preto, São Paulo state, Brazil, where he spent all his formative years. He is the son of Manoel Ferreira Jardim Filho and Deocélia Bassoteli Jardim, and he has one sister, Débora Ferreira Jardim.

After graduating from high school at COC Sistema de Ensino, in Ribeirão Preto, SP in December 2003, he became part of the first class of Economics at the São Paulo School of Economics, at the Getúlio Vargas Foundation (FGV-EESP). During his time in college, he focused on the study of the history of economic thought, writing a thesis on the relationship between the American pragmatic philosophy and Economics. During this time he also started a degree in Philosophy at the University of São Paulo (USP), where he studied at night. Unfortunately, Hegel did not allow him to conclude his studies.

He entered the MA program in Economics at the University of São Paulo in January of 2007, after placing 6th in the national entrance exam. At USP, he widened his knowledge of History of Economic Thought and Macroeconomics. In his Master thesis, he built a theoretical model to argue how financial constraints by middle and low income families could create the rise in consumption of durable goods observed during price freezes in Brazil and in other countries. It was at USP that he was exposed for the first time to cutting edge research in Economics, and it was there that he decided to apply to PhD programs abroad.

In July 2010, Eduardo enrolled at Duke University in Durham, NC, to pursue a doctoral degree in Economics. During his time at Duke, he was very pleased to work under the supervision of his advisor, Patrick Bayer, on topics related to the importance of dynamic models in Urban Economics. His research included understanding how agents price the cost of crime in a neighborhood, when crime rates are changing over time; how retail and service stores who benefit from spillovers between themselves agglomerate within the city; and how households with preferences over their neighborhood composition sort in the city if they have rational expectations about how the neighborhoods will change. This thesis is the combination of those different efforts. After graduation, Eduardo will be taking a position of Economist at Amazon, in Seattle.