

THE UNIVERSITY OF CHICAGO

THE EFFECTS OF CREDIT SUPPLY SHOCKS AND NEIGHBORHOOD SPILLOVERS ON  
HOUSING INVESTMENT: EVIDENCE FROM CHICAGO

A DISSERTATION SUBMITTED TO  
THE FACULTY OF THE DIVISION OF THE SOCIAL SCIENCES  
IN CANDIDACY FOR THE DEGREE OF  
DOCTOR OF PHILOSOPHY

KENNETH C. GRIFFIN DEPARTMENT OF ECONOMICS

BY  
MERU BHANOT

CHICAGO, ILLINOIS

JUNE 2019

## DEDICATION

To my family and Mariah Farbo for their unwavering love and support.

## ACKNOWLEDGMENTS

I would like to thank my thesis committee comprising Jonathan Dingel (chair), Amir Sufi, and Steven Levitt for their support and feedback throughout the process. I would also like to thank colleagues and faculty at the University of Chicago who provided feedback at various points in the research process, including Kyle Kost, John Grigsby, Felix Tintelnot, Luc Anselin, Stephane Bonhomme, Julia Kochinsky, Rodrigo Adao, Chad Syverson, Joseph Vavra, Eric Zwick, and Alexander Torgovitsky. Thanks as well to my mentors at the IMF, Lev Ratnovski and Divya Kirti, who provided excellent feedback and guidance on the HMDA data and robustness checks. Additional thanks to the teams at NHGIS, the Chicago Open Data Portal, the National Archives (for HMDA data), BuildFax, and ZTRAX for access to their data and support staff, and their willingness to make incredible data broadly available to researchers and the public.

## ABSTRACT

This paper studies the effect of credit supply in the home mortgage market on households' investment in housing quality and the resulting effect on neighborhood development. I construct shocks to refinancing credit supply and credit standards, and demonstrate using difference-in-differences and instrumental variables approaches with Chicago administrative data that negative shocks depress housing investment and spillover to nearby tracts. I then present a dynamic model that rationalizes these results through investment incentives of profit-maximizing and credit-constrained building managers catering to consumers whose housing demand depends on house and neighborhood quality. The model best fits the data when households place equal weight on house and neighborhood quality - suggesting an important role for neighborhood effects in investment incentives - and matches growth rates from 2010-2016 in model-implied housing quality with a correlation of 0.53, and in annual rental prices with a correlation of 0.48. I use the model to simulate the impact of Chicago's selected "opportunity zones" on city-wide house prices, and construct counterfactual opportunity zones that have a larger impact both on the city as a whole and on distressed tracts in the city.

# CONTENTS

<b>List of Figures</b>	<b>ix</b>
<b>List of Tables</b>	<b>xi</b>
<b>1. Introduction</b>	<b>1</b>
1.1. Related Literature and Policy . . . . .	8
1.2. Outline . . . . .	13
<b>2. Data description</b>	<b>14</b>
2.1. Housing Investment Data: The Chicago Open Data Portal . . . . .	14
2.2. Housing Credit Data: The Home Mortgage Disclosure Act . . . . .	19
2.2.1. Data Overview . . . . .	19
2.2.2. Data Summary . . . . .	19
<b>3. Constructing and Verifying Credit Shocks</b>	<b>22</b>
3.1. The Credit Supply Shock . . . . .	22
3.1.1. The Credit Supply Shock: Construction . . . . .	22
3.1.2. The Credit Supply Shock: Estimation . . . . .	25
3.1.3. The Credit Supply Shock: Validation . . . . .	32
3.2. The Credit Standards Shock . . . . .	36
3.3. Application to Chicago . . . . .	45
<b>4. Empirical Results</b>	<b>47</b>
4.1. Notation and Econometrics . . . . .	47
4.2. Threats to Identification . . . . .	52
4.3. Effects of Credit Supply and Standards within Tracts . . . . .	53
4.4. Spillovers in Investment . . . . .	57

<b>5. Model</b>	<b>65</b>
5.1. Setup	65
5.1.1. Workers' Housing Demand	67
5.1.2. Building Mangers	69
5.1.3. Solving the Bellman	72
5.1.4. Credit Constraints	73
5.2. Functional Form Assumptions	75
5.2.1. Imposing Functional Forms	76
5.2.2. Imposing Credit Constraints	81
<b>6. Model Estimation and Calibration</b>	<b>84</b>
6.1. Calibration	85
6.2. Estimation Steps	95
6.3. Simulation	96
<b>7. Model Applications</b>	<b>101</b>
7.1. Chicago Opportunity Zones	102
7.1.1. Background Information and Model Relationship	102
7.1.2. Baseline Results	105
7.1.3. Selecting Opportunity Zones: "Random" and "Single Tract" approaches	108
<b>8. Conclusion</b>	<b>117</b>
<b>References</b>	<b>118</b>
<b>A. Data Appendix</b>	<b>126</b>
A.1. Chicago Open Data Portal	126
A.1.1. Building Permits	126
A.1.2. Other Chicago Data	127
A.2. HMDA Details	130

A.3. Top 10 Entities in Chicago - Market Share by Tract in 2006 . . . . .	133
A.4. Other Data Sources . . . . .	137
A.4.1. BuildFax . . . . .	137
A.4.2. Zillow ZTRAX . . . . .	137
<b>B. Entity-year FE Details</b>	<b>139</b>
<b>C. Additional Specifications</b>	<b>141</b>
C.1. Reduced Form Specifications . . . . .	142
C.2. Individual Instruments . . . . .	144
<b>D. Math Appendix</b>	<b>145</b>
D.1. Workers' Housing Demand . . . . .	145
D.2. Building Manager's Problem Details . . . . .	146
D.3. Proof: Symmetric Steady State is Unique in Long-Run without constraints . . . . .	147
D.4. Cobb-Douglas Case . . . . .	147
D.4.1. Comparative Statics . . . . .	148
<b>E. Model Calibration Details</b>	<b>151</b>
E.1. ACS Data Correction . . . . .	151
E.2. ZTRAX on Land and Assessment Values . . . . .	152
<b>F. Model Estimation Details</b>	<b>154</b>
F.1. Estimation Equations . . . . .	154
<b>G. Additional Simulations</b>	<b>157</b>
G.1. Property Tax or Mortgage Rate changes . . . . .	157
G.2. Menu or Investment Cost changes . . . . .	160
G.3. Building Amenities: The Obama Library . . . . .	163

\*

## LIST OF FIGURES

1.1. Median Year Structure Built - Chicago . . . . .	2
1.2. National Permit Values by Type . . . . .	3
1.3. Cook County Permit Values by Type . . . . .	3
2.1. Housing Investment per HH-Year . . . . .	16
2.2. Housing Investment per HH-Year in Crisis and Recovery . . . . .	18
2.3. Housing Investment Residuals in Crisis and Recovery . . . . .	18
2.4. Originated Home Purchase and Refinancing Loans - Chicago . . . . .	20
2.5. Median Household Income 2006-2010 ACS - Chicago . . . . .	21
3.1. Level of Geographic Dispersion of Lenders . . . . .	24
3.2. Relative Explanatory Power of Entity and County FEs Annually . . . . .	30
3.3. County-Level Crisis Supply and Demand Shocks . . . . .	33
3.4. Cook County Number of Loan Applications and Rejections . . . . .	38
3.5. Cook County Denials by Primary Reason . . . . .	39
3.6. Percentage of Lending in 2010 done by 2004-2006 Chicago Lenders . . . . .	44
3.7. Refinancing Credit Shock Residuals 2007-2010 . . . . .	46
4.1. Distance Weighting Example . . . . .	48
4.2. Coefficient Plot of Spillovers across time and space . . . . .	60
5.1. Steady State: Visualizing Convergence . . . . .	80
5.2. Maintenance Funds relative to the cost of full maintenance . . . . .	82
6.1. First two amenity principal components, unrotated . . . . .	89
6.2. Distance disamenity in dollars per year . . . . .	94
6.3. Model implied amenities $\zeta$ in quality units . . . . .	96
6.4. Euler loss conditional on $\alpha_q$ . . . . .	97
6.5. Optimal Policy function with $\zeta = 0$ . . . . .	98
6.6. Ratios of constrained $q_{t+1}^c$ . . . . .	99
6.7. Difference between model and data growth rate in annual rental cost . . . . .	100

7.1. Opportunity Zones in Chicago with Selection Criteria . . . . .	103
7.2. Opportunity Zone Simulation: Price Impact Relative to Baseline . . . . .	107
7.3. “Random” Approach distribution: selecting and targeting all tracts . . . . .	111
7.4. “Random” Approach distribution: selecting and targeting below median tracts . . . . .	111
7.5. Comparison of selected tracts to best counterfactuals . . . . .	115
7.6. Comparison of simulated impact of selected tracts to best counterfactuals . . . . .	116
A.1. Lending Market Share of Rank 1 and 2 Lenders . . . . .	133
A.2. Lending Market Share of Rank 3 and 4 Lenders . . . . .	134
A.3. Lending Market Share of Rank 5 and 6 Lenders . . . . .	134
A.4. Lending Market Share of Rank 7 and 8 Lenders . . . . .	135
A.5. Lending Market Share of Rank 9 and 10 Lenders . . . . .	135
A.6. Lending Market Share of Rank 15 and 20 Lenders . . . . .	136
B.1. Within-Year Significance of entity-year FEs for Top Entities . . . . .	139
B.2. Within-Year SD of entity-year FEs for Top Entities . . . . .	140
B.3. Entity FEs in 2009 for Top Entities . . . . .	140
E.1. Assessment Information: 2010 Tax Year . . . . .	153
G.1. Varying $\omega$ : Price Impact . . . . .	159
G.2. Investment cost curve for variants of $\rho$ and $A_m$ . . . . .	161
G.3. Varying $\rho$ and $A_m$ : Price Impact . . . . .	162
G.4. Varying $A_m$ : Price Impact . . . . .	164
G.5. Price impact of Obama Presidential Center after 5 years . . . . .	165
G.6. City Price Impact of 1SD Shocks to Amenities after 5 years (% , PW)" . . . . .	167
H.1. Opportunity Zone Simulation: Year-by-Year Price Impact . . . . .	169
H.2. Opportunity Zone Simulation: Year-by-Year Price Impact without constraints . . . . .	170

## LIST OF TABLES

3.1. County-Level Adj $R^2$ from FE Regressions on DHS Change in Lending . . . . .	26
3.2. Adj $R^2$ from FE Regressions on DHS Change in Lending . . . . .	29
3.3. Tract-Level Adj $R^2$ from FE Regressions on DHS Change in Lending . . . . .	31
3.4. Adj $R^2$ from FE Regressions on DHS Change in Median Borrower Characteristics . . . . .	34
3.5. FE Correlation: Lending and Borrower Characteristics 2005-2015 . . . . .	35
3.6. Correlation between Entity-Year FEs and entity health measures . . . . .	36
3.7. Simple Rejection Prediction - Cook County . . . . .	42
4.1. Summary Statistics . . . . .	52
4.2. Housing Investment (Dollars per HH) on Standardized Credit Supply . . . . .	54
4.3. Housing Investment (Dollars per HH) on Credit Shortfall . . . . .	55
4.4. Effects of Credit Shortfall on Housing Investment . . . . .	57
4.5. Spillovers in Housing Investment across space . . . . .	58
4.6. Spillovers in Housing Investment across time and space . . . . .	58
4.7. IV: Instrumented Spillovers . . . . .	61
4.8. IV: Lag Instrumented Spillovers . . . . .	62
6.1. Baseline Parameter Values . . . . .	94
7.1. Price impact of Chicago's selections compared to best counterfactuals . . . . .	112
C.1. Lagged Credit Supply Shocks on Building Code Violations . . . . .	141
C.2. Lagged Credit Supply Shocks on Abandoned Housing . . . . .	142
C.3. Reduced Form Spillovers . . . . .	143
C.4. Lag Reduced Form Spillovers . . . . .	143
C.5. IV: Instrumented Spillovers with Single Instrument . . . . .	144
C.6. IV: Lag Instrumented Spillovers with Single Instrument . . . . .	144

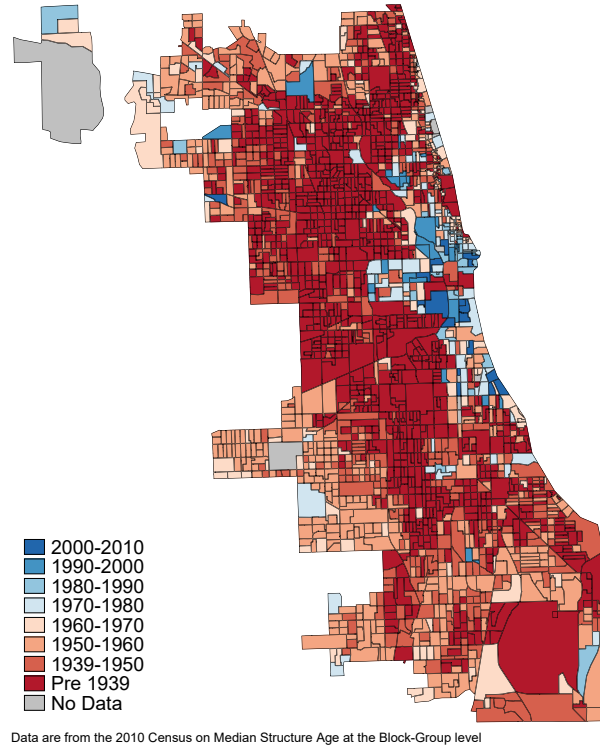
# 1. INTRODUCTION

How do local investment incentives and national housing credit supply shocks interact to affect the way in which neighborhoods grow, decline, and change? On the one hand, cities are driven by their local economy, comprising local demand for goods and services, and neighborhood success depends on local factors, such as housing quality, safety, and school quality. On the other, cities are elements of an increasingly globalizing world, subject to the effects of financial markets, domestic and international trade, and a plethora of other brought on by technological innovation. In this paper, I focus on one aspect of homeowners' and landlords' financial decisions that is complicated by these local and national forces: whether or not to invest in the quality of their home.

In many ways, the home investment is a household-level decision: a husband and wife discussing whether to invest their savings in a new kitchen, a single mother forgoing maintenance on her house to support her child instead, or a landlord cutting short-term costs by patching a leaky roof rather than replacing it. However, as a universally relevant micro phenomenon, it quickly becomes the driving force behind the built environment in all cities. Houses are not rebuilt from scratch annually, but are instead the result of many decades and cycles of construction, improvement, maintenance, decline, and disrepair by one or more owners. Figure 1.1 shows that the majority of Chicago - aside from the downtown and some recently-gentrifying suburbs - was built before World War II. These investment cycles comprise many households' decisions at different points in time, and are thus influenced both by household characteristics, such as savings, future life plans, or job stability, and by neighborhood characteristics, such as school quality, shops and restaurants, local public goods, and crime rates.

Housing investment is a major fraction of household spending. Gyourko and Tracy (2006) note that, according to the American Housing Survey, maintenance and repairs amount to \$1,750 per household per year nationally. Data from the 2013 AHS for Chicago suggest that this number is \$3,750 in Chicago from 2012-2013, with households in the Chicago area spending over \$9 bil-

Figure 1.1: Median Year Structure Built - Chicago



lion in this period. Haughwout et al. (2013) further note that, according to the Census Bureau's aggregate residential improvement expenditures series, flow maintenance averages about 45% of the magnitude of new residential construction, peaking during the boom at around \$145.6 billion in 2006 before declining to \$112.0 billion in 2010. These official figures, which amount to 3% to 6% of household income, likely understate the true economic cost of residential improvement, which often includes many forms of undocumented minor projects or substantially undervalued sweat equity.<sup>1</sup>

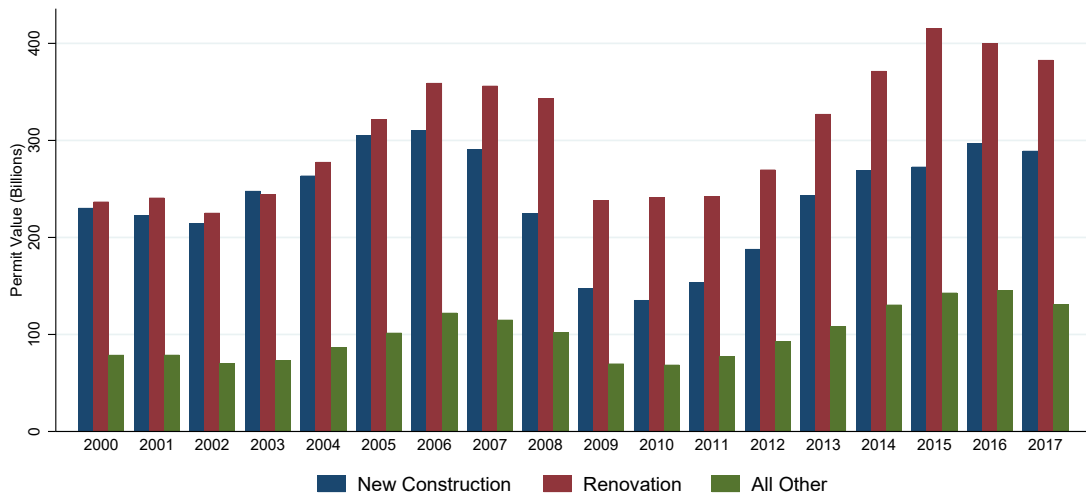
Figure 1.2 and figure 1.3 show that renovation expenses have matched or outstripped new construction permit values both nationally and in Chicago for almost all years since 2000.<sup>2</sup> Unlike

<sup>1</sup>Many building permit applications list \$1 as the cost of repairs when individuals do it themselves, severely understating the real economic cost of maintenance. Haughwout et al. (2013) note that the CEX data suggests that households use only their own labor for projects 37% of the time, and supply some of their own labor 3% of the time. The 2013 AHS data for Chicago provide similar estimates, with 915,250 projects done by professionals compared to 384,850 done as DIY, with professional projects costing roughly four times as much. The extent to which these costs are incorporated into survey or administrative measures of investment is uncertain.

<sup>2</sup>The national exception is 2003, in which renovation and new construction expenses are very similar, and 2004-2007 in Chicago, during the housing boom.

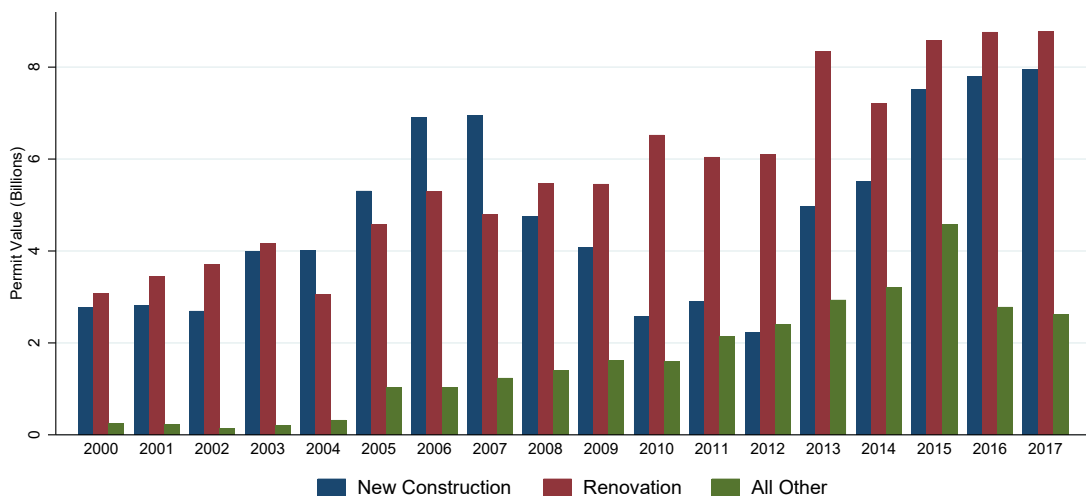
new construction, which is beyond the budget of a majority of American households, renovation expenses can vary depending on the project and can be undertaken by households with much lower savings due to having lower fixed and variable costs. Because renovation and maintenance expenditures are economically large and feasible forms of investment for many households, they have the potential to be affected directly by financial forces that shock household balance sheets.

Figure 1.2: National Permit Values by Type



Source: BuildFax

Figure 1.3: Cook County Permit Values by Type



Source: BuildFax

This paper explores the driving forces behind this housing investment process in the city of

Chicago.<sup>3</sup> I focus on Chicago because it has a diverse demographic distribution, comprises a number of adjacent yet heterogeneous neighborhoods, and has a relatively simple and featureless geography. Together, these features create an ideal environment in which to study the effects of credit supply on a variety of neighborhood types, and to see if these effects spill over to spatially adjacent neighborhoods that may have significantly different demographics. To measure housing investment and disinvestment, I use building permit, building code violation, and abandoned housing data from the city of Chicago. Though I'm agnostic about the specific spillover mechanism in the analysis, in theory I interpret spatial spillovers of minor repairs and major renovations as a combination of two factors that are often distinct in policy debate. First, an economic incentive; households' perception of the value of investing is dependent on the quality of their neighborhood, and is specifically driven by consumer preferences for both neighborhood and housing quality. Second, a "broken windows" or "physical disorder" hypothesis; the prevalence of disorder such as unkempt yards, abandoned buildings, and broken windows within a neighborhood discourages attentiveness to one's own house. The former mechanism is often framed as an investment decision on the part of individual homeowners, while the latter is social, psychological, and generally attributed to neighborhoods as a whole. However, in this paper and to the economist generally, both hypotheses are the same. Investment decisions are based on expectations about future home value and rental income, which are partially informed by the same neighborhood forces that underlie the broken windows hypothesis. Specifically, the level of physical disorder provides potential investors with information about past and present neighborhood conditions, which is relevant to the economic investment decision if these conditions persist through time.

Estimating spillovers directly is susceptible to a number of econometric concerns highlighted by Manski (1993), such as reflection - an endogeneity problem associated with estimating the effect of a spatial lag of a variable on itself - and correlated unobservables - where common

---

<sup>3</sup>Chicago is part of Cook County, Illinois. For much of the analysis, I focus on the city of Chicago proper. For some, I reference Cook County (FIPS 17031), which includes Chicago and a few surrounding suburbs.

spatial shocks can look like spillovers, but changes in the variable and its spatial lag are instead both caused by another variable. To determine a causal spillover channel, I turn to national refinancing lending data to understand how mortgage originators' exposure to the housing crisis affected their national strategy, and therefore the supply of credit and the borrower standards they applied in Chicago tracts. These credit supply and standards changes have the potential to affect households' ability to refinance, directly influencing their flow income, net worth, and cash available for housing investment. In other words, these shocks are primarily interpreted as affecting household balance sheets, similar to the balance sheet and net worth channels relating financial effects to real outcomes in Mian et al. (2013) and Mian and Sufi (2014). Households or landlords that are unable to refinance their mortgages because of changing credit conditions are less likely to make investments in their house, whether for their own flow benefit or to improve the house's sale or rental value.<sup>4</sup>

To measure credit supply and standards shocks, I use Home Mortgage Disclosure Act (HMDA) data on the universe of home mortgage loan flows from 1992-2015 to construct two census tract-year-level measures, with a focus on refinancing credit: a shift-share shock identified by a lender's national health and a credit standards shock based on changes in lending standards by Chicago's lenders both within Chicago and nationally during the 2007-2009 financial crisis. The former shock is based on a methodology also employed by Greenstone et al. (2014) to study the effects of small business credit supply on employment and by Mondragon (2017) to study the effects of housing credit supply on employment. The latter is novel, and specifically leverages the housing crisis and its effects on lender liquidity insofar as it drives lending standards.

Overall, the instrumental variable approach combines these pieces to understand the full effect of credit supply forces - both the direct effect of credit on households, and the subsequent spillover effect within and across neighborhoods. In this approach, I trace the effect of neigh-

---

<sup>4</sup>There is scope for home mortgage supply shocks acting as a proxy for business credit supply shocks - i.e. being a good measure of local bank health, rather than explicitly housing credit. However, Greenstone et al. (2014) find that the supply of small business credit has minimal effect on real outcomes, suggesting that this alternative channel may be small, and any effects of bank credit supply indeed operate through households.

boring tracts' credit supply shocks, controlling for a tract's own credit supply shock, through neighboring tracts' investment to investment levels in the initial tract. Empirically, I find that credit supply plays a substantial role in home investment, with a one standard deviation (1SD) refinancing credit supply shock increasing total home investment by \$41 per household per year, and renovation investment by \$49 per household per year in the baseline, up from means of \$221 and \$172 respectively. There is evidence that negative credit supply shocks lead to substitution from expensive investments (renovation) to more minor investments (easy permits). Tightening credit standards also play an important role controlling for refinancing credit supply, with a 1SD credit shortfall leading to a decline in housing investment of between 6-15%. In addition, I find small negative relationship between refinancing credit supply and residential building code violations and housing abandonment in the subsequent year. The former is driven by the intensive margin - an increase in the number of violations per house rather than the number of houses violating - suggesting that negative credit supply shocks may be particularly harmful to households who are already credit constrained, and become unable to perform even basic maintenance after a shock. This is corroborated by the abandoned housing result. In addition to these direct, first stage effects I find a strong role for spillover effects, with a 1% increase in lending predicting a 0.748-0.895% increase in investment in the IV estimation depending on investment type. Even estimating this specification with temporal lags leads to a modest positive effect of between 0.229-0.414%, suggesting some persistence across time.

Based on these empirical results, I present a dynamic model of housing investment that incorporates the spillovers and implied credit constraints present in the empirical analysis. The model extends the monocentric city model by adding a dynamic maintenance decision by building managers - landlords or homeowners - that depends on the quality of nearby properties through workers' preferences, and therefore willingness to pay, for neighborhood quality. By modeling maintenance as a dynamic process, I critically allow for spillovers to play a central role in the evolution of cities through intertemporal and spatial incentives. If a building

manager's maintenance decision depends on the neighborhood quality that she expects in the future, then her dynamic problem is intrinsically linked with the simultaneous problem being solved by her neighbors. These spatial linkages, combined with building managers' dynamic decisions, lead to an important role for strategic complementarities across space. Testing the existence of these complementarities and understanding their magnitude is an important first step in evaluating policies that focus on improving neighborhoods through improving building quality or subsidizing housing development in one part of the neighborhood. For instance, if strategic complementarities are weak, then improving the quality of one tract in a city is unlikely to incentivize investment in nearby tracts in that neighborhood or in surrounding neighborhoods, and is unlikely to be a cost-effective policy. I calibrate this model to the city of Chicago, obtaining a correlation of 0.48-0.53 (not targeted) between actual and model growth rates of housing quality and prices from 2010-2017. I then structurally estimate the magnitude of consumers' preference for the quality of their specific census tract relative to surrounding tracts, and find that the model fits best when own-tract preferences are 0.55 relative to neighboring-tract preferences of 0.45. This suggests an important role for neighborhood effects in pricing and investment decisions, as well as noticeable spillover effects for tract-level shocks or policies.

Using the calibrated model, it is possible to simulate the effects of a of tract and city-level policies that affect taxes, investment costs, amenities, or commuting costs. To demonstrate this, I estimate the impact of Chicago's "opportunity zones" on medium-run housing quality and prices. Opportunity zones are census tracts selected by state governments and approved by the US Treasury, subject to a per-state quota, in which investors can defer capital gains on investment. For the city of Chicago, 136 of the 806 census tracts were selected, primarily in distressed areas with above median unemployment and poverty rates, and below median family income. I find that the city's 136 selected census tracts lead to a simulated increase in house prices in the city as a whole of 4.44% after 5 years of the policy, and 11.87% in distressed tracts specifically relative to a baseline simulation without the policy. Using a "random" and a "single tract"

approach, I run a large number of counterfactual simulations in which I select alternative census tracts subject to a variety of selection and targeting criteria, and find that impacts of 8.44% and 17.27% on the city and distressed tracts respectively are possible depending on the policy targets. In particular, I find that there are many counterfactuals that outperform the city's selection both at the city and distressed tract levels, even taking into account potential political limitations on tracts that are eligible for selection. Overall, the model provides a useful benchmark both for understanding the magnitude of impacts that can be expected from policy interventions and for guiding where the city should target policy interventions to have the largest impact.

### 1.1. RELATED LITERATURE AND POLICY

In addition to the papers previously mentioned, this paper builds on a diverse literature on the theory of city equilibrium and development, the role of housing as a durable good, gentrification, the existence of coordinated development frictions, and the real effects of credit shocks through debt overhang and their effect on household balance sheets. In this section, I highlight some of the most closely related papers on these topics.

First, it's worth noting that most spatial models of cities, even those that make extensive theoretical contributions to the baseline monocentric city model along other dimensions, do not take the maintenance and renovation decisions of homeowners and landlords into account. This is true despite the fact that the monocentric city model has been studied extensively in the urban literature, and has been extended along myriad other dimensions. Starting with the baseline monocentric model summarized by Duranton and Puga (2015), the literature has considered altering the city's shape from a line to a circle (Lucas and Rossi-Hansberg (2002)) to a network of blocks connected by transportation infrastructure (Ahlfeldt et al. (2015)) to incorporating radial commuting highways (Anas and Moses (1979), Baum-Snow (2007)). Similarly, the distribution of land use has been extended from monocentric to endogenous determination of land use by

firms and workers (Lucas and Rossi-Hansberg (2002), Fujita and Ogawa (1982)). Commuting costs have been considered as both monetary, time costs, and heterogeneous costs across individuals (Behrens et al. (2015)). Spatial utility benefits such as local public goods (de Bartolome and Ross (2003)) and heterogeneous amenities (Brueckner and Rosenthal (2009)), as well as heterogeneous residents (Braid (1981), Määttänen and Terviö (2014), Landvoigt et al. (2015)) have been studied in detail. In large part, maintenance has been overlooked because many existing models are static and focus primarily on the thought experiment of building a city once from the ground up, with the quantity - or quality - of housing chosen once. Maintenance, an inherently dynamic process, is difficult to incorporate in such a model.

A large theoretical literature also analyzes the developers' problem, again with a focus on new construction rather than maintenance and renovation. These papers consider how developers make investment decisions while taking into account that they will recover their investment as a discounted stream of rents. Anas (1978) characterizes a solution to this problem in the context of a monocentric city with a myopic developer that does not internalize changing demand conditions. Capozza and Helsley (1989) and Capozza and Helsley (1990) extend this work to include developers with perfect foresight and developers with uncertainty respectively, which capture preemptive development in growing cities that does not occur with a myopic developer and show that it is forestalled slightly by uncertainty. These models are suggestive of the importance of coordination and the potential for spillover effects in models that combine space, development, and gains to coordinated investment. Building on this work, Fujita (1982), Wheaton (1982), and Turnbull (1988) demonstrate that the existence of heterogeneity in housing types can generate leapfrog development that is commonly observed in empirics: certain areas further from the central business district are developed earlier. Finally, Braid (2001) combines perfect foresight and redevelopment, but ignores maintenance, and shows that under certain functional form assumptions, the developer's optimal redevelopment problem at any distance can be solved by applying an appropriate transformation of the solution to the problem at unit distance. This literature provides the backdrop for the model in this paper, and relies on many

of the same neighborhood-related incentives and disincentives to investment. However, the inclusion of dynamic maintenance is either implicitly or explicitly (as in Braid (2001)) ignored in all of these models.

Substantial previous work has been done on durable housing and its relevance for city growth and decline. Particularly relevant is Glaeser and Gyourko (2005), who note that there is an asymmetry between growing and declining cities due to housing being a durable good. The authors show empirically that when a city is growing, its growth is primarily reflected in population, while city decline is largely seen in prices. This is because it is easier to build houses to fit demand than to wait for houses to deteriorate. This paper implies the importance of a maintenance channel in moderating the rate of decay of the housing stock in a way that varies across cities with different fundamentals. In a growing city, depreciation is worth staving off, while in a declining city depreciation may be optimal because demand is not high enough to support the current quality or quantity of housing.

Relatedly, there is a large and related literature on neighborhood externalities. For instance, Guerrieri et al. (2013) analyze gentrification and house prices under the broad assumption that rich people like living near one another. The authors are intentionally agnostic about potential underlying mechanisms, many of which have been proposed in other papers, such as decreased crime, higher amenities, and better schools. I consider another mechanism in this paper - housing quality, maintenance, and investment.

Most closely related to my work is Bayer et al. (2016), who document neighborhood spillovers and investment contagion at a very granular spatial level. They find that during the housing boom before 2007, novice home buyers (speculators) are drawn into investing in housing - either to hold or to “flip” - during the boom as a direct result of observing investing activity in their own neighborhoods, only to lose money when the bubble bursts. These results are corroborated by Mian and Sufi (2018), who add that these marginal buyers had lower credit scores and higher default rates, and helped trigger the mortgage default crisis. Though the channel of

investment due to neighborhood effects is similar, my work differs by considering renovation investments by landlords or homeowners generally, either using previously owned property or by specifically focusing on investment after the purchase itself, throughout the later boom, bust, and recovery.

There is also a small but critically related literature that considers the effects of large shocks on development. Siodla (2015) studies the 1906 San Francisco fire, while Hornbeck and Keniston (2017) study the Great Boston Fire of 1872. Similarly, Autor et al. (2014) analyzes the effects of the sudden elimination of rent control in Cambridge, Massachusetts and its effects on rental prices. All three papers show coordinated development due to overcoming previous economic redevelopment frictions. In the first two papers, “creative destruction” of multiple adjacent lots stimulates a need for simultaneous investment, while in the latter the promise of future rental increases incentivizes investment and raises the values of never-controlled prices, suggesting that the market internalizes ongoing and future neighborhood quality changes into home prices. Intuitively, these are suggestive of a mechanism that is conceptually central to this paper as well; growth may be restricted when building managers do not - or are not able to - internalize positive spillover effects from their own investment.

Finally, Glaeser et al. (2015) show that Google Street View data can be used to find clear correlations between visual housing quality and metrics like income and home valuation. While these authors don’t explicitly address spillovers or financial channels, they provide data-driven evidence that observable housing quality is a metric that is fundamentally tied to important measures of economic welfare and housing prices. This suggests that neighborhood quality, and changes to neighborhood quality due to housing investment, are observable neighborhood features that homeowners and landlords can use when making their investment decisions.

In addition to these urban results, the effect of financial market shocks on households generally has been a central area of research since the 2007-2010 financial crisis. As previously noted, Mian et al. (2013) and Mian and Sufi (2014) demonstrate that housing market turmoil can have

a crippling effect on household balance sheets through home values, and that this can severely affect households' marginal propensity to consume. Similarly focusing on home values, Haughwout et al. (2013) and Melzer (2017) demonstrate that homeowners in negative equity or at risk of default reduce investment in their property because increases in the house value resulting from housing investments are likely to go to the lender rather than the household. This implies that refinancing credit shocks that limit homeowners' ability to escape debt overhang can have a substantial effect on investment. Mondragon (2017) shows that credit supply shocks can even affect local employment through depressed demand by constrained households even controlling for home prices, suggesting that credit supply can play a direct role through its effect on households' liquidity rather than solely through debt overhang. Benmelech et al. (2017) demonstrate that home purchases and housing market churn itself can have a stimulative effect on home investment expenditures, providing evidence that households spend money on home improvements both just before and in the year after a transaction. This suggests a role for home purchase credit shocks affecting housing investment. Finally, Gyourko and Tracy (2006) highlight the importance of saving and dis-saving in a house as a way for households to smooth consumption, noting that maintenance and repair expenditures respond to transitory income shocks. This is critically related to my analysis, given that one can interpret refinancing credit supply shocks as potential shocks to a household's annual net income.

Overall, I seek to analyze how this work on the effects of financial forces on households affects cities through the lens housing investment, and to provide results that are relevant for policymakers seeking to understand neighborhood change, how to incentivize development, and where to invest in urban amenities in a way that maximizes impact. The role of public investment in improving neighborhood quality and spurring private investment is a central concept in public policy, and the core mission of organizations like Reinvestment Fund. These organizations work on a city-by-city basis to understand in which neighborhoods investment would be most effective, and seek to use local stimulus to overcome coordinated investment frictions similar to those discussed by Hornbeck and Keniston (2017). This paper complements these

efforts by providing a framework to think about the financial and local forces that limit private investment of homeowners and landlords and to assess their empirical importance in a way that is generalizable to many cities.

## 1.2. OUTLINE

The subsequent sections (i) review the data used for the empirical analysis, (ii) construct credit supply shocks and validate that they capture supply effects, (iii) discuss empirical results on the determinants of housing investment, (iv) present a dynamic model of housing investment, (v) apply a variety of data sources and the model to estimate the extent to which consumers demonstrate preference for surrounding tracts' quality using a moment-matching approach, (vi) simulate the model using historical data, to test its fit, and (vii) simulate the 10-year spatial impact of Chicago's opportunity zones using the model relative to a baseline.

## 2. DATA DESCRIPTION

In this section, I describe the two primary datasets that are used for empirical analysis. Administrative data for the city of Chicago are provided via the Chicago Open Data Portal, and include address-day-level data on housing investment and neglect in the form of building permits, building code violations, and abandoned housing. Home mortgage lending data are obtained from the Home Mortgage Disclosure Act (HMDA), which includes loan-level information on accepted and rejected home purchase and refinancing loans, including the year and census tract of the loan.

Two other proprietary datasets, BuildFax BuildFax (2018) and Zillow ZTRAX Zillow (2017), are also used in the analysis.<sup>5</sup> BuildFax provides aggregated county-month building permit data from 2000-2016 for over 900 US counties to provide a national perspective on housing investment.<sup>6</sup> Zillow's ZTRAX data provides extensive national home transaction and assessment microdata, which is used to classify Chicago building permits as residential or non-residential and to estimate parameters for calibration. These datasets are secondary to this paper and are described in appendix A.

### 2.1. HOUSING INVESTMENT DATA: THE CHICAGO OPEN DATA PORTAL

From the Chicago Open Data Portal, I use data on building permits, building code violations, crime, and abandoned housing. All datasets are obtained at the event-level (latitude-longitude-day), and are collapsed spatially to tract-year to match the most granular data identifiers in HMDA.

The primary dataset for measuring housing investment is administrative building permit data, which consists of all permits issued by the city of Chicago from 2005-2016.<sup>7</sup> After substantial

---

<sup>5</sup>BuildFax data was obtained courtesy of the John and Serena Liew Fellowship Fund at the Fama-Miller Center for Research in Finance, University of Chicago Booth School of Business.

<sup>6</sup>BuildFax data covers over 200 million US citizens, and counties included in the data are generally from larger cities for which permitting is more standardized and available historically.

<sup>7</sup>Data is available beyond 2016, but newer data were not necessary because HMDA data is released with a lag.

cleaning using natural language processing and a match to Zillow's ZTRAX to identify residential permits, there are 270,119 residential permits, of which 157,362 are easy permits, renovation, and porch construction.<sup>8</sup> Easy and renovation permits are differentiated by the extent of the project involved and what must be submitted to the department of buildings to proceed with the project. Easy permits are for minor repairs that don't require plans to be submitted, and are the most common kind of permit. Renovation/alteration permits are for more major projects that require plans or contractor information to be submitted. I also construct a measure of total housing investment, which combines easy permits, renovations, and porch construction, the latter of which is too minor to be considered independently.

There are two relevant outcomes in the data - the count and value of permits applied for by type, where value is defined as the estimated cost of the project reported in the permit application. This cost is a preferable measure of investment because it clearly differentiates projects by size, but is substantially noisier. Specifically, there are incentives to under-report to avoid fees, and limited review of whether the dollar value listed in the permit application matches the work being done. I run a series of lasso regressions on the words in the permit descriptions to adjust the values for permits with extremely low estimated costs, leveraging cost estimates from permits with similar descriptions. Details on data construction are provided in appendix A.1.

In addition to potential under-reporting of estimated costs to lower permit fees, one of the primary concerns with permit data is that it only captures a small subset of residential investment - the majority of investments made by households are either (i) done by hand or (ii) employ a professional but don't require a permit. Comparing the permit data from July 2011 - June 2013 to the 2013 American Housing Survey's question on households' home improvement spending which focus on this time period for cities as a whole suggests that of the \$7.07 billion of self-

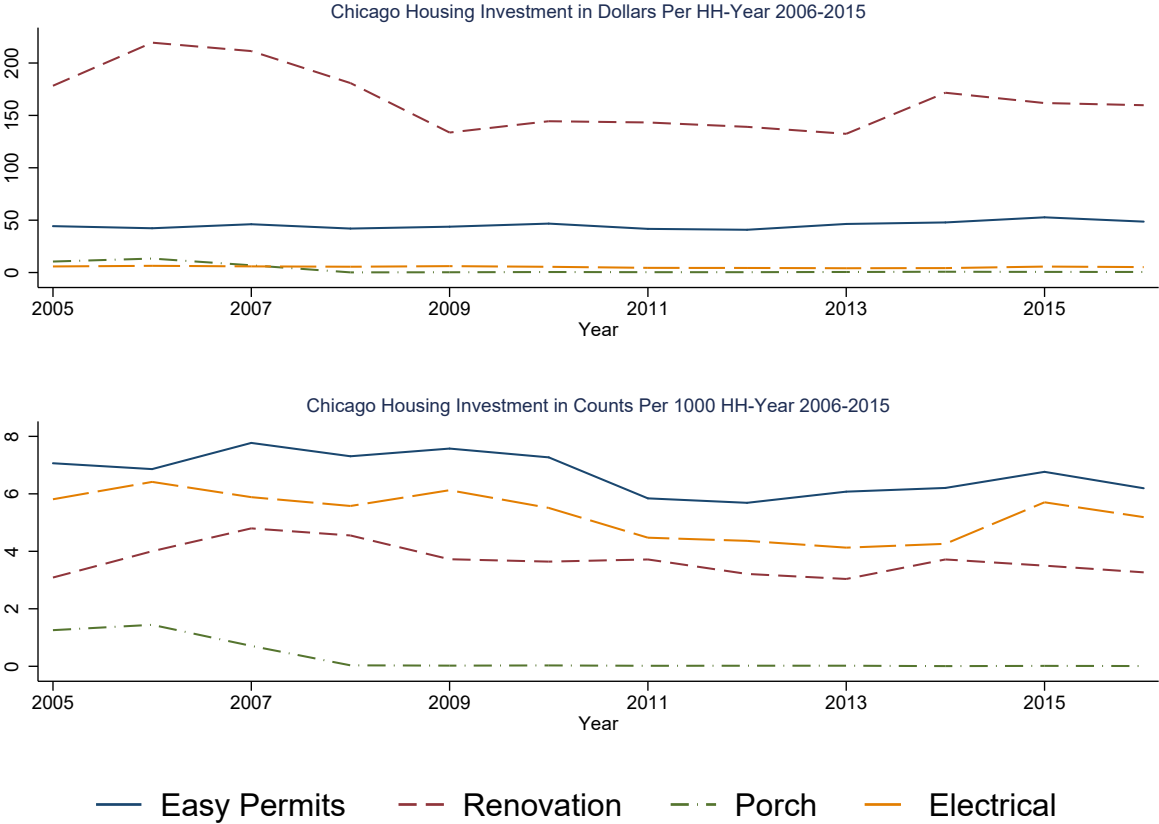
---

Data from 2000-2005 were obtained via FOIA, but data from 2004 and earlier cannot be matched to later data due to system and record keeping changes in 2004. Data from 2006-Present is publicly available on the Chicago Open Data Portal.

<sup>8</sup>I do not use new construction for many reasons: new construction (i) requires available land that does not exist in all tracts, (ii) is often done on a larger scale than other permits (i.e. multifamily dwellings) in cities, making it beyond the scope of most households, and (iii) is often non-residential, so in many tract-years new construction is 0, reflecting land supply frictions or substantial fixed costs.

declared home improvement activity in remodeling, renovations, interior additions, exterior additions, and other additions, between \$0.45 billion and \$1.05 billion are reported as either easy or renovation permits.<sup>9</sup> Therefore, permits capture between 6% to 14% of self-reported home investment activity. Despite this shortcoming, permits are desirable because they are both available at much more granular spatial units (address vs city), and at a higher frequency (daily vs biennial), than the AHS survey data. Furthermore, projects involving permits are more likely to entail real value added to a home, and to be highlighted in future home sales, making permit spending a better measure of capitalized household investment than household spending, which may include personalization expenses as discussed in Benmelech et al. (2017).

Figure 2.1: Housing Investment per HH-Year



<sup>9</sup>This depends on whether extremely high cost permits (above 95th percentile in size), which may be more closely associated with companies renovating apartment buildings, are winsorised. The 2013 AHS survey was run in the summer of 2013 from April to September, and asks respondents at the time of the interview how much they spent in the last two years, so July 2011- June 2013 is an approximation.

Figure 2.1 shows the four primary permit types available to current home owners. In my analysis, I focus on easy and renovation permits, because these represent the vast majority of housing investment dollars, as seen in the left pane. Electrical permits are common, but are often very low cost, and are generally accompanied by an easy or renovation permit if they actually reflect a significant investment. Porch permits are rare, and largely undocumented after 2007. For this reason, they are not considered separately in the analysis. Between the two major permit types, renovation is substantially larger in dollars of investment, but easy permits are more frequent. In the analysis, I consider three permit outcomes: easy permit values, renovation permit values, and total housing investment, which includes easy, renovation, and porch permits together.

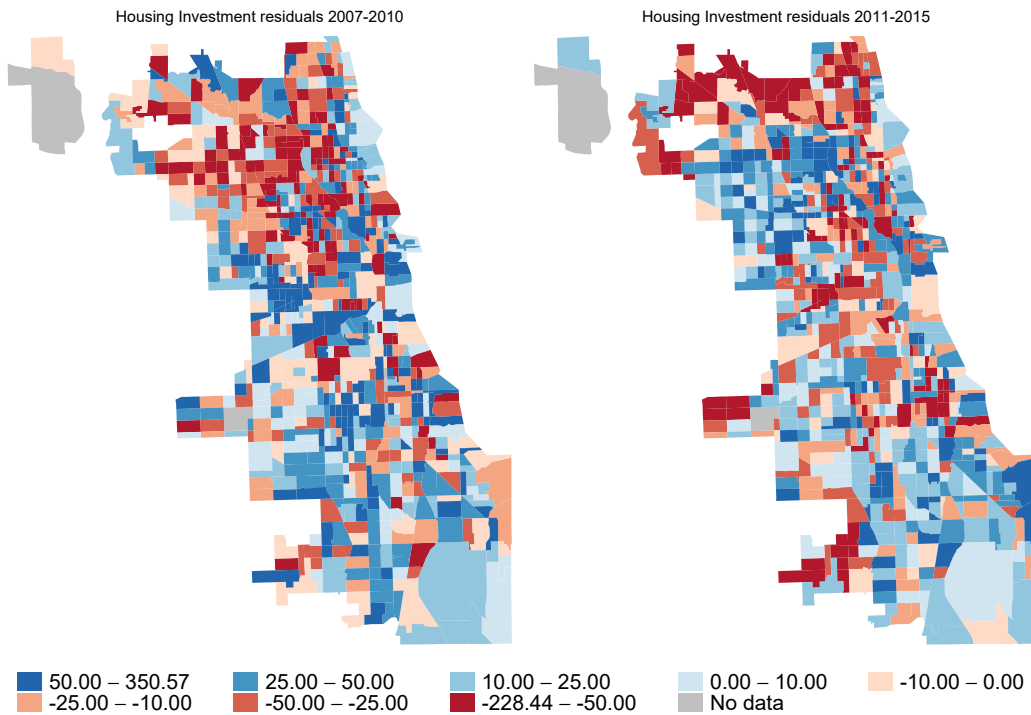
Figure 2.2 shows the spatial distribution of housing investment during the crisis and recovery, highlighting the heterogeneity that is persistent but evolving through both periods. This spatial heterogeneity suggests that the time series averages in figure 2.1 are insufficient for understanding Chicago's housing investment: neighborhoods within the city have significantly different rates of investment, and may respond differently during times of economics stress. Figure 2.3 depicts the same time periods and investment variable, but residualized by tract and year-level fixed effects to control for these consistent patterns. Controlling for time invariant tract-level heterogeneity and city-level annual fluctuations, there is substantial residual spatial heterogeneity in the effects that differs across the crisis and recovery and does not reflect Chicago's underlying demographics. Understanding the driving forces behind this residual pattern is the primary question of interest.

In addition to building permits, I consider some ex-post measures of lack of building investment, such as building code violations, and 311 reports of abandoned buildings. Further details are provided in appendix A.1.

Figure 2.2: Housing Investment per HH-Year in Crisis and Recovery



Figure 2.3: Housing Investment Residuals in Crisis and Recovery



## 2.2. HOUSING CREDIT DATA: THE HOME MORTGAGE DISCLOSURE ACT

### 2.2.1. DATA OVERVIEW

Credit supply and standards shock variables are derived from Home Mortgage Disclosure Act (HMDA) data starting in 2004, the first year in which standardized RSSD ID identifiers for lenders were widely reported, and continuing through 2015.<sup>10</sup> For some tables and figures, data going back to 1993 is used to provide historical context.<sup>11</sup> HMDA provides loan-level data on accepted and rejected home mortgage loan applications regarding loan size, loan type, lien status, census tract, MSA, applicant and co-applicant ethnicity, race, and sex, whether the loan was sold to another institution within the year, and other optional reporting information. Details about the data source and data cleaning process, including how to obtain the correct ultimate parent RSSD ID for each lender, the intentional lack of merger adjustment, and handling changes in geography across time are provided in appendix A.2.

### 2.2.2. DATA SUMMARY

The HMDA data from 2004-2015 paints a stark picture of the housing market in the aggregate. Figure 2.4 shows the fall and rise in home mortgage lending through the crisis and into the recovery. By contrast, refinancing did not experience a similar pattern for two reasons. First, interest rates dropped substantially during this time, incentivizing qualified homeowners to refinance if able. Second, many government programs, including the Making Home Affordable (MHA) and Hardest Hit Fund (HHF) programs helped underwater borrowers refinance their homes even if their income or credit score would have otherwise made them ineligible for new rates. This is particularly true in 2009 and 2012, when each of these programs respectively hit their peaks. Chicago, as the major city in a state included in the Hardest Hit Fund program, experiences a trend roughly similar to that of the US at large, with steady refinancing through

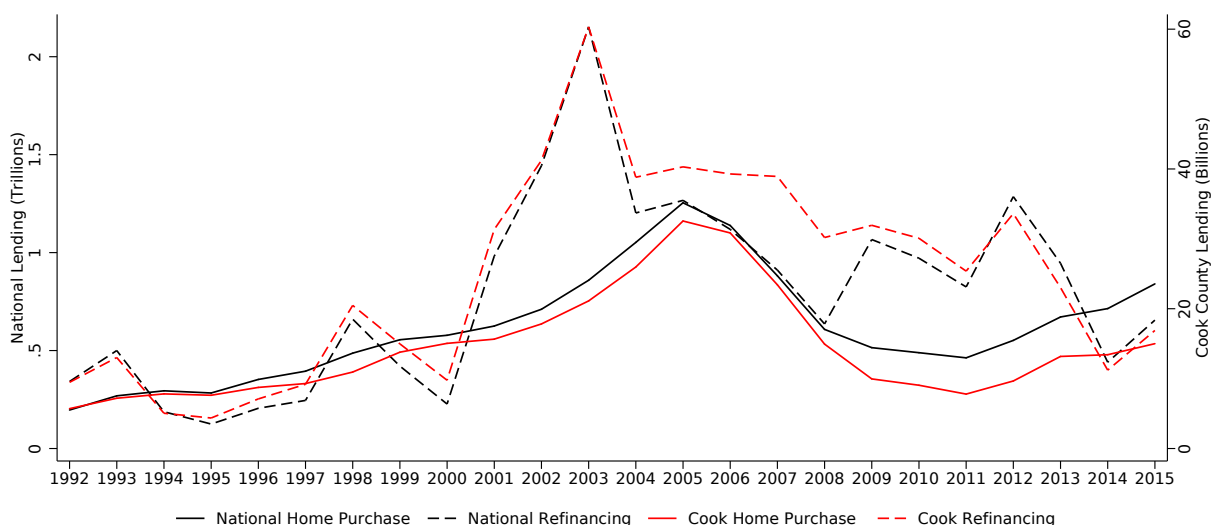
---

<sup>10</sup>RSSD ID stands for Replication Server System Database ID, and is the unique identifier assigned to financial institutions by the Federal Reserve.

<sup>11</sup>Matching of RSSD IDs from this period uses the Avery Files, provided publicly by Bob Avery and Neil Bhutta. Data from before 2004 is not used for the main analysis because Chicago permit data start in 2005.

the crisis, and a decline at the end, when HHF has already been used by eligible homeowners and interest rates have risen.

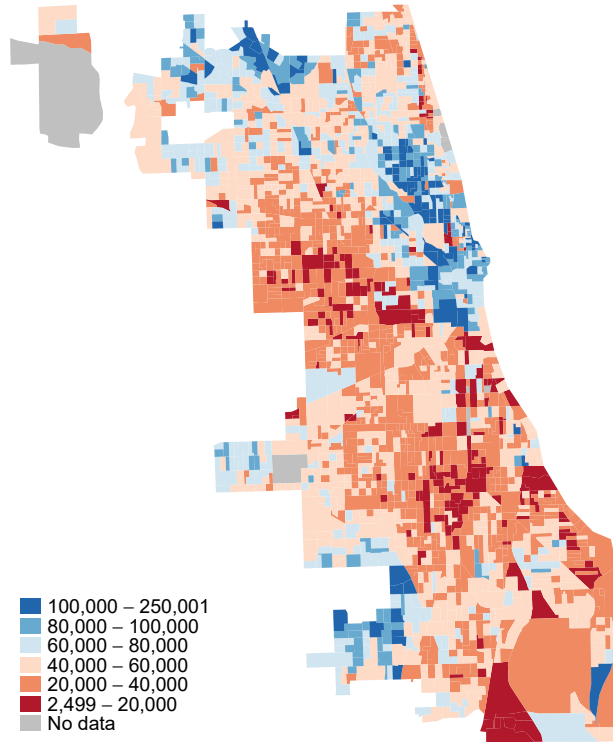
Figure 2.4: Originated Home Purchase and Refinancing Loans - Chicago



These aggregate patterns mask important heterogeneity within the city of Chicago. Chicago, like many American cities, is organized around neighborhoods that have substantially different makeup in terms of race, income, age, and other demographic characteristics. Figure 2.5 depicts the stark income dimension of this heterogeneity, with some parts of the city well below the poverty line, while others are among the richest census tracts in the US. This heterogeneity implies that it is insufficient to understand how Chicago households as a whole responded to credit supply shocks, as different neighborhoods in the city face drastically different economic conditions and challenges.

In addition to these demographic differences, there is substantial heterogeneity in the mortgage lending environment in different parts of the city. The market share and spatial distribution of the ten largest entities is depicted in appendix A.3. Overall, there are (i) commercial bank lenders, whose lending behavior is spatially correlated with their physical branch locations in the city, (ii) national lenders who are spread across the city, but focus on certain neighborhoods via word-of-mouth or other advertising, local demographics, or a non-bank storefront, and (iii)

Figure 2.5: Median Household Income 2006-2010 ACS - Chicago



other non-bank lenders with obvious and specific focus on lending only to subsets of the city, also potentially because of advertising.<sup>12</sup> Lenders with lower levels of total assets who fall into the third category above tend to be more spatially concentrated in their lending and only operate in a subset of US states; presumably to focus on neighborhoods in cities where they have a competitive information or local expertise advantage over large lenders.

This spatial heterogeneity in lending is critical for the construction of credit shocks that affect parts of the city differently, as each location's shock will depend on the portfolio of lending entities available to it, as defined by recent lending patterns. In a space as compact as a city, identifying differences in credit shocks between neighboring census tracts rely on there being a large number of lending entities, each with a unique lending distribution across the city.

---

<sup>12</sup>Commercial bank locations are not random, but are generally decided pre-crisis and either maintained or closed through the crisis and recovery. These closures are captured in the credit supply shock as a negative shock.

### 3. CONSTRUCTING AND VERIFYING CREDIT SHOCKS

In this section, I construct two different versions of credit shocks using HMDA data. The first shock is a refinancing credit supply shock, following an adjusted shift-share methodology used in Greenstone et al. (2014) to construct credit supply shocks in small business lending, and also applied to housing markets in Mondragon (2017). The second is a credit standards shock, which uses changes in acceptance and rejection rates before and after the financial crisis within the city of Chicago as a measure of the relative stringency of financial institutions' standards in different parts of the city. For the former shock, I show how it is constructed, estimate it, and visualize it. I then provide a test to verify that it is a reasonable measure of credit supply that excludes credit demand and demonstrate that it is (i) positively correlated with entity stock returns and (ii) negatively correlated with entity reliance on pre-crisis short term debt.<sup>13</sup>

#### 3.1. THE CREDIT SUPPLY SHOCK

##### 3.1.1. THE CREDIT SUPPLY SHOCK: CONSTRUCTION

To construct a credit supply shock, I follow Greenstone et al. (2014) in identifying supply and demand as fixed effects from a regression at the entity ( $i$ ) - areal unit ( $j$ ) - year ( $t$ ) level, where areal units can be defined either as counties or census tracts. Specifically, the specification takes the following form:

$$\text{LendingChange}_{ijt} = \alpha_{it} + \beta_{jt} + \epsilon_{ijt} \quad (3.1)$$

Where  $\text{LendingChange}_{ijt}$  is the Davis et al. (1996) (DHS) growth rate in the dollar value of lending by a bank in a given areal unit:

$$\text{LendingChange}_{ijt} = \frac{\text{Lending}_{ijt} - \text{Lending}_{ijt-1}}{.5 * (\text{Lending}_{ijt} + \text{Lending}_{ijt-1})}$$

---

<sup>13</sup>Short term debt reliance is one of many potential measures of exposure to the financial crisis.

Use of the DHS growth rate is a slight departure from Greenstone et al. (2014), who use log differences. Mathematically, DHS growth rates are a second order approximation of the log difference for growth rates around zero. In practice, DHS growth rates are similar in that they are a symmetric growth rate measure, but have the advantage of being able to incorporate entry (value of 2) and exit (value of -2) which is common for lenders at granular areal units.

$\alpha_{it}$  are entity-year fixed effects, and are identified by a consistent pattern of entity lending across counties. These fixed effects are interpreted as entity supply of credit. Similarly,  $\beta_{jt}$  are areal unit-year fixed effects, identified from consistent pattern of lending within areal unit by all entities. The specification is weighted by  $Lending_{ij,t-1}$ , the entity's previous period lending in the county.<sup>14</sup>

Intuitively, the Frisch-Waugh-Lovell theorem allows us to interpret this methodology in two steps. First,  $\beta_{jt}$  controls for common lending patterns by all entities within an areal unit-year - a proxy for local demand for lending. The identifying assumption required here is that reduced demand in a county is spread homogeneously, or at least randomly, across lenders. If this is not the case, then even after controlling for areal unit-year fixed effects, some residual variation in lending may still be due to demand effects that are unevenly distributed across lenders, presumably because lenders cater to different borrowers that are heterogeneous effected by demand shocks. I provide a test of this identifying assumption in the subsequent section.

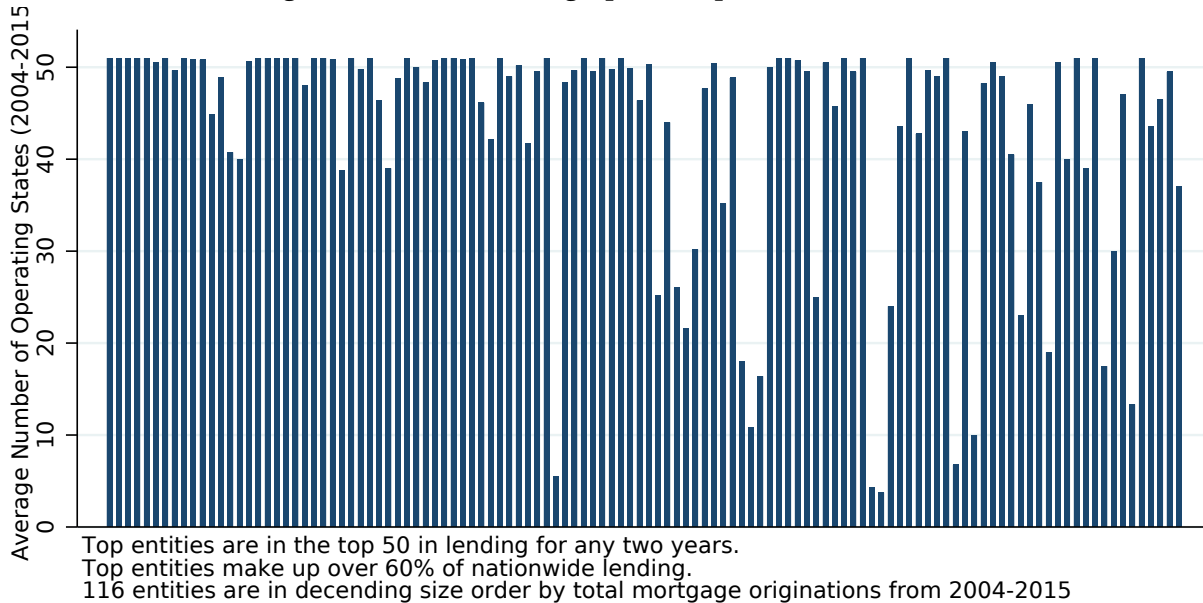
Once local demand is controlled for,  $\alpha_{it}$  captures consistent patterns in lending by each entity across areal units within a year. For  $\alpha_{it}$  to be identified, it's critical to have sufficient geographic variation in lending within entity. Figure 3.1 shows that the largest entities in the data are almost all national, with some large regional lenders. Importantly, it's not necessary that *all* entities are national;  $\alpha_{it}$  are well estimated for local lenders because the presence of national banks

---

<sup>14</sup>Weighting by previous period lending omits entering lenders from being included in the specification. In some robustness exercises, alternative weights, such as the denominator of DHS changes  $.5 * (Lending_{ij,t} + Lending_{ij,t-1})$  are used. However, previous period lending is used as a baseline because all weights that include entering lenders necessarily depend on their lending volumes in periods  $t$  and later, which are positively correlated with growth rates.

helps accurately distinguish local demand effects from national entity supply effects, allowing for identification of entity-year effects for community lenders.

Figure 3.1: Level of Geographic Dispersion of Lenders



After obtaining  $\alpha_{it}$ , following Greenstone et al. (2014) I convert this from a measure at the entity-year ( $i, t$ ) level to one at the areal unit-year level ( $j, t$ ), using previous year lending shares by lender  $i$  in the areal unit  $j$ ,  $\rho_{ijt-1}$ .

$$\hat{\alpha}_{jt} = \sum_i \rho_{ijt-1} \hat{\alpha}_{it} \quad \sum_i \rho_{ijt-1} = 1 \quad \forall j, t \quad (3.2)$$

Notably, this transformation does not require that  $j$  match the geographic level at which  $\hat{\alpha}_{it}$  are estimated. For my baseline analysis, I estimate  $\alpha_{it}$  using county-year level data, and then project this to the tract-year level. The results are qualitatively similar running the first stage at the tract-year level, but noisier, as tract-year lending data features much more entry, exit, and statistical error.<sup>15</sup>

<sup>15</sup>Note that the use of DHS changes makes entry and exit better handled in my methodology than it would be if lending were calculated as a log difference. However, at the tract level, the large fluctuations in lending relative to the baseline level combined with entry and exit make the relative magnitude of these DHS changes important, which is not well suited to an approximation around a growth rate of zero. Therefore while DHS changes handle entry and exit in that these observations are not dropped as they would be with a log difference, less

$\alpha_{it}$  can also be constructed using data on only part of the country. For my analysis on the city of Chicago, I omit Cook county (FIPS 17031) from the initial specification to ensure that the entity-year FEs are estimated using data that can only be correlated with the outcome measures through the lenders.<sup>16</sup> I standardize both the supply shock ( $\hat{\alpha}_{jt}$ ) and the demand shock ( $\hat{\beta}_{jt}$ ) to have a weighted mean of 0 and a weighted SD of 1 across the national sample.

### 3.1.2. THE CREDIT SUPPLY SHOCK: ESTIMATION

Table 3.1 provides a breakdown of the adjusted  $R^2$  of specifications for county-year-entity level data, adding fixed effects one-by-one in working up to equation (3.1) in specification (7). Specifications (1)-(3) show the fraction of variation explained just by year, county, and entity fixed effects independently. Year fixed effects are important for refinancing credit because they proxy for national interest rate fluctuations. County fixed effects alone explain very little variation, because much of the country from 2004-2015 experiences both a decline and a rise in lending, with relatively few counties experiencing growth or decline throughout the sample. Entity fixed effects are of moderate importance for both lending types because some lenders consistently grew or declined in the sample period. Specification (4) combines these uninteracted fixed effects to provide a baseline for the subsequent specifications. (5) and (6) build on (4) by including all the same fixed effects, as well as county-year and entity-year FEs respectively. In these specifications, entity-year FEs (6) explain dramatically more of lending changes than county-year FEs (5), suggesting a substantial role for credit supply. Specification (7), which includes both county-year and entity-year FEs, also supports this conclusion. Overall, the explanatory power of entity-year level effects (from (4) to (6)), without any geographic information, is staggering, and suggests that lenders' financial health and willingness to invest in housing year-to-year plays a massive role in lending behavior nationwide.

The results in table 3.1 are robust a variety of alterations to the specification. In particular,

---

entry and exit still leads to more precise estimation.

<sup>16</sup>Cook county contains City of Chicago, as well as a handful of surrounding suburbs.

Table 3.1: County-Level Adj  $R^2$  from FE Regressions on DHS Change in Lending

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	Observations
All	0.084	0.006	0.151	0.269	0.320	0.743	0.780	5,639,230
Home Purchase	0.056	0.011	0.156	0.251	0.284	0.688	0.714	3,725,497
Refinancing	0.175	0.006	0.132	0.334	0.392	0.748	0.791	4,425,979
Year FE	Yes	No	No	Yes	Yes	Yes	Yes	
County FE	No	Yes	No	Yes	Yes	Yes	Yes	
Entity FE	No	No	Yes	Yes	Yes	Yes	Yes	
County-Year FE	No	No	No	No	Yes	No	Yes	
Entity-Year FE	No	No	No	No	No	Yes	Yes	

the results are the same after subsetting to only the largest lenders, dispelling concerns that it's the number of entities that drives the importance of entity-year FEs through overfitting.<sup>17</sup> In addition to subsetting to loans originated by only large entities, only small entities, or only banks, I consider the following dimensions of robustness in table 3.2:

- **Conforming vs. Jumbo/Non-Conforming.** Conforming loans are those that fall below the loan size threshold such that Fannie Mae and Freddie Mac are willing to purchase them. Despite the fact that these GSEs faced substantial problems during the housing crisis and ultimately required a federal bailout by the FHFA, banks perceive that loans sold to GSEs don't need to be held on balance sheet and ultimately do not require the same liquidity commitment by the lender. Therefore, one might expect lenders to act differently towards conforming and non-conforming loans when their balance sheet is more or less liquid. Some specifications also subset to lenders for whom over 5% of lending is jumbo - well above the national average - as these lenders are demonstrably different in their ability to hold large loans that cannot be sold to GSEs on their balance sheet.
- **Held vs Not Held.** Held loans are those that are originated and held, rather than originated and sold on within the same year either to a GSE or a private securitizer. Loans that are not held are generally originated by an entity who knows the loan can easily be sold on, and a fee can be collected for originating the new mortgage. By contrast, loans that are held

<sup>17</sup>In this robustness check, there are over 3000 counties and below 200 lenders, and entity-year FEs are still the primary explanatory force.

require equity from the lender itself, and reflect a bet that the mortgage will not become delinquent. These two possible loan actions imply very different intentions - and business models - by the originator, and could easily be differently affected in times of crisis, when originating loans for a fee may still be profitable if a buyer is already lined up, but holding loans on balance sheet may not be tenable.

- Sold to GSE. Loans that are immediately sold to GSEs are more likely to fluctuate with the health of GSEs than the health of the originator. Related to this, I include specifications in which I treat GSEs as the de-facto lender for loans that are originated by another entity and immediately sold to a GSE.
- Above or Below Median Income Borrowers. Borrowers' characteristics affect the risk of a loan, and during times of crisis, lenders may differentially tighten standards on different classes of borrowers. This is explored more when constructing the credit standards shock and verifying the credit supply shock.
- Alternate weighting. In the baseline specification, I use the previous year's lending amount as a weight. This importantly drops entities that are newly lending in a county-year, because their weight is zero. If I instead construct a weight using the entity's overall lending in the county across the sample period, the results are not substantively different.

Table 3.2 demonstrates that the importance of entity-year FEs is robust to all subsampling and changes to the specifications. Furthermore, the changes in the importance of entity-year FEs, as measured by the difference in  $R^2$  from specification (4) to (6), match closely with it being a good measure of credit supply. For instance, changes in conforming and held loans are substantially better explained by entity-year FEs than are conforming and not-held loans, reflecting the fact that loans that are not held provide relatively little information on the lender, since these are flipped so quickly that no liquidity is required. Similarly, entity-year FEs explain more variation in lending for large and very large entities than for small entities, consistent with the intuition that large lenders pursue a national lending strategy that is subject to national credit availability.

Notably, the results are not driven by overfitting due to an extremely large number of entity-year fixed effects. In addition to the use of adjusted  $R^2$  to penalize extraneous variables, the results are identical when subsetting to only large lenders, despite the fact that there are far more counties than large lenders.

Table 3.2: Adj  $R^2$  from FE Regressions on DHS Change in Lending

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	Observations
All	0.084	0.006	0.151	0.269	0.320	0.743	0.780	5,639,230
Conforming	0.079	0.005	0.145	0.259	0.313	0.729	0.771	5,553,631
Non-Conforming	0.150	0.036	0.148	0.348	0.393	0.711	0.755	953,901
Held	0.082	0.010	0.171	0.285	0.325	0.750	0.777	2,987,033
Not Held	0.089	0.006	0.153	0.267	0.316	0.744	0.780	4,215,377
Conforming, Held	0.077	0.009	0.164	0.274	0.320	0.741	0.772	2,901,567
Conforming, Not Held	0.083	0.006	0.150	0.258	0.307	0.735	0.773	4,180,465
Non-Conforming, Held	0.095	0.037	0.173	0.310	0.341	0.701	0.741	631,437
Non-Conforming, Not Held	0.214	0.036	0.122	0.402	0.444	0.735	0.771	511,147
Above Median Income	0.078	0.008	0.147	0.261	0.307	0.715	0.750	4,593,223
Below Median Income	0.084	0.008	0.147	0.269	0.317	0.713	0.750	4,013,023
Sold To GSE	0.128	0.006	0.160	0.302	0.353	0.764	0.810	1,951,601
Home Purchase	0.056	0.011	0.156	0.251	0.284	0.688	0.714	3,725,497
Refinancing	0.175	0.006	0.132	0.334	0.392	0.748	0.791	4,425,979
Jumbo Lenders: Conforming	0.081	0.005	0.141	0.261	0.316	0.742	0.788	4,166,293
Jumbo Lenders: Non-Conforming	0.154	0.034	0.146	0.349	0.394	0.717	0.761	877,204
GSEs as One Lender	0.151	0.004	0.179	0.340	0.422	0.740	0.821	3,115,492
Large Entities	0.196	0.004	0.070	0.289	0.312	0.774	0.797	1,466,918
Very Large Entities	0.208	0.003	0.053	0.290	0.308	0.782	0.809	973,637
Without Large Entities	0.123	0.010	0.172	0.324	0.349	0.766	0.775	4,209,957
Without Very Large Entities	0.121	0.008	0.165	0.314	0.340	0.763	0.773	4,703,237
Bank And BHC	0.140	0.006	0.113	0.280	0.347	0.713	0.771	2,216,941
Weighted by Average of t and t-1 Lending	0.103	-0.000	0.021	0.162	0.223	0.706	0.746	7,978,675
Year FE	Yes	No	No	Yes	Yes	Yes	Yes	
County FE	No	Yes	No	Yes	Yes	Yes	Yes	
Entity FE	No	No	Yes	Yes	Yes	Yes	Yes	
County-Year FE	No	No	No	No	Yes	No	Yes	
Entity-Year FE	No	No	No	No	No	Yes	Yes	

The relative importance of entity FEs relative to county FEs at an annual level is depicted in figure 3.2. The figure shows the adjusted  $R^2$  for the baseline specification, pooling home purchases and refinances, year-by-year with county FEs, entity FEs, and both. There are three main takeaways from this figure. First, entity FEs clearly explain more variation in all years than county FEs. Second, entity FEs better explain variation during the crisis (2007-2009) than during the recovery (2010-2015). Finally, the power of entity FEs to explain lending changes is relatively consistent from 1993-2009, after which it falls. This is suggestive of a shift in home mortgage lending that is not discussed in this paper, but could be due to the rise of “Fin-Tech” changing the financial marketplace, new regulatory conditions due to the Dodd-Frank Act, or other changes to this market since the housing crisis. Appendix B analyzes the statistical significance and variance of entity-year FEs, and how these vary across years in the sample.

Figure 3.2: Relative Explanatory Power of Entity and County FEs Annually

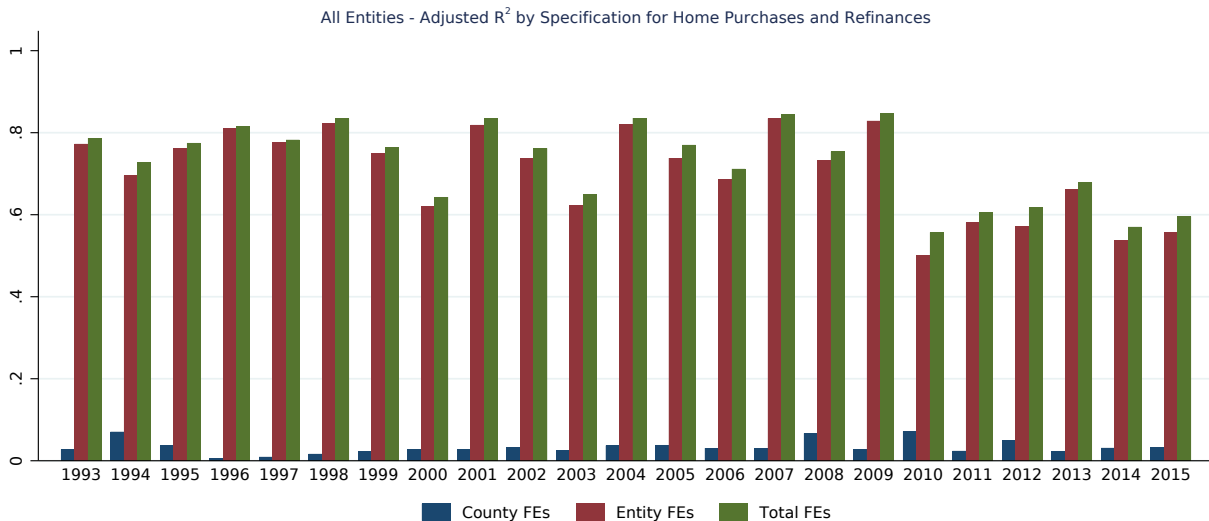


Table 3.3 presents these regressions at the more granular entity-tract-year level. Year-on-year changes in lending are substantially noisier at the this level of disaggregation, as there are over 60,000 tracts in the US, so this specification risks substantial over-fitting of local demand effects. Furthermore, tract level regressions feature more changes involving “entry” or “exit” from a tract that could be driven by chance rather than true supply effects, and which are handled

relatively poorly by equation (3.1), as even DHS changes do not perfectly handle changes from and to zero.<sup>18</sup> Despite this vast disparity in the number of fixed effects, entity-year effects in specification (6) relative to (4) still explain more than tract-year effects in specification (5) relative to (4). Thus, although supply effects explain a much smaller fraction of the total variance in these specifications, this is almost fully driven by substantially more variation at the tract level, and dramatically stronger controls for local demand. In practice, the estimated entity-year FEs derived from the tract-level specification are highly correlated with those estimated from the county-level specification. In the subsequent analysis, the  $\hat{\alpha}_{it}$  values from the county-level regression are used for these reasons, though results are qualitatively similar when constructing the shock using tract-level specifications.

Table 3.3: Tract-Level Adj  $R^2$  from FE Regressions on DHS Change in Lending

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	Observations
County Home Purchase	0.056	0.011	0.156	0.251	0.284	0.688	0.714	3,725,497
County Refinancing	0.175	0.006	0.132	0.334	0.392	0.748	0.791	4,425,979
Tract Home Purchase	0.019	0.046	0.142	0.210	0.259	0.353	0.399	28,681,040
Tract Refinancing	0.091	0.027	0.149	0.267	0.351	0.428	0.503	35,446,660
Year FE	Yes	No	No	Yes	Yes	Yes	Yes	
Areal FE	No	Yes	No	Yes	Yes	Yes	Yes	
Entity FE	No	No	Yes	Yes	Yes	Yes	Yes	
Areal-Year FE	No	No	No	No	Yes	No	Yes	
Entity-Year FE	No	No	No	No	No	Yes	Yes	

Figure 3.3 depicts the broad patterns captured by this methodology. The first figure is the supply shock, or entity-year FE, calculated for each year of the crisis, weighted by the previous year market share of the entity in each county, averaged across the four crisis years, and normalized to have weighted mean zero. The second figure is the corresponding demand shock, or county-year FE. The two maps are strikingly different, yet contain some intuitive features. In both, Florida is generally red, suggesting both that demand fell in Florida and that banks operating in Florida generally cut lending nationally. This matches closely with Florida's central role in the sub-prime lending crisis. By contrast, New York, New Jersey, and coastal California have

<sup>18</sup>In addition, lots of entry and exit is also problematic for weighting the specification properly, as previous period lending is used as a weight, and is relatively more noise than signal in the tract-level specification.

negative credit supply shocks, but positive demand shocks. This follows from these areas being largely served by large, national banks that were hit hardest by the collapse of financial markets and the deteriorating quality of mortgage backed securities. These lenders were thus the most likely to rein in their lending nationally despite relatively strong demand conditions in many of their primary lending markets.

### 3.1.3. THE CREDIT SUPPLY SHOCK: VALIDATION

One crucial concern with using equation (3.1) to estimate credit supply shocks is that entity-year fixed effects may capture an element of demand if demand varies at the entity-year level rather than the county-year level. This could be the case if certain types of lenders lend to different types of borrowers, and changes in borrower types' demand for credit is heterogeneous, forcing these lenders to adjust their lending to new borrower classes in which they may have less expertise.

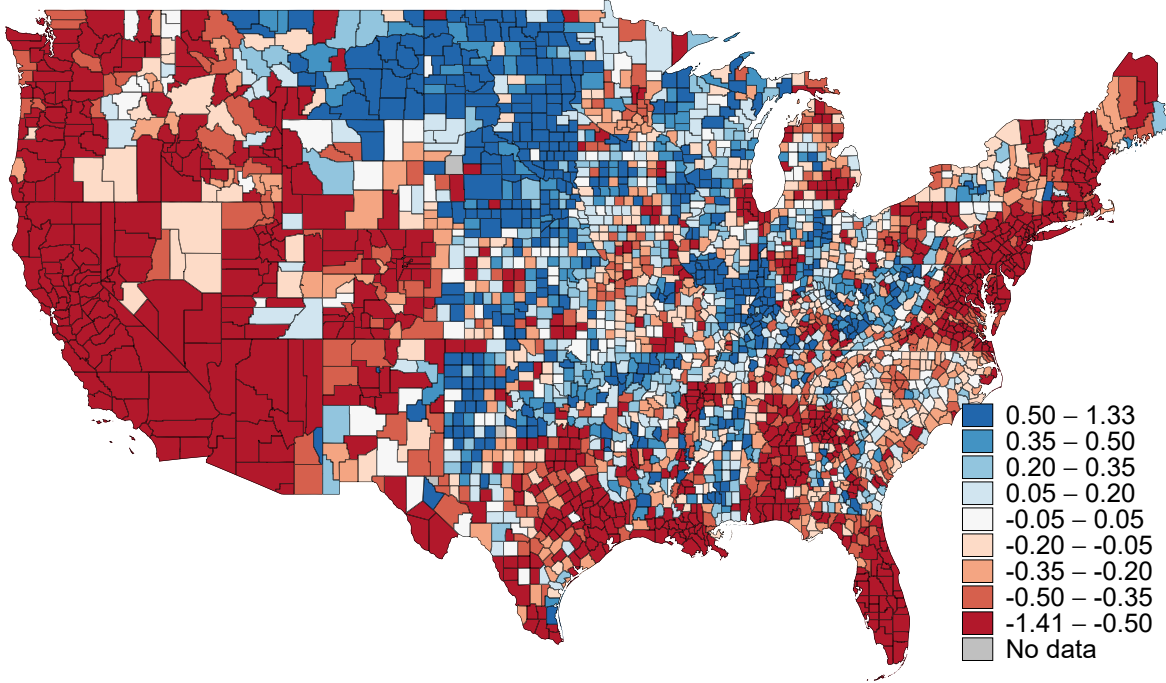
One way to test this is to run a specification similar to equation (3.1), but with changes in borrower characteristics as an outcome, rather than changes in lending. Though HMDA provides only limited data about borrower characteristics, and in particular does not provide information such as credit score, it does provide two variables that reasonably measure borrower creditworthiness: borrower income and borrower loan-to-income ratio (LTI). In constructing the credit standards shock, I demonstrate that these variables are significant predictors of acceptance or rejection of a loan application. Furthermore, median LTI is highly correlated at the county level with similar measures such as the median debt-to-income ratio (DTI) used in Mian et al. (2013).

In table 3.4 I run the following specifications to understand whether entity-year fixed effects have any explanatory power for changes in borrower composition:

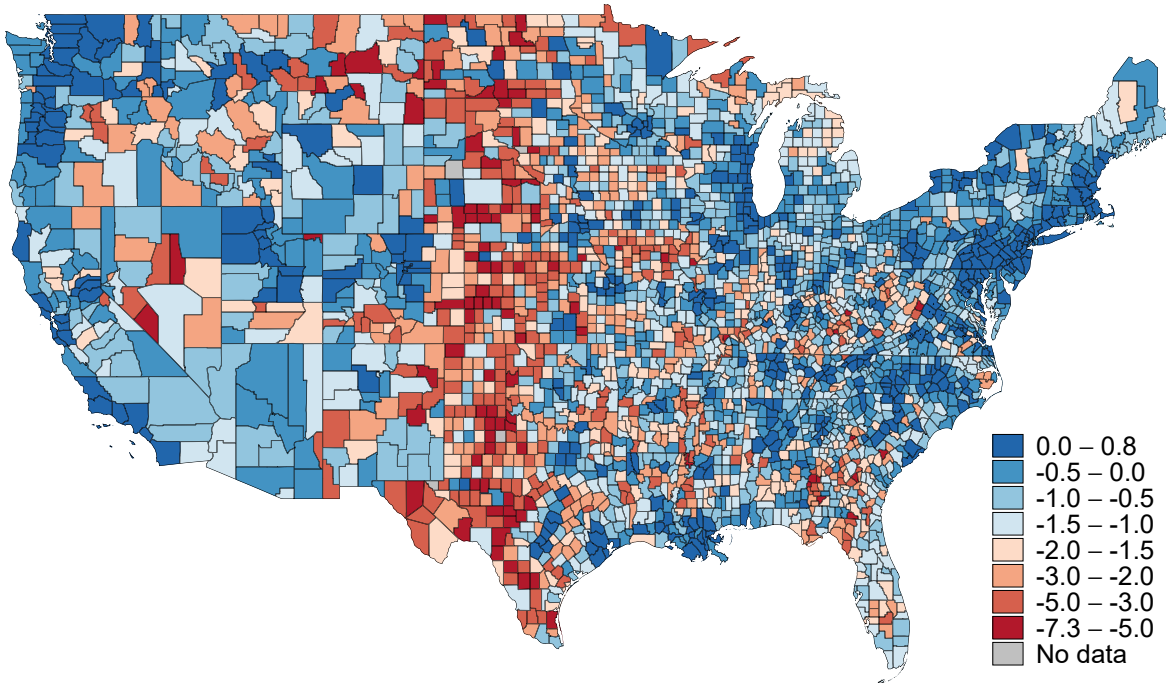
$$\ln \text{MedianIncomeChange}_{ijt} = \alpha_{it} + \beta_{jt} + \epsilon_{ijt}$$

Figure 3.3: County-Level Crisis Supply and Demand Shocks

Standardized Supply Shock 2007-2009 By County



Standardized Demand Shock 2007-2009 By County



Note: Blue indicates more positive fixed effects, meaning a positive shock relative to the rest of the country. Effects are constructed using the county-year-entity specification.

$$\ln \text{MedianLTIChange}_{ijt} = \alpha_{it} + \beta_{jt} + \epsilon_{ijt}$$

Table 3.4 shows reason for concern - entity-year FEs explain a modest portion of borrower

Table 3.4: Adj  $R^2$  from FE Regressions on DHS Change in Median Borrower Characteristics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	Observations
Income	0.015	0.001	0.025	0.036	0.060	0.243	0.268	3,535,422
LTI	0.032	0.001	0.031	0.060	0.084	0.300	0.324	3,535,424
Year FE	Yes	No	No	Yes	Yes	Yes	Yes	
County FE	No	Yes	No	Yes	Yes	Yes	Yes	
Entity FE	No	No	Yes	Yes	Yes	Yes	Yes	
County-Year FE	No	No	No	No	Yes	No	Yes	
Entity-Year FE	No	No	No	No	No	Yes	Yes	

characteristics. However, in order for this to be indicative of demand-effects in our credit supply shock, it must be that the entity-year fixed effects from these borrower specifications are correlated with the entity-year fixed effects from the lending specification. The intuition here is simple: changes in borrower composition at the entity-year level can be caused by two primary things: changes in entity lending strategy, and changes in available borrowers. The latter is the primary concern, but only reflects a change in demand if new borrowers have different aggregate levels of credit demand than old borrowers. If this were the case, we'd expect positive entity-year FEs from the borrower specification, indicative of either higher income or higher LTI amongst borrowers, to be correlated (positively for income, negatively for LTI) with entity-year FEs from the lending regression, because changes in borrower composition would imply substantial entity-year level changes in the level of credit demand. Table 3.5 shows that these correlations are both only slightly positive and insignificant over the sample period, with and without fixed effects.

An alternative method for confirming that entity-year FEs capture credit supply is to demonstrate that they are correlated with other variables that generally reflect entity health. For entities that are publicly traded, I use their RSSD ID and the Federal Reserve Bank of New York's

Table 3.5: FE Correlation: Lending and Borrower Characteristics 2005-2015

	(1)	(2)	(3)	(4)	(5)
Entity-Year FE in Income Regression	0.361 (0.624)			0.963 (0.652)	
Entity-Year FE in LTI Regression		0.231 (0.354)			0.117 (0.270)
Adj $R^2$	-0.000	0.003	0.219	0.252	0.247
Adj Within $R^2$	-0.000	0.003	0.000	0.007	-0.000
Year FE	No	No	Yes	Yes	Yes
Entity FE	No	No	Yes	Yes	Yes
Weighted	Yes	Yes	Yes	Yes	Yes
Observations	859	859	932	852	852

Standard errors in parentheses

Specifications are at the Entity-Year level, and are clustered by year

Specifications subset to lenders who are in the top 50 in any year of the sample

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

RSSD to Permco match to match entity-year FEs to compare entity-year FEs with stock performance, with the expectation of a strong positive correlation. However, one might expect this correlation even if entity-year FEs captured demand - declining mortgage credit demand in a public lender's lending markets could be seen as a sign of weakness by investors, and lead to a similar decline in stock performance. An alternative measure that more closely reflects supply is the lender's pre-crisis short term debt reliance. During the crisis, short-term debt markets weakened substantially, and entities with a strong reliance on this type of funding for liquidity would be more likely to cut their mortgage lending.

Table 3.6 shows a positive and significant correlation between entity-year FEs and stock returns, with better results during the crisis than the recovery. This could be either because strong correlations during the crisis are easily driven by a number of entities declining substantially, both in their lending behavior and in their stock value, or because entity-year FEs are less well estimated during the recovery, as noted in figure 3.2. Pre-crisis net short term debt, defined as net fed funds sold and securities purchased over total debt, is strongly negatively correlated with entity-year FEs during the crisis. Finally, pre-crisis tier 1 capital ratio is only very weakly pos-

itively correlated with entity-year FEs.<sup>19</sup> This weak correlation is not surprising - tier 1 capital ratios vary significantly with bank size and strategy, and are only a weak proxy of bank health in the cross-section.<sup>20</sup>

Table 3.6: Correlation between Entity-Year FEs and entity health measures

	Crisis	Recovery	Crisis	Crisis
Crisis Stock Return	0.542*** (0.047)			
Recovery Stock Return		0.163*** (0.036)		
Pre-Crisis Net Short Term Debt			-5.042*** (0.873)	
Pre-Crisis Tier 1 Capital Ratio				0.105 (1.386)
Adj $R^2$	0.240	0.055	0.067	-0.002
Lend Weight	Yes	Yes	Yes	Yes
Observations	416	332	451	403

Standard errors in parentheses

Pre-crisis net short term debt is defined as the fraction of net fed funds sold and securities purchased over total debt including deposits in 2006.

All variables winsorised at 1% and 99% levels

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

### 3.2. THE CREDIT STANDARDS SHOCK

The credit supply shock constructed in the previous section leverages national changes in entity mortgage lending *volumes* to estimate credit supply. An alternative approach is to consider changes in the credit *standards* of mortgage lenders across time within Cook County, which contains the city of Chicago. Specifically, here I leverage the fact that HMDA contains both accepted and rejected loan applications, and their characteristics, to discern which banks within Chicago are becoming differentially more stringent towards their borrowers. Though

<sup>19</sup>Tier 1 capital ratios are obtained from SNL, which sources these from multiple regulatory forms depending on the entity type.

<sup>20</sup>These regressions are run cross sectionally, using average values across years, because of potential heterogeneity in the precise years in which entities experience declines in lending or stock value during the crisis. The results are similar when run as a clustered panel regression.

both shocks ultimately reflect the same lender incentives, they differ in that the credit supply shock controls for demand through areal unit-year FEs and ex-post tests, a more macro approach, while the credit standards shock uses explicit information on the micro characteristics of potential borrowers available to the lender.

The primary concern with using rejected loans is one of sample selection - loan rejections do not represent the universe of individuals who consider applying for a loan. Indeed, it's possible that potential borrowers meet with a lender and are discouraged from submitting a formal application, or that borrowers who would like to purchase or refinance a home don't attempt to do so because they expect to be rejected. Figure 3.4 shows that through the crisis and recovery, there is a substantial dropoff in both applications and rejections, consistent with this kind of informal rejection potentially becoming more common during the crisis, when borrowers' characteristics weakened. However, credit standards can still be roughly estimated using the characteristics of the population of borrowers that do end up applying, so long as enough borrowers with diverse characteristics are accepted and rejected. This implicitly assumes that borrowers not in the data would either have been certainly rejected had they tried to apply, or did not want to apply; i.e. that the sample contains enough borrowers that want a loan and are at least marginally acceptable candidates.

HMDA provides some information on the reason that loans were denied, as shown in figure 3.5. However, these reasons, such as "insufficient collateral" and "credit application incomplete" are qualitative rather than quantitative, and lenders are not required to maintain the same standards when making these distinctions across years or even across loan applications within year. As a result, this data provides little information about changing standards during the crisis and recovery, or about relatively credit stringency across entities.

HMDA also provides, both directly and indirectly, a number of quantitative borrower characteristics that lenders may use when evaluating loan applications because of their effect on the loan's riskiness. Specifically, I consider inflation-adjusted applicant income, LTI, lien status,

Figure 3.4: Cook County Number of Loan Applications and Rejections

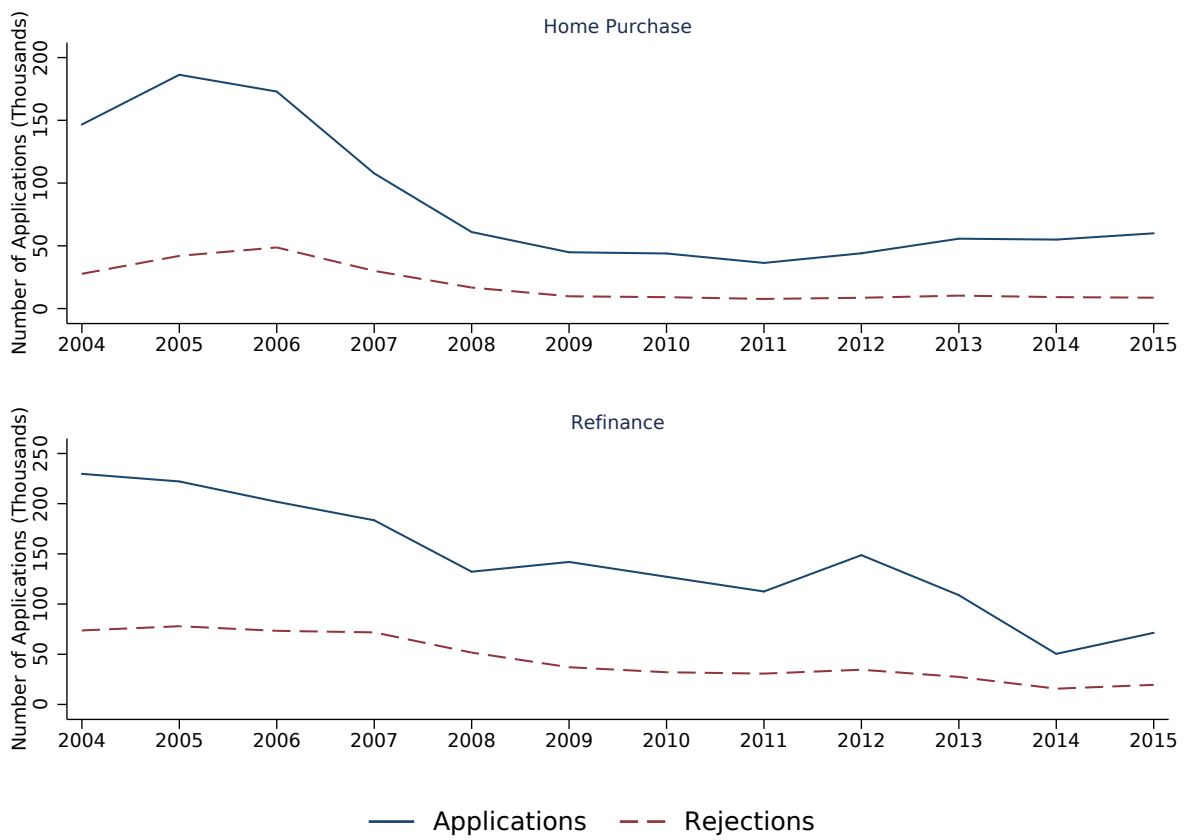
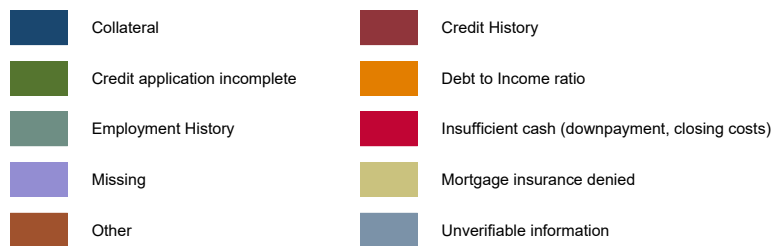
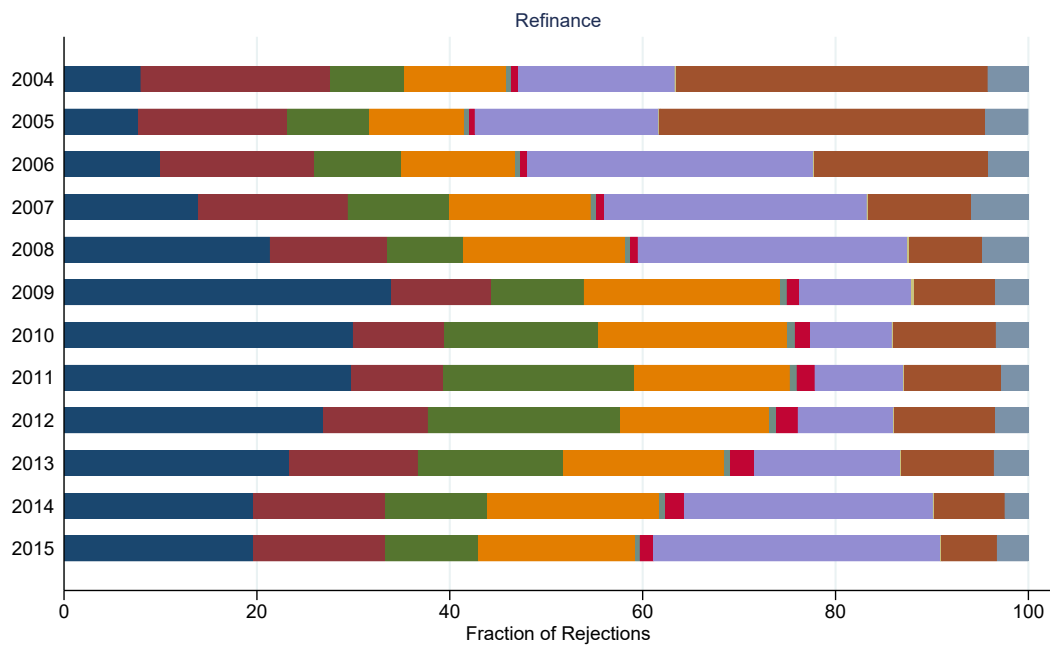
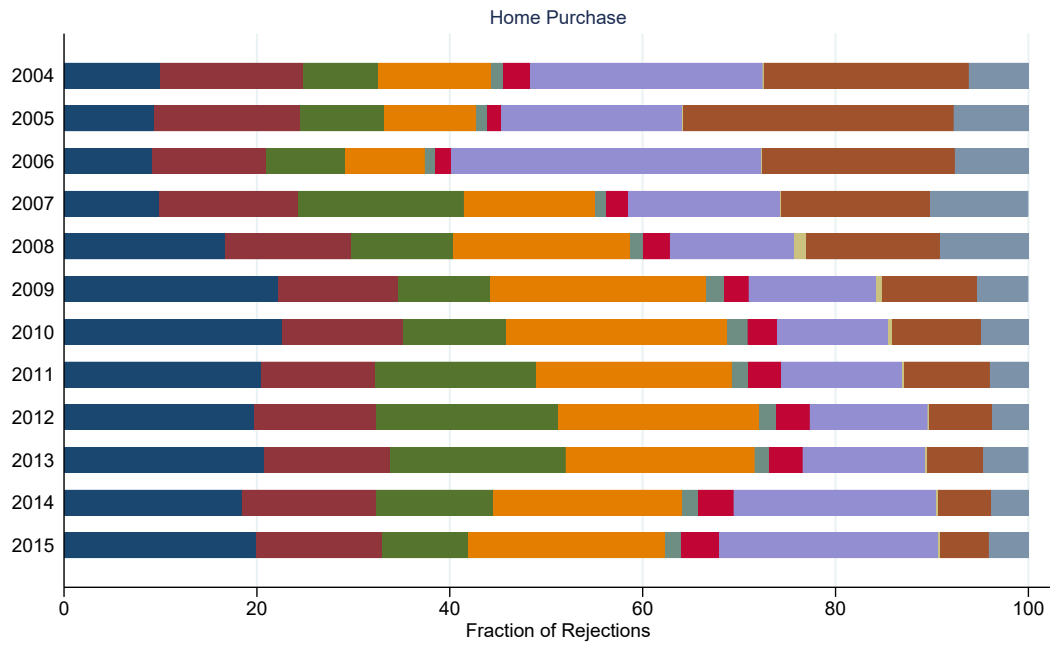


Figure 3.5: Cook County Denials by Primary Reason



jumbo status, applicant and co-applicant demographics, owner-occupier status, loan preapproval, and the size and geographic extent of the lender.<sup>21</sup>

In table 3.7, I use loan-level data to predict the binary rejection variable using these characteristics in Cook County Illinois during the pre-crisis period (2004-2006) using both a linear probability model and a logit model. I consider variants of the specification either including or omitting entity fixed effects, as well as using all applications or omitting rejections for which qualitative reasons were given.<sup>22</sup> The borrower characteristics results are relatively consistent and intuitive across the specifications, with higher applicant income, lower LTI, conforming loans, and first liens all implying a higher likelihood of acceptance. The dummy that the home will be owner-occupied has no effect in the logit specifications, and conflicting effects in the linear model, though applications for which this information was not provided are more likely to be rejected, suggestive of incomplete information provided by the applicant.

In addition to borrower characteristics, lenders are classified as as Community (Small) Local, Large or Mid Local, Mid Non-Local, Small Non-Local, and Large Non-Local. This includes an asset threshold, defined by regulatory reporting thresholds (globally systemically important (GSIB), Mid-Size, and Community), and a geographic dispersion threshold, defined by a Herfindahl index (Local, Non-Local). The size cutoffs are \$50 billion in assets for GSIB status, and \$10 billion in assets for Mid-Size status. The dispersion cutoff is set at a Herfindahl index of 0.3 at the state level, so that lenders that do the majority of their lending in a few states or less are classified as local.<sup>23</sup> Larger size and non-local lender status are absorbing classifications, meaning that lenders are classified as being larger or more geographically dispersed if

---

<sup>21</sup>Applicant and co-applicant demographics, such as race and ethnicity, are included to control for unobservable characteristics, and are not a considered primary factor in determining rejection. Indeed, Congress' motivation in 1975 for signing HMDA into law was to insure that discrimination via loan rejection on the basis of these characteristics was not occurring.

<sup>22</sup>These reasons are in figure 3.5. Loan applications are kept in this version of the specification if they (i) Are accepted (ii) Are rejected for a reason listed as "Other" (iii) Are rejected and no reason is reported ("Unknown Reason").

<sup>23</sup>0.3 was selected based on the distribution of HHI values seen in the data. The vast majority of loan-weighted lenders exist below 0.2 and above 0.9 in any given year, so 0.3 was selected as a value that is between these two and consistently yields the same classifications across loan types.

they achieve any of the cutoffs in a single year. For example, Large Non-Local is the baseline in table 3.7, and contains all entities that are large enough to be considered GSIBs in at least one year of operation, and which have an Herfindahl index of geographic state concentration below the 0.3 threshold in at least one year. Large and Mid-Size Local lenders are combined into a single classification for clarity because these lenders are rare.

Relative to large national lenders - for which the dummy variable is omitted - local lenders of all sizes are more likely to accept an application. Holding geographic extent constant, results vary on whether smaller entities are more or less likely to reject. These results fit with a hypothesis that local lenders are better at understanding intangibles about borrowers' circumstances, and may be better able to assess when a given borrower with ostensibly poor fundamentals is actually not risky.

Using the estimated model coefficients from this pre-crisis period, I predict expected lending during the crisis and recovery (2007-2010, 2011-2015) out-of-sample. Since the logit and linear probability model give probabilities of rejection, this entails multiplying the probability of acceptance with the loan amount, and aggregating within tract-year.<sup>24</sup> These values are compared against the true level of accepted loans, with the difference defined as the "credit shortfall." Mathematically, this measure is defined in equation (3.3), where  $L_{jt}$  is the set of loan applications in location  $j$  at time  $t$ , and the estimated coefficients in table 3.7 are used to construct  $P(\text{RejectPrecrisis})$ .

$$CS_{jt} = \sum_{l \in L_{jt}} ((1 - \text{Reject}_l) * \text{Value}_l - (1 - P(\text{RejectPrecrisis})_l) * \text{Value}_l) \quad (3.3)$$

This approach is simple and incomplete - it does not control for how acceptance criteria vary across lenders, aside from a "level effect" captured by the entity fixed effects in some specifications, and by the more aggregate lender type in other specifications. Instead, it captures aggregate changes in lending standards by all lenders before and after the crisis, and how they

---

<sup>24</sup>Linear probability model probabilities outside the [0,1] interval are truncated.

Table 3.7: Simple Rejection Prediction - Cook County

	(1)	(2)	(3)	(4)	(5)
Log applicant income	-0.016*** (0.002)	-0.046*** (0.002)	-0.000 (0.001)	-0.475*** (0.014)	-0.426*** (0.018)
Loan to income ratio	0.027*** (0.001)	0.024*** (0.001)	0.011*** (0.001)	0.067*** (0.006)	0.030*** (0.008)
Jumbo loan	0.018*** (0.003)	0.019*** (0.004)	0.006* (0.002)	0.071** (0.024)	0.056+ (0.029)
Subordinate lein	0.111*** (0.003)	0.093*** (0.004)	0.040*** (0.002)	0.358*** (0.022)	0.022 (0.027)
No owner-occupied info	0.332*** (0.010)	0.357*** (0.010)	0.327*** (0.011)	1.643*** (0.053)	1.502*** (0.062)
Not owner-occupied	0.048*** (0.003)	-0.009** (0.003)	0.023*** (0.002)	0.103*** (0.018)	-0.019 (0.021)
Small local lender		-0.168*** (0.003)		-1.266*** (0.021)	-0.888*** (0.028)
Large/Mid-Size local lender		-0.099*** (0.004)		-0.606*** (0.028)	-1.413*** (0.056)
Mid-size non-local lender		0.026*** (0.003)		0.133*** (0.013)	0.073*** (0.020)
Small non-local lender		0.032*** (0.002)		0.341*** (0.009)	1.111*** (0.012)
Constant				1.082*** (0.078)	-0.081 (0.096)
$R^2$	0.231	0.105	0.364		
Pseudo $R^2$				0.041	0.080
Observations	610870	610870	502717	610870	502717
Loan Purpose	Refinance	Refinance	Refinance	Refinance	Refinance
Entity FE	Yes	No	Yes	No	No
Demographic FEs	Yes	Yes	Yes	No	No
Sample	Full	Full	Other Only	Full	Other Only
Reg Type	Linear	Linear	Linear	Logit	Logit

Standard errors in parentheses

All specifications are clustered at the Tract level

+  $p < 0.10$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

differentially affect tracts depending on their borrower composition.

To capture lender-level changes in credit strictness, I consider a second approach, which estimates pre-crisis credit shortfall at the entity level using all loan applications in all census tracts for each Chicago lender. To define Chicago lenders, I take the top 50 lenders in Chicago from 2004-2006, as well as five aggregates of the remaining lenders by lenders' geographic and size types (Community Local, Large or Mid Local, Mid Non-Local, Small Non-Local, and Large Non-Local). For each lender, I predict their application rejections in all tracts, including outside of Chicago, using a variety of specifications mirroring table 3.7. I then construct a second measure of credit shortfall (CS2) at the  $(j, t)$  level using an identical method to equation (3.3) but at the lender  $i$  level, then sum across all  $N$  lenders. Unlike the previous approach, this approach leverages lender-specific coefficients when predicting rejection, and incorporates lenders potentially tightening or loosening standards along different borrower characteristic margins. Mathematically, this equation is defined in equation (3.4).

$$CS2_{jt} = \sum_{i=1}^N \sum_{l \in L_{ijt}} ((1 - \text{Reject}_l) * \text{Value}_l - (1 - P(\text{RejectPreCrisis}))_l) * \text{Value}_l \quad (3.4)$$

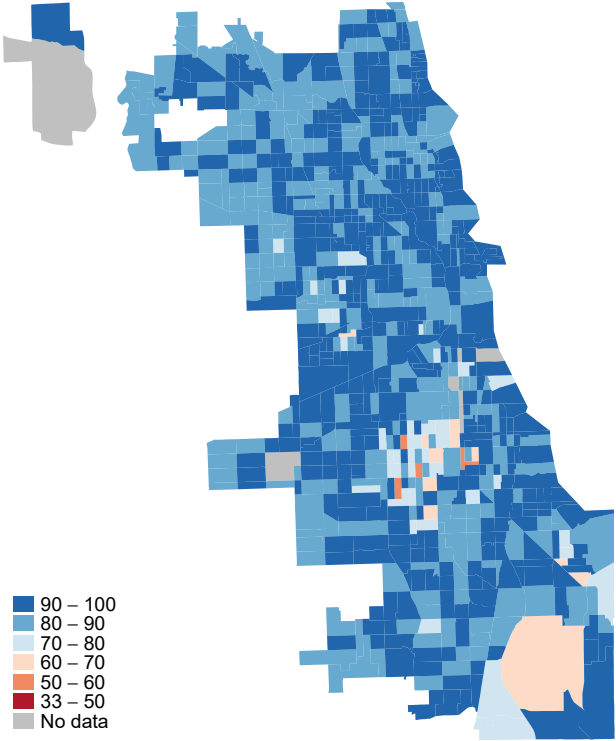
One shortcoming of this approach is that it handles entry and exit poorly. Lenders who lend in Chicago after 2006, but not from 2004-2006, are not in the sample. To correct for this, I construct a variable that measures the percentage of lending by continuing lenders in tract  $j$  at time  $t$ , defined according to equation (3.5).

$$\text{PercContinuing}_{jt} = 100\% * \frac{\text{Lending By 2004-2006 Chicago Lenders}_{jt}}{\text{Lending}_{jt}} \quad (3.5)$$

This variable is both interacted with CS2 to construct the primary credit standards shock and included as an independent covariate alongside the interaction in the subsequent regressions. The interaction captures the extent to which credit standards of these pre-crisis lenders ultimately affects local borrowers. Specifically, if the composition of lenders changes dramatically

during the crisis, then the standards of these pre-crisis lenders is less important and this interaction shrinks this variable towards zero. Independently, PercContinuing's sign is uncertain; churn of mortgage lenders could be a good or bad sign, since new lenders could imply either (i) a relatively booming housing market, with existing lenders failing to meet borrower demand, or (ii) a substantial decline in the health of lenders that cater to a given tract, that may only be imperfectly filled in by newcomers to Chicago's market. In practice, this correction is minor for the vast majority of Chicago, and PercContinuing seems to be mostly a positive indicator. Figure 3.6 shows the percentage of lending done by pre-crisis lenders in 2010, i.e. lenders that survived the whole crisis, and for much of Chicago the values are above 80%. The exception to this is primarily in Back-of-the-Yards and Englewood, two neighborhoods that are relatively poor and featured a wide variety of subprime lenders in the pre-crisis period that failed during the crisis.

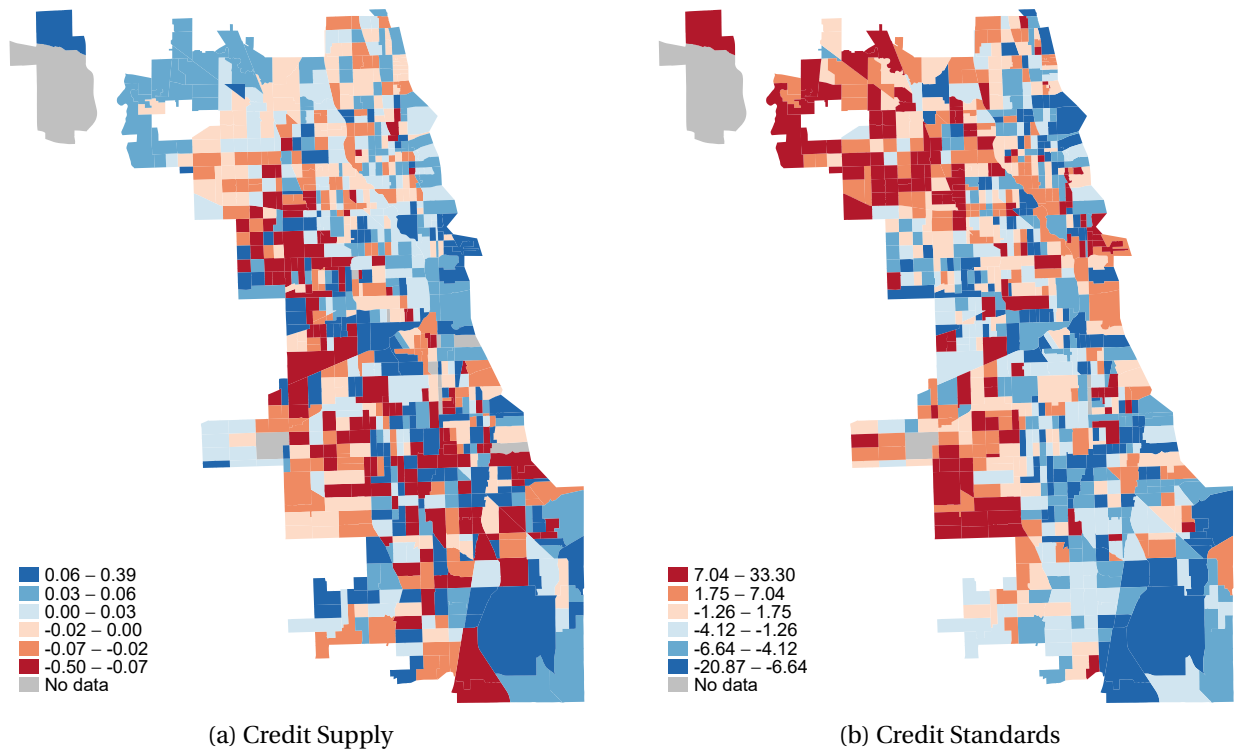
Figure 3.6: Percentage of Lending in 2010 done by 2004-2006 Chicago Lenders



### 3.3. APPLICATION TO CHICAGO

Using the methodology outlined in the previous sections, I construct tract-year level shocks for the city of Chicago. For the credit supply shock, I estimate the credit supply shock using national entity-county-year level data, excluding Cook County, convert it to the tract-year level, and standardize, weight, and winsorise the shock at the tract-year level. Within the city of Chicago, I residualize the shock with respect to tract-level fixed effects to remove trends in lending growth within tracts over the sample periods. Figure 3.7 depicts the resulting credit supply and credit standards shocks, averaging over the crisis period, such that redder tracts reflect those with lower credit supply or tighter credit standards. The credit supply shock displays substantial heterogeneity across the city, and does not obviously match Chicago's demographics, which are highly spatially correlated. The income dimension of this spatial correlation is depicted in figure 2.5. Note that the magnitude of effects is small relative to the normalization largely due to mean-reversion effects associated with averaging across years for these figures. The credit standards shock, with negative values (blue) indicating areas where credit standards tightened relatively less for identical borrowers, are more spatially correlated, and are a closer match to the demographic distribution of Chicago.

Figure 3.7: Refinancing Credit Shock Residuals 2007-2010



## 4. EMPIRICAL RESULTS

In this section, I present the estimating equations and empirical results used to understand the effect of these credit forces on housing investment, and how this shock allows for the estimation of neighborhood spillover effects. I begin by formalizing the notion of a tract’s neighbors using the spatial weights matrix  $\mathbf{G}$  and describe how this is used in the estimating equations to estimate spillovers, with a focus on econometric concerns. Using this framework, I estimate the model in stages. I begin by estimating the within-year effects of credit supply shocks on housing investment, future building code violations, and future housing abandonment within a census tract. I then present the effects of credit standards, using both the simple aggregate and the lender-level approaches. I find that the effect of both credit supply and credit standards are significant even controlling for one another. I then use the spatial weights matrix to perform (i) OLS regressions that demonstrate correlations in investment and code violations across space and (ii) instrumental variables regressions using neighboring tracts’ credit supply shocks as an instrument for neighbors’ investment, both with and without temporal lags. The latter is the primary result, leveraging the fact that refinancing credit shocks estimated using the rest of the US are strong predictors of housing investment at the tract level, and can be used in an IV approach to alleviate the reflection problem concerns that generally plague the OLS estimation of spillover effects.

### 4.1. NOTATION AND ECONOMETRICS

To estimate spillovers in investment, I define the spatial weights matrix  $\mathbf{G}$ .  $\mathbf{G}$  is  $N \times N$  matrix, where  $N$  is the number of locations: census tracts in this analysis. Elements in  $\mathbf{G}$  are defined by:

$$g_{ii} = 0 \quad \forall i \quad (4.1)$$

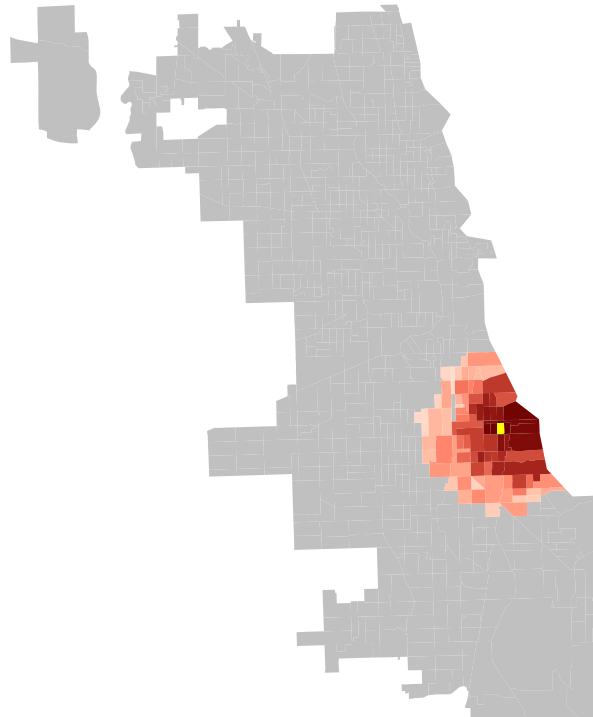
$$g_{ij} = \frac{I\{d_{ij} \leq \bar{d}\}}{d_{ij}^2} s_j \gamma_i \quad \forall j \neq i \quad (4.2)$$

$$\sum_i g_{ij} = 1 \tag{4.3}$$

$s_j$  is the size weighting attributed to neighbor  $j$ , 2010 Census housing units in this analysis, and  $d_{ij}$  is the distance between the two locations in miles,  $\bar{d}$  is the cutoff beyond which weights are assumed to be zero, and  $\gamma_i$  is a scaling factor selected so that each row of  $\mathbf{G}$  sums to one. Figure 4.1 depicts these spatial weights for a tract in Hyde Park, Chicago (green) with the assumed baseline range  $\bar{d} = 3$  miles. Empirical results presented in the subsequent sections are robust to alternative forms of spatial weights, such as rook and queen weights, and different distance cutoffs  $\bar{d}$ . For clarity of exposition, I focus on weighted inverse squared distance weights which most closely match previous empirical findings, such as Esteban et al. (2010), that housing externalities decline with distance.

**Figure 4.1: Distance Weighting Example**

Inverse Distance Weights Relative to Tract 17031410700 (yellow)



The spatial weights matrix  $\mathbf{G}$  is used to construct spatial lags (henceforth “Slag”) of data vectors,

according to the following matrix multiplication:

$$y_{n(i)} := \sum_j g_{ij} y_j \quad \text{Summation Form}$$

$$\vec{y}_{n(\cdot)} := \mathbf{G}\vec{y} \quad \text{Matrix Form}$$

Where subscript  $n(i)$  refers to the neighbors of location  $i$ , and  $\vec{y}_{n(\cdot)}$  refers to a vector of  $y_{n(i)}$  values.

Contemporaneous OLS regressions using spatial lags are subject to reflection problem concerns noted by Manski (1993), whereby neighbors' contemporaneous values of the outcome variable in the construction of the spatial lag are endogenous by construction if the model is true, since their value depends on the outcome. This can be seen from the baseline spatial lag model in Anselin (1988), where the matrix  $\mathbf{GG}$  will not have a zero diagonal:

$$\vec{y} = \rho\mathbf{G}\vec{y} + \mathbf{X}\beta + \vec{\epsilon} \quad \text{Model with spatial weights matrix } \mathbf{G}$$

$$\vec{y} = \rho\mathbf{G}(\rho\mathbf{G}\vec{y} + \mathbf{X}\beta + \vec{\epsilon}) + \mathbf{X}\beta + \vec{\epsilon} \quad \text{Substitute model into RHS}$$

$$\vec{y} = \rho^2\mathbf{GG}\vec{y} + (\rho\mathbf{G} + \mathbf{I})(\mathbf{X}\beta + \vec{\epsilon}) \quad \text{Simplify}$$

$\mathbf{GG}\vec{y}$  is the second spatial lag of  $\vec{y}$ . Unlike in models with time dependence, for which the second temporal lag contains information from two periods prior, the second spatial lag contains information from neighbors-of-neighbors in the current period. This leads to a fundamental endogeneity problem because one of each tract's neighbors' neighbors is itself. This manifests mathematically as a non-zero diagonal of the  $\mathbf{GG}$  matrix, meaning that  $y_i$  is present on both sides of the above equation, in  $y_i$  itself on the left, and in the corresponding element of  $\mathbf{GG}\vec{y}$  on the right.

Historically, there have been multiple approaches for handling this reflection problem, which are outlined clearly in Anselin (2009). One is to construct the reduced form, solving the initial

equation for  $\vec{y}$  and thus eliminating  $y$  from the right-hand side of the equation entirely. Some simple matrix algebra provides the following specification, which can be estimated with maximum likelihood (ML) or generalized method of moments (GMM).

$$\vec{y} = (\mathbf{I} - \rho\mathbf{G})^{-1}(\mathbf{X}\beta + \vec{\epsilon})$$

This approach relies heavily on the accuracy of the initial model specification because of the reliance on the matrix inversion to fully remove  $\vec{y}$  from the right side. As a result, it is sensitive to the accuracy of spatial weights matrix  $\mathbf{G}$ , which cannot be easily tested.

A second approach is to find an appropriate instrument  $\mathbf{Z}$  for the endogenous variable  $\mathbf{G}\vec{y}$ . Kelejian and Robinson (1993) prove that 2SLS estimates  $\hat{\rho}$  and  $\hat{\beta}$  are consistent, and suggest the use of spatial lags of the exogenous variables -  $\mathbf{GX}$ ,  $\mathbf{GGX}$ ,  $\mathbf{GGGX}$ , etc. - as instruments.

A third, and somewhat unconventional, approach is to consider whether the data generating process may be such that an alternative model in which the spatial effects are lagged rather than contemporaneous may be true:

$$\vec{y}_t = \rho\mathbf{G}\vec{y}_{t-1} + \mathbf{X}_t\beta + \vec{\epsilon}_t$$

This solves the reflection problem if the new model assumption is correct, but estimating this equation directly depends very strongly on this assumption, which is impossible to prove explicitly. In this sense, the lag model is only appropriate if the problem lends itself to a lagged interpretation.

I employ the second and third approaches. First, I demonstrate that the results are robust to using lagged rather than contemporaneous neighbors' investment, controlling for two-way fixed effects. Second, I employ an instrumental variables approach, identifying the spillovers of housing investment across space by instrumenting neighbors' changes in investment using their credit shocks ( $\mathbf{GX}$ ), controlling for the tract's own credit shocks ( $\mathbf{X}$ ) and two-way fixed ef-

fects.

The baseline empirical model can be described by the following system of matrix equations.<sup>25</sup>

$$\text{Invest}_{n(j)t} = \text{FE}_{jt} + \mathbf{X}_{jt}\boldsymbol{\beta} + \mathbf{X}_{n(j)t}\boldsymbol{\mu} + \epsilon_{jt} \quad (4.4)$$

$$\text{Invest}_{jt} = \text{FE}_{jt} + \mathbf{X}_{jt}\boldsymbol{\beta} + \widehat{\text{Invest}}_{n(j)t}\boldsymbol{\rho} + \epsilon_{jt} \quad (4.5)$$

Where the notation follows:

$$\begin{aligned} y_{n(j)t} &= Gy_{jt} \\ \mathbf{X} &= \left[ \hat{\alpha}_{jt}, (\widehat{\text{CS2}}_{jt} * \text{FC}_{jt}), \text{FC}_{jt} \right] \\ \text{FE}_{jt} &= \tilde{\gamma}_j \tilde{\mathbf{1}}'_t + \tilde{\mathbf{1}}_j \tilde{\boldsymbol{\delta}}'_t \end{aligned}$$

Specifically,  $n(j)$  refer to weighted spatial averages of the neighbors of tract  $j$  defined by a spatial weight matrix  $G$ .  $\mathbf{X}$  comprises the credit supply variables.  $\hat{\alpha}_{jt}$  is the entity-year credit supply shock  $\hat{\alpha}_{it}$ , projected to the tract-year level according to equation (3.2),  $\widehat{\text{CS2}}_{jt}$  is the lender level of credit shortfall defined according to equation (3.4), and  $\text{FC}_{jt}$  is a control for the fraction of continuing pre-crisis lenders through the crisis and recovery.  $\text{FE}_{jt}$  comprises time  $t$  and tract  $i$  fixed effects.  $\epsilon_{jt}$  is an error term, which is always two-way clustered at the neighborhood-year and tract levels.<sup>26</sup>  $\mathbf{X}_{n(j)t}$ , the credit supply shock of a tract's neighbors, is the omitted instrumental variable. Fixed effects and  $\mathbf{X}_{jt}$ , the tract's own credit supply shock, are the primary controls. Table 4.1 summarizes the main variables used in estimation at the annual level, excluding spatial lags.

<sup>25</sup>In practice, when testing the first stage effects of credit shocks on investment, the estimation is done at the  $j$  level rather than the  $n(j)$  level to provide direct results about credit and investment independent of a spillover channel, and credit shocks are added to the model sequentially (i.e. not all of  $\mathbf{X}$  is included at once) for clarity of exposition. Results are qualitatively similar with both  $j$  and  $n(j)$  level variables included simultaneously or when  $\text{Invest}_{n(j)t}$  is actually used as an outcome rather than  $\text{Invest}_{jt}$ .

<sup>26</sup>91 neighborhoods of Chicago are assigned to tracts by first matching blocks to neighborhoods, and then assigning tracts to neighborhoods according to plurality of housing share.

Table 4.1: Summary Statistics

	mean	sd	p25	p50	p75
Housing Investment Per HH	221.22	187.57	96.96	165.03	284.70
Renovation Investment Per HH	172.12	176.19	53.79	116.77	229.35
Easy Permit Investment Per HH	45.71	27.88	25.93	41.38	60.36
Refinancing Credit Supply Shock	0.24	0.35	0.08	0.25	0.46
Home Purchase Credit Supply Shock	0.13	0.27	-0.02	0.16	0.34
Dollar Credit Shortfall Per HH	21.31	19.25	7.03	15.24	30.23
Perc Continuing Lenders since Pre-Crisis	89.89	9.59	85.33	90.99	98.83

## 4.2. THREATS TO IDENTIFICATION

Before presenting the results of the specifications presented in equation (4.4), it's worth considering explicit threats to this spillover identification strategy. Specifically, the primary concern is spatial correlation in the matrices comprising  $\mathbf{X}_{jt}$ , the spatial lags of which are used as the instrumental variables, which is driven by spatial correlation in lenders' lending shares across the city.

To understand this concern, I focus on the case of the projected credit supply shock  $\hat{\alpha}_{jt}$  specifically. Recall equation (3.2), which is used to convert the national entity-year shocks to tract-level shocks using the lagged entity-tract-year lending share. In matrix form, this equation is:

$$\hat{\alpha}_{j,t} = \rho_{ji,t-1} \hat{\alpha}_{i,t} \quad \forall t \in T$$

Where  $\hat{\alpha}_{j,t}$  is a  $J$  length vector at time  $t$ ,  $\rho_{ji,t-1}$  is a  $J \times I$  matrix of lending shares of lender  $i$  in tract  $j$  at time  $t - 1$ , such that the rows of  $\rho_{ji,t-1}$  sum to 1, and  $\hat{\alpha}_{i,t}$  is a  $I$  length vector of entity-year FEs for year  $t$ .

The tract-lender-year shares  $\rho_{ji,t-1}$  pose the primary threat to identification. As seen in appendix A.3, lending shares for individual lenders are spatially correlated because they tend to focus advertising and physical branches in specific parts of the city in which they have expertise. This direct spatial targeting, combined with lingering effects such as word-of-mouth, leads

to persistent spatial correlation in lending behaviors. Computing the Moran's I statistic for the spatial correlation in the top 25 lenders' shares in 2005 yields an average scaled statistic of 0.210, and a median of 0.182, statistically significant at the 1% level for all lenders, implying consistent positive spatial sorting. The construction of  $\hat{\alpha}_{j,t}$  relies on a matrix multiplication involving these spatially correlated lending matrices, resulting in credit shocks that are also spatially correlated with a significant scaled Moran's I statistic of 0.276 in the same year. However, many of these correlations are persistent through time. Controlling for two-way fixed effects that are present in all specifications, crucially tract fixed effects, this statistic declines to 0.071 in 2005. In the primary specifications presented in equation (4.4), own credit supply shocks are included as a control when using neighbors' credit supply as an instrument to address this residual spatial correlation - effectively using the part of neighbors' credit supply shock that is orthogonal to the tract's own shock as the true instrument. Therefore, although there is spatial correlation in lending, it's much weaker when tract-level fixed effects are included, and the main specifications seek to control for any residual correlation with the own shock control.

Though this exercise focuses on one specific year and instrument, it provides an example of how spatial correlation in lender shares is mitigated in the baseline specification. In the subsequent sections, I build up to estimating the model in equation (4.4) by (i) considering the first stage effects of credit on investment ignoring spatial effects, (ii) understanding the spatial correlations in investment, without instrumental variables, and finally (iii) estimating the IV to understand the strength of spillover effects, using refinancing credit shocks as instruments.

### 4.3. EFFECTS OF CREDIT SUPPLY AND STANDARDS WITHIN TRACTS

Table 4.2 presents the effect of credit supply shocks, constructed using data excluding Cook County, on different forms of housing investment using a non-spatially lagged form of the first

stage and omitting credit standards:

$$\text{Invest}_{jt} = \beta \hat{\alpha}_{jt} + \delta_t + \gamma_j + \epsilon_{jt}$$

Table 4.2: Housing Investment (Dollars per HH) on Standardized Credit Supply

	(1)	(2)	(3)	(4)	(5)	(6)
	HInvest	HInvest	Renov	Renov	Easy	Easy
Refinancing Credit Supply Shock	41.334*** (11.271)		48.816*** (11.177)		-5.783*** (2.026)	
Home Purchase Credit Supply Shock		12.217 (10.488)		14.065 (10.139)		-0.651 (1.731)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Tract FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	7235	6742	7235	6742	7235	6742

Standard errors in parentheses

Standard errors are two-way clustered at the Census Tract and Neighborhood-Year levels

Variables are winsorised at the 1 and 99th percentile to handle extreme outliers

Observations are weighted by 2010 Census housing units

Specifications use credit shock variables constructed without Cook County in the sample

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Specifications (1) and (2) show that a 1SD refinancing (home purchase) credit supply shock leads to a \$41 (\$12) increase in housing investment per household-year relative to a baseline of \$221. Specifications (3)-(6) decompose this into renovation and easy permit investment, showing that increases in renovations are even larger, but are partially offset for refinancing shocks by a decline in easy permit projects. Specifically, specifications (3) and (4) show that a 1SD refinancing (home purchase) credit supply shock leads to a \$49 (\$14, but insignificant) increase in renovation investment per household-year relative to a baseline of \$172. This suggests that refinancing credit may facilitate some substitution from minor to major projects by making homeowners more confident in their balance sheets. These specifications are robust to the inclusion of controls for time-varying tract quality measures, such as crime. Appendix C shows that adverse credit supply shocks also increase future building code violations and housing abandonment. Together, these findings demonstrate that credit supply both (i) helps homeowners undertake positive housing investment and (ii) helps constrained homeowners maintain prop-

erties that may otherwise severely depreciate.

Table 4.3 presents the effect of the basic credit standards measure (CS) defined in equation (3.3), on investment using the following specification:

$$\text{Invest}_{jt} = \beta_2 \widehat{\text{CS}}_{jt} + \delta_t + \gamma_j + \epsilon_{jt}$$

Table 4.3: Housing Investment (Dollars per HH) on Credit Shortfall

	(1)	(2)	(3)	(4)	(5)
Simple Dollar Credit Shortfall per HH	-0.197 (0.219)	0.040 (0.217)	-0.708*** (0.165)	-0.221 (0.203)	-0.611*** (0.158)
Refinancing Credit Supply Shock	41.220*** (11.248)	41.369*** (11.277)	35.313*** (10.966)	40.636*** (11.230)	35.221*** (10.970)
Year FE	Yes	Yes	Yes	Yes	Yes
Tract FE	Yes	Yes	Yes	Yes	Yes
Shock Regression Type	Linear	Linear	Linear	Logit	Logit
Shock Loan Subsetting	All	All	Other Only	All	Other Only
Shock Bank FE	Yes	No	Yes	No	No
Mean Credit Shortfall	3.0	2.9	19.4	2.7	18.0
SD Credit Shortfall	10.8	11.2	18.9	14.3	19.7
Observations	7235	7235	7235	7235	7235

Standard errors in parentheses

Standard errors are two-way clustered at the Census Tract and Neighborhood-Year levels

Variables are winsorised at the 1 and 99th percentile to handle extreme outliers

Observations are weighted by 2010 Census housing units

Variables transformed so that credit shock coefficients can be interpreted as:

A 1 dollar credit shortfall per HH leads to a \_ dollar change in housing investment

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Here we see a significant effect in specifications for which credit shortfall is determined using only rejections for which specific rejection reasons were unspecified, omitting potentially confounding loan rejections for which clear grounds for rejection were presented. Specification (3) for instance implies that a 1SD credit shortfall leads to a decline in housing investment by 13.38%, controlling for the refinancing credit supply shock. However, there are two important caveats. First, the result is not robust to different ways of constructing the credit shortfall, particularly in specifications (1), (2), and (4). Second, this correlation is identified off of the differ-

ential extent to which borrowers in different tracts are affected by changing standards relative to the pre-crisis period. It may be that these underlying differences are why investment falls; the same things that make banks not want to lend to potential borrowers, such as illiquidity, are the same reasons people may not want to invest in their homes. This is not fully corrected by the inclusion of tract fixed effects, since this concern is at the  $(j, t)$  level - the concern is both that the composition of tracts differs ( $j$  level), but also that this differing composition can lead to different exposure to changes through the crisis and recovery ( $t$  level). However, this remains suggestive that changes in lending standards may have unequal effects on different parts of the city depending on how close borrowers in a given census tract are to the acceptance margin.

Table 4.4 improves on table 4.3 by instead using credit standards estimated at the lender level according to equation (3.4), and controlling for the continuity of lenders through time, reflecting the fact that if existing lenders tighten standards, there is scope for entry into the tract's lending market. Specifically, the specification is now:

$$\text{Invest}_{jt} = \beta_2 \widehat{\text{CS}}_{jt} * \text{FC}_{jt} + \omega \text{FC}_{jt} + \delta_t + \gamma_j + \epsilon_{jt}$$

Specifications (1)-(4) give consistent results, and suggest an effect of a 1SD credit shortfall on housing investment of between 5.96% and 14.57%, controlling for credit supply. From table 4.2 and table 4.4, I conclude that, ignoring spatial effects, national refinancing credit supply shocks significantly affect households' housing investment. In the next section, I incorporate spatial effects, and show that there are clear spatial correlations in investment that are driven in part by credit supply.

Table 4.4: Effects of Credit Shortfall on Housing Investment

	(1)	(2)	(3)	(4)
Dollar Credit Shortfall per HH x Frac Continuing Lenders	-0.602** (0.238)	-0.767*** (0.165)	-0.546** (0.231)	-0.715*** (0.161)
Perc Continuing Lenders since Pre-Crisis	-0.266 (0.294)	-0.037 (0.291)	-0.236 (0.294)	-0.045 (0.291)
Refinancing Credit Supply Shock	38.945*** (11.068)	32.935*** (10.854)	38.705*** (11.066)	33.375*** (10.870)
Year FE	Yes	Yes	Yes	Yes
Tract FE	Yes	Yes	Yes	Yes
Shock Regression Type	Linear	Linear	Logit	Logit
Shock Loan Subsetting	All	Other Only	All	Other Only
Mean Credit Shortfall	3.4	19.6	3.7	19.7
SD Credit Shortfall	9.9	19.0	11.2	19.6
Observations	7235	7235	7235	7235

Standard errors in parentheses

Standard errors are two-way clustered at the Census Tract and Neighborhood-Year levels

Variables are winsorised at the 1 and 99th percentile to handle extreme outliers

Observations are weighted by 2010 Census housing units

Variables transformed so that credit shock coefficients can be interpreted as:

A 1 dollar credit shortfall per HH leads to a \_ dollar change in housing investment

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

#### 4.4. SPILLOVERS IN INVESTMENT

Table 4.5 (unlagged) and table 4.6 (lagged) show that there are strong positive correlations in housing investment across time and space, controlling for year and tract fixed effects. The first specifications in table 4.5 and table 4.6 show that a \$1 increase in housing investment in neighboring tracts is correlated with a \$0.774 increase in investment in a tract contemporaneously, and a \$0.410 increase in investment in a tract in the subsequent year. Despite the use of time lags, fixed effects, and clustering at the tract and neighborhood-year levels, there is always some concern as noted by Manski (1993) that this result is driven by the reflection problem or, more likely, by correlated unobservables. However, this baseline correlation is robust to time frame used to construct the lag (years or an average of the previous 6 months), level of aggregation (blocks, block groups, tracts), and the form of the spatial weights (different distances, as well as rook weights, queen weights, and nearest neighbor weights). The existence of spatial correlation in investment controlling for twoway effects is extremely robust, but causality remains uncertain.

Table 4.5: Spillovers in Housing Investment across space

	HInvest	Renov	Easy
Slag Housing Investment Per HH	0.774*** (0.053)		
Slag Renovation Investment Per HH		0.767*** (0.053)	0.001 (0.005)
Slag Easy Permit Investment Per HH		0.100 (0.203)	0.779*** (0.042)
Year FE	Yes	Yes	Yes
Tract FE	Yes	Yes	Yes
Observations	8760	8760	8760

Standard errors in parentheses

Standard errors are two-way clustered at the Census Tract and Neighborhood-Year levels

Variables are winsorised at the 1 and 99th percentile to handle extreme outliers

Observations are weighted by 2010 Census housing units

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Table 4.6: Spillovers in Housing Investment across time and space

	HInvest	Renov	Easy
Lag Housing Investment Per HH	0.085*** (0.022)		
Lag Slag Housing Investment Per HH	0.410*** (0.054)		
Lag Renovation Investment Per HH		0.081*** (0.025)	0.001 (0.002)
Lag Slag Renovation Investment Per HH		0.427*** (0.054)	-0.023*** (0.007)
Lag Easy Permit Investment Per HH		0.146* (0.079)	0.009 (0.019)
Lag Slag Easy Permit Investment Per HH		0.062 (0.280)	0.519*** (0.044)
Year FE	Yes	Yes	Yes
Tract FE	Yes	Yes	Yes
Observations	8030	8030	8030

Standard errors in parentheses

Standard errors are two-way clustered at the Census Tract and Neighborhood-Year levels

Variables are winsorised at the 1 and 99th percentile to handle extreme outliers

Observations are weighted by 2010 Census housing units

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

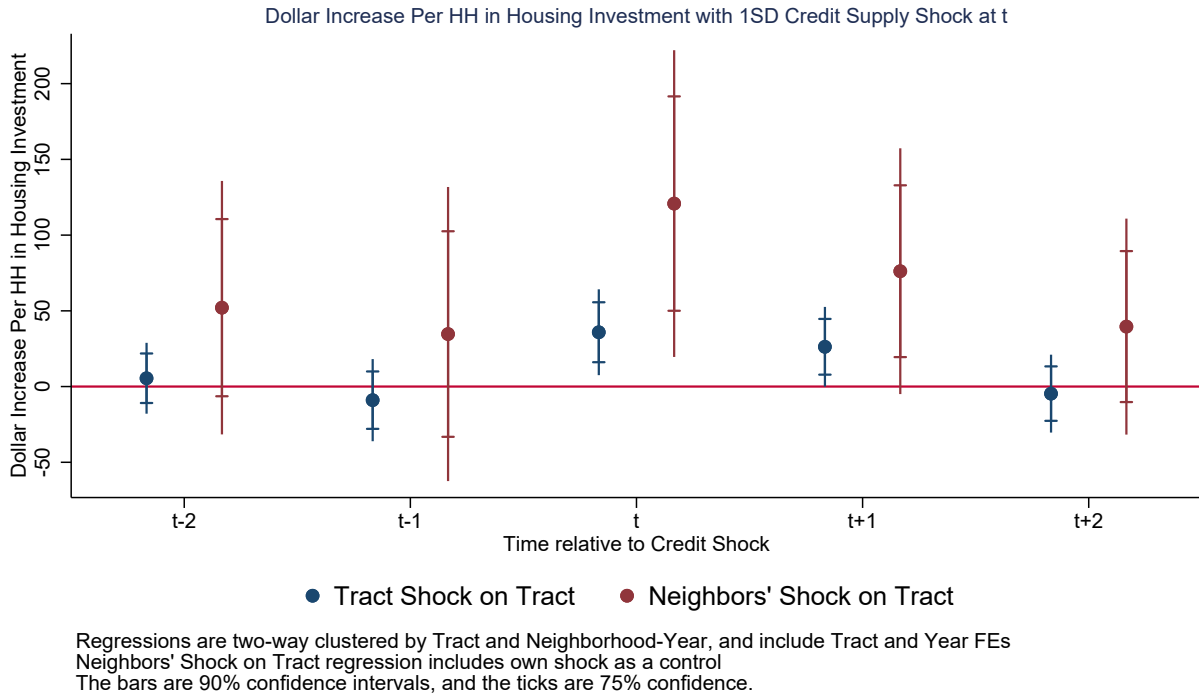
Building off of these correlations, I return to the effects of credit supply both directly on investment, and indirectly through neighbors' investment. To understand the temporal and spatial effects, as well as to test for pre-trends, I combine spatial weights with the credit supply shock. Figure 4.2 presents a reduced form coefficient plot, visualizing the  $\beta$  and  $\kappa$  coefficients from the following regression:

$$\text{Invest}_{jt} = \sum_{\tau=t-2}^{t+2} \beta_{\tau} \text{CS}_{j\tau} + \sum_{\tau=t-2}^{t+2} \kappa_{\tau} \text{CS}_{n(j)\tau} + \gamma_j + \delta_t + \epsilon_{jt}$$

Specifically, the blue dots refer to the  $\beta$ s, and the red dots refer to the  $\kappa$ s, both controlling for fixed effects  $\gamma_j$  and  $\delta_t$ . With this visualization, there are multiple goals. First, all effects at  $t - 2$  and  $t - 1$  should be indistinguishable from zero, since current credit supply shocks at  $t$  should have no effect on housing investment at  $t - 2$  or  $t - 1$ . If these effects were not zero, the primary concern would be that credit supply shocks capture neighborhood-level effects that may be persistent across a few years, rather than being shocks that are relevant themselves and time  $t$  specific: a problem of correlated unobservables. Second, I expect to see a positive effect on impact at  $t$ , potentially declining to zero in future years  $t + 1$  and  $t + 2$ . Note that this effect is not plotted as cumulative, so effects declining to zero simply means that credit supply shocks at  $t$  have no effect on investment at time  $t + 2$  controlling for the effect of these shocks at times  $t - 2$  through  $t + 1$ . The points on the figure are exactly the regression coefficients themselves. The blue dots show that the credit supply shock increases investment within the tract, consistent with the baseline specification. This effect persists in the subsequent year, before becoming insignificant a year later. The red dots show the coefficient on neighbors' credit supply shocks, and are suggestive of an increase in investment when one's neighbors experience a positive credit supply shock, controlling for one's own shock, particularly at time  $t$ . Standard errors in this figure are larger than in the main regressions because (i) this specification includes 5 different time-lags simultaneously, meaning that the coefficients are estimated using only residual differences controlling for these other lags and (ii) the need for two leads and two lags shortens

the sample by four years, leading to substantially less data for estimation.<sup>27</sup>

Figure 4.2: Coefficient Plot of Spillovers across time and space



Building on the OLS evidence provided by running the first stage, second stage correlations, and reduced form above, I now turn to an IV approach to assess the causal strength of the spillovers suggested in the correlations in table 4.5 and table 4.6. Specifically, these specifications use the full empirical model in equation (4.4) to estimate the causal spillover effect: the extent to which neighbors' investment drives tract investment, instrumented by neighbors' credit shocks, and controlling for the tract's credit shocks.

$$\text{Invest}_{n(j)t} = \text{FE}_{jt} + \mathbf{X}_{jt}\beta + \mathbf{X}_{n(j)t}\mu + \epsilon_{jt}$$

$$\text{Invest}_{jt} = \text{FE}_{jt} + \mathbf{X}_{jt}\beta + \widehat{\text{Invest}}_{n(j)t}\rho + \epsilon_{jt}$$

Table 4.7 presents these specifications using contemporaneous data, while table 4.8 uses a

<sup>27</sup>The former is intentional, while the latter is a necessary consequence of the method. See ?? for an example of this same methodology, with similarly large confidence intervals.

Table 4.7: IV: Instrumented Spillovers

	HInvest	Renov	Easy
Slag Housing Investment Per HH	0.895*** (0.182)		
Slag Renovation Investment Per HH		0.876*** (0.149)	
Slag Easy Permit Investment Per HH			0.748*** (0.168)
Year FE	Yes	Yes	Yes
Tract FE	Yes	Yes	Yes
Kleibergen-Paap F-stat	23.197	35.855	35.698
Stock-Yogo crit value: 10% max IV bias	9.080	9.080	9.080
Stock-Yogo crit value: 10% max IV size	22.300	22.300	22.300
Anderson-Rubin weak-id robust joint p-value	0.001	0.000	0.001
Hansen J p-value	0.498	0.480	0.812
Observations	7161	7161	7161

Standard errors in parentheses

Specifications control for own credit supply shock, own credit shortfall, and frac continuing lenders

Credit shortfall measure uses linear specification, subset to loans with unspecified rejection reason

Standard errors are two-way clustered at the Census Tract and Neighborhood-Year levels

Variables are winsorised at the 1 and 99th percentile to handle extreme outliers

Observations are weighted by 2010 Census housing units

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Table 4.8: IV: Lag Instrumented Spillovers

	HInvest	Renov	Easy
Lag Slag Housing Investment Per HH	0.229 (0.179)		
Lag Slag Renovation Investment Per HH		0.298* (0.156)	
Lag Slag Easy Permit Investment Per HH			0.414** (0.180)
Year FE	Yes	Yes	Yes
Tract FE	Yes	Yes	Yes
Kleibergen-Paap F-stat	23.629	34.313	34.158
Stock-Yogo crit value: 10% max IV bias	9.080	9.080	9.080
Stock-Yogo crit value: 10% max IV size	22.300	22.300	22.300
Anderson-Rubin weak-id robust joint p-value	0.129	0.042	0.023
Hansen J p-value	0.087	0.050	0.194
Observations	6422	6422	6422

Standard errors in parentheses

Specifications control for own credit supply shock, own credit shortfall, and frac continuing lenders

Credit shortfall measure uses linear specification, subset to loans with unspecified rejection reason

Standard errors are two-way clustered at the Census Tract and Neighborhood-Year levels

Variables are winsorised at the 1 and 99th percentile to handle extreme outliers

Observations are weighted by 2010 Census housing units

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

lagged specification, both omitting controls from the table for clarity. In the contemporaneous regression, I find a strong role for spillover effects, with a 1% increase in lending predicting a 0.748-0.895% increase in investment in the IV estimation depending on investment type. Even estimating this specification with lags leads to a modest positive effect of between 0.229-0.414%, suggesting some persistence across time. These results are roughly similar to the OLS results, with contemporaneous IV results actually being slightly larger in magnitude, while lagged results are slightly smaller and less significant. Together, these suggest a multiplier of almost 2 for the effects of refinancing credit on investment - the magnitude of the direct effect is matched almost exactly by the size of the resulting spillover effect over two years. The reduced form specifications are presented in appendix C.1. Specifications in which each instrument is used independently of the other yield similar results and are presented in appendix C.2.

Because the IV strategy relies on a credit supply shock using data from outside of Chicago and a spatial formulation, there are some ex-ante concerns about weak instruments and instrument exogeneity respectively. Specifically, weak instruments could be a concern if the spatial correlation in lending is too high, so that the neighbors' shock instrument does not contain relevant information when controlling for the own tract supply shock. In Frisch-Waugh terms, if the neighbors' shocks are too well predicted by the own tract shocks (high spatial correlation), then the residual of this specification, which is the instrument, could easily be mostly error. Instrument exogeneity is a concern because of the spatial nature of the problem and the spatial averaging used in constructing the instrument. In particular, if neighbors' shocks are more precisely estimated than own tract shocks (because they are an average of many neighbors), the own tract controls could fail, and neighbors' shocks could be almost a direct proxy for the own tract shock. This would show up as an instrument exogeneity concern - neighbors' shocks would be informative about the tract's investment through other channels than the neighbors' own investment. These two concerns are almost opposites in this particular context; the former reflects the possibility that the instrument is too weak to survive the controls, while the latter could arise if the neighbors' shock is so precisely estimated that the own shock control intended

to address spatial correlation concerns fail.

Despite these concerns, standard IV tests detect almost no issues with the identification, except marginal issues with power and exogeneity in the lagged specification. A large Kleibergen-Paap rk LM statistic rejects underidentification, and the Kleibergen-Paap rk Wald F statistic exceeds the 5% or 10% maximal IV critical value across all specifications. Though we cannot test instrument exogeneity directly, the p-value for Hansen's J statistic fails to reject the null with p-values above 0.48 in all contemporaneous specifications. This suggests that the excluded instruments are appropriately independent of the second-stage error, or that over-identifying restrictions hold. The J statistic p-value is lower for lagged specifications, and marginally rejects the model for renovation permits relative to a 5% threshold ( $0.050 = 0.050$ ). In the lagged specification on general home investment, the Anderson-Rubin test suggests that the instruments may be weak relative to a 5% threshold ( $0.129 > 0.050$ ). Therefore overall, the contemporaneous model exhibits no observable IV concerns, while the lagged specification's instrument is marginally weak for only one of the three specifications.

## 5. MODEL

### 5.1. SETUP

The empirical results presented in the previous section suggest that (i) maintenance activity of a household is positively predicted by the maintenance activity of its neighbors, (ii) credit supply constraints can prevent building managers from investing in housing, and (iii) the effects of credit supply and constraints can spill over across space. In this section, I develop a model that combines the monocentric city model with a dynamic model of building managers' maintenance decisions to capture these features. The model centers around building managers' dynamic housing investment decisions as they seek to maximize their rental income from tenants who value both housing and neighborhood quality. Specifically, the model contains two types of agents: building managers who choose a level of maintenance in each period, and workers who work in the city and bid for housing in each period. Building managers seek to maximize profit. Workers seek to maximize the utility value of their consumption bundle, which comprises both housing and non-housing consumption.

The city is represented by a two dimensional coordinate plane, where locations are denoted by a pair  $(x, y)$ . I normalize  $(0, 0)$  to be the central business district (CBD) without loss of generality. The city is assumed to be small and open relative to the national economy, and workers are freely mobile within and across cities. Individuals consume a numeraire good  $c$ , with price  $p_c$  normalized to 1. Time is discrete and runs indefinitely. There are  $N_t$  individuals living in the city at time  $t$ . Workers are heterogeneous in their skill levels, which are drawn from a distribution  $\mathcal{S}$ .  $\mathcal{S}$  is assumed to have nonzero support from  $[0, \bar{s}]$ , where  $\bar{s}$  is large. The wage workers obtain in a given city  $l$  is defined as the product of their city-specific productivity  $A_l$  and their skill level  $s$ :

$$w_l^s = A_l s$$

Workers of type  $s$  obtain utility  $\bar{U}^s$  in equilibrium, where  $\bar{U}^s$  is fixed in all periods and repre-

sents the outside option of moving to other cities or rural areas for a worker with skill  $s$ .<sup>28</sup> This outside option plays a critical role in determining house prices in the city, as it bounds workers' willingness to forgo consumption to pay for housing.

Heterogeneous skill levels imply heterogeneous wages, which leads workers to demand different housing quality and crucially to have different marginal values of quality. This has the effect of increasing the marginal value of investment at the neighborhood level - improvements across the neighborhood will lead to gentrification, whereby higher skilled workers displace lower skilled workers and pay a premium for the high housing and neighborhood quality.

Combining these assumptions implies that workers of different skill levels from a national pool compete with one another for housing in the city, bidding up prices until units of a given quality are inhabited by the worker type  $s$  that is willing to pay the most. This bid-rent approach allows the model to remain tractable, but leans heavily on the assumption that there is an endless supply of workers at each skill level  $s$  who are able to bid for housing.<sup>29</sup> This strong assumption limits the scope of policy questions that the model can address. In particular, changing the national skill distribution  $\mathcal{S}$  does not affect equilibrium because the city is assumed sufficiently small that only the domain of  $s$ , rather than its distribution, is relevant; house prices do not respond to the quantity of houses in the city that are of the same quality, nor to the density of buyers of a given skill level, since there are always more potential buyers at each skill level. Effectively, the market is frictionless for building managers - they can always find a renter, and the renter they find is always of the type  $s$  with the highest willingness to pay for that quality location. The effect of loosening this assumption can be roughly approximated by increasing the cost of investment for net improvements, decreasing the marginal value of investment and forcing building managers to internalize the menu costs associated with finding a new renter who is willing to pay a higher rent that matches the new housing quality.

---

<sup>28</sup>There is scope for the model to handle a system of cities, in which case workers of each type choose the city that maximizes their utility. The outside option is required here because of the focus on a single city.

<sup>29</sup>In particular, this will be crucial for deriving equation (5.15), which is used to make the supply side of the model analytically tractable.

### 5.1.1. WORKERS' HOUSING DEMAND

In each period, workers with skill level  $s$  choose a consumption level  $c$  of the numeraire good with price  $p_c = 1$  and a location  $(x, y)$  in which to live by solving:

$$\max_{c, (x, y)} u(Q_{x, y}, c) \quad \text{s.t.} \quad A_l s - \tau d_{x, y} = c + P(Q_{x, y}, d_{x, y})$$

where utility is assumed to be increasing and strictly concave in each of its arguments.  $d_{x, y}$  is the distance from the house to the CBD.  $\tau > 0$  denotes the monetary cost of commuting.  $A_l$  is the city-specific productivity, such that wages are  $A_l s$ .  $P(Q_{x, y}, d_{x, y})$  is the price of housing, defined as a function of location quality and distance. Location quality  $Q_{x, y}$  can be further decomposed as follows:

$$Q_{x, y} = f(q_{x, y}, \bar{q}_{x, y}, \zeta_{x, y}, \bar{\zeta}_{x, y})$$

$q_{x, y}$  is *housing quality*, defined as the quality of the building in which the individual is choosing to live.  $\bar{q}_{x, y}$  is *neighborhood quality*, defined as the quality of nearby locations' houses.  $\zeta_{x, y}$  are non-housing amenities available to individuals living at  $(x, y)$ .  $\bar{\zeta}_{x, y}$  are the non-housing amenities of  $(x, y)$ 's neighbors. I assume the following partially-additive functional form for housing quality:

$$Q_{x, y} = f(q_{x, y} + \zeta_{x, y}, \bar{q}_{x, y} + \bar{\zeta}_{x, y})$$

This implies that a each location has a "local" quality  $q_{x, y} + \zeta_{x, y}$  that is defined as the sum of its housing and non-housing amenities. A location's full quality  $Q_{x, y}$ , which defines its desirability to workers and its price, then depends on a combination of this local quality, and the local quality of its neighbors.

Dropping the  $(x, y)$  location subscripts, the problem can be written as:

$$\max_{c, Q, d} u(Q, c) \quad \text{s.t.} \quad A_l s - \tau d = c + P(Q, d)$$

The general solution is provided in appendix D. For the purposes of obtaining housing demand, only the utility equalization equation is necessary:

$$\bar{U}^s \geq u(Q, A_l s - \tau d - P(Q, d))$$

Importantly, this equation holds with equality only if type  $s$  workers inhabit this location in equilibrium. However, this equation can be written with equality by changing  $P(Q, d)$  to  $P^s(Q, d)$ , the willingness to pay of type  $s$  workers for housing of quality  $(Q, d)$ , where  $P^s(Q, d) \leq P(Q, d)$ :

$$\bar{U}^s = u(Q, A_l s - \tau d - P^s(Q, d))$$

This equation can be solved for  $P^s(Q, d)$  as a function  $f$  of the other parameters, where  $f$  depends on the specific functional form of  $u(\cdot)$ .

$$P^s(Q, d) = f(Q, d, \bar{U}^s, A_l, s, \tau) \tag{5.1}$$

Prices  $P(Q, d)$  are determined by applying a bid-rent procedure in which workers compete for housing. The type  $s$  worker with the highest willingness to pay for a given quality and distance determines the equilibrium price.

$$P(Q, d) = \max_s \{P^s(Q, d)\} \tag{5.2}$$

Note that if there were only one worker type, prices would be directly determined by the utility equalization equation. With a continuum of types, prices are instead determined competitively by workers seeking a house that trades off consumption and housing costs in a way that internalizes their specific outside option.

### 5.1.2. BUILDING MANAGERS

Each plot of land  $(x, y)$  is run by a building manager. No distinction is made between the building being owned by a landlord or being owner-occupied (managed by the tenants) - the model treats both as having the same investment incentives.<sup>30</sup> As a result, it is assumed that buildings can be transferred to new ownership via sales that reflect the present discounted value of the property, but that such transfers do not change the problem the current manager solves in period  $t$ . Building managers face a dynamic problem - in each period they decide the extent to which they want to invest in their property to counteract depreciation rate  $\delta$ . Specifically, each building manager's problem, suppressing coordinates  $(x, y)$ , is defined by the following Bellman equation:

$$V(q, \bar{q}, d, \zeta, \bar{\zeta}) = (1 - \omega)P(q + \zeta, \bar{q} + \bar{\zeta}, d) + \max_{m \geq 0} \left\{ -c(m) + \beta V((1 - \delta)q + m, E[\bar{q}_{t+1} | \bar{q}], d, \zeta, \bar{\zeta}) \right\}$$

The Bellman equation consists of a few basic components.  $(1 - \omega)P(q + \zeta, \bar{q} + \bar{\zeta}, d)$  reflects the one-period rental income associated with the current quality and location, subject to a non-maintenance flow cost rate  $\omega$ , which includes taxes and mortgage interest rates. Note that non-housing amenities  $\zeta$  and  $\bar{\zeta}$ , such as parks, crime, and school quality are assumed not to respond to changes in the housing stock in the short and medium run. In a long-run model, amenities  $\bar{\zeta}$  comprising  $\zeta$  and  $\bar{\zeta}$  would likely respond to housing quality.<sup>31</sup>

$\bar{q}$  and  $\bar{\zeta}$  are the housing and non-housing quality of the surrounding neighborhood respectively, as noted in the consumer's problem. Generally, this can be any function of neighbors' qualities that has the property that closer neighbors are assigned relatively more weight on the

<sup>30</sup>In practice, many differences that are outside of the scope of the model are possible. Landlords may, for instance, have more scope for skimping on maintenance because of sticky rental prices, or may be better able to coordinate investment across multiple units.

<sup>31</sup>For example, increases in housing quality may lead to changes in the crime rate, greater incentives for business investment, or increased pressure on local government to invest in amenities such as parks.

neighborhood quality of a given location:

$$\bar{q}_{x,y} = \bar{q}_{x,y} \left( q_{\{X,Y\} \setminus (x,y)} \right) \quad \bar{\zeta}_{x,y} = \bar{\zeta}_{x,y} (\zeta_{\{X,Y\} \setminus (x,y)})$$

In a continuous space, this can be thought of as an exponential discounting form, similar to that used in Lucas and Rossi-Hansberg (2002), with  $\delta_q$  as the exponential discounting rate:

$$\bar{q}_{x,y} + \bar{\zeta}_{x,y} = A \int_0^\infty \int_0^{2\pi} \left( q_{d \cos \phi, d \sin \phi} + \zeta_{d \cos \phi, d \sin \phi} \right) e^{-\delta_q d} \delta \phi \delta d \quad (5.3)$$

In practice, most empirical analysis uses discrete (block groups, tracts, or zip codes) rather than continuous space. In this case,  $\bar{q}$  is the product of the spatial weights matrix  $\mathbf{G}$  and  $q$  as used in the empirical analysis.<sup>32</sup> Stacking the  $\bar{q}$  values for each discrete location, and stacking the vector of qualities  $q$  and location-specific amenities  $\zeta$  similarly, neighborhood qualities can be defined through the weights matrix  $G$  as:

$$\vec{\bar{q}} = \mathbf{G} \vec{q} \quad \vec{\bar{\zeta}} = \mathbf{G} \vec{\zeta} \quad (5.4)$$

As noted in the empirical section,  $\mathbf{G}$  is assumed to have a number of standard properties. First, each row of  $\mathbf{G}$  sums to 1 as a normalization, which ensures that  $\bar{q}$  and  $q$  are measured on the same scale.<sup>33</sup> Next, the diagonal of  $\mathbf{G}$  is zero, indicating that neighborhood quality for a given location is a function of its neighbors, not of itself.<sup>34</sup>

The expected quality of the surrounding neighborhood in the next period,  $E[\bar{q}_{t+1} | \bar{q}]$ , reflects

<sup>32</sup>See Gibbons et al. (2015). In spatial econometrics, this is sometimes called the  $W$  matrix.

<sup>33</sup>In particular, if all locations have the same quality  $q$ , then  $\bar{q}$  will be equal to this same value in all locations.

<sup>34</sup>For example, in a block-grid structure - where each block has exactly four neighbors -  $\bar{q}$  could simply be defined as the mean quality of the block's four neighbors in Rook-contiguity space:

$$\bar{q}_{x,y} = \frac{(q_{x,y-1} + \zeta_{x,y-1}) + (q_{x-1,y} + \zeta_{x-1,y}) + (q_{x,y+1} + \zeta_{x,y+1}) + (q_{x+1,y} + \zeta_{x+1,y})}{4} \quad (5.5)$$

In this case, the weights matrix takes the form of an  $N \times N$  matrix (where  $N$  is the number of blocks), where each row comprises  $N - 4$  zero values, and 4 values of  $1/4$ .  $G$  of this form is often called a contiguity matrix, and it's not necessary that each block have exactly 4 neighbors. In this particular case, the contiguity matrix allows us to construct spatial averages of quality.

the interdependence of building managers' decisions, and therefore of the Bellman equations, at each location in each period. To make this tractable, I remove uncertainty by assuming a rational expectations condition that imposes perfect foresight; building managers solve their problem assuming that the neighborhood quality in the subsequent period will be  $\bar{q}_{t+1}$ , and are correct in this assumption in equilibrium:

$$\vec{\bar{q}}_{t+1} = \mathbf{G}\vec{\bar{q}}_{t+1} \quad (5.6)$$

Computationally, this condition can be thought of as the result of an iterative process. Initially, managers solve their problem assuming the neighborhood will not change, then repeatedly and collectively internalize that this is incorrect and adjust their expectations. For the majority of the analysis, building managers solve their problem as a function of  $E[\bar{q}_{t+1}|\bar{q}]$ , often simplified notationally to  $\bar{q}_{t+1}$ . Equation (5.6) is then used to construct an equilibrium optimal policy. Though building managers look ahead to the next period, they do not internalize the future behavior of other building managers, and what it implies about the long-run future of neighborhood quality.

$c(m)$  is the convex cost of maintenance per unit of quality. Maintenance costs are assumed to be convex to reflect the intuition that maintaining a property at a cursory level is somewhat easy: fixing readily apparent problems involves relatively little search cost. By contrast, maintaining a building thoroughly requires searching for less visible concerns and fixing those as well. This search cost increases with the level of maintenance. Explicitly,  $c(m)$  satisfies:

$$c'(m) > 0 \quad c''(m) \geq 0 \quad c(0) = 0$$

$\delta \in (0, 1)$  is the natural rate of housing depreciation in the absence of maintenance, which is assumed to be homogeneous across locations. The actual rate of depreciation is impacted by the maintenance decision.  $\beta \in (0, 1)$  is the discount factor, which is related to the interest rate

by  $\beta = 1/(1+r)$ .

### 5.1.3. SOLVING THE BELLMAN

Recall that the Bellman Equation at each location  $(x, y)$  is

$$V(q, \bar{q}, d, \zeta, \bar{\zeta}) = (1 - \omega)P(q + \zeta, \bar{q} + \bar{\zeta}, d) \\ + \max_{m \geq 0} \left\{ -c(m) + \beta V((1 - \delta)q + m, E[\bar{q}_{t+1} | \bar{q}_t], d, \zeta, \bar{\zeta}) \right\}$$

The problem can be rewritten in terms of choosing next period's housing quality  $q_{t+1}$  rather than choosing a level of maintenance  $m$ :

$$V(q, \bar{q}, d, \zeta, \bar{\zeta}) = (1 - \omega)P(q + \zeta, \bar{q} + \bar{\zeta}, d) \\ + \max_{q_{t+1} \geq (1 - \delta)q} \left\{ -c(q_{t+1} - (1 - \delta)q) + \beta V(q_{t+1}, E[\bar{q}_{t+1} | \bar{q}_t], d, \zeta, \bar{\zeta}) \right\}$$

This frames the problem as a standard dynamic investment problem, where the building manager treats housing as her capital stock, and determines the extent to which to build new capital between periods to maximize rents given her expectations about next period's neighborhood quality.<sup>35</sup> The solution to this problem is provided in appendix D, and yields the following dynamic Euler equation:

$$c'(q_{t+1} - (1 - \delta)q_t) = \beta(1 - \omega)(P_q(q_{t+1} + \zeta, E[\bar{q}_{t+1} | \bar{q}_t] + \bar{\zeta}, d) \tag{5.7}$$

$$+ (1 - \delta)c'(q_{t+2} - (1 - \delta)q_{t+1})) \tag{5.8}$$

Equation (5.7) consists of three primary components.  $c'(q_{t+1} - (1 - \delta)q_t)$  denotes the marginal cost of an additional unit of quality next period.  $\beta(1 - \omega)(P_q(q_{t+1} + \zeta, E[\bar{q}_{t+1} | \bar{q}_t] + \bar{\zeta}, d))$  is the marginal benefit of an additional unit of quality next period insofar as it contributes to a higher rental price in the next period.  $\beta(1 - \omega)(1 - \delta)c'(q_{t+2} - (1 - \delta)q_{t+1})$  denotes the continuation ben-

<sup>35</sup>See, for instance, Stokey et al. (2004) Section 5.9.

efits of having performed maintenance now, which decreases the need for future maintenance. All terms are weakly positive given the lower bound constraint  $q_{t+1} \geq (1 - \delta)q_t$ .

Setting  $q = q_t = q_{t+1} = q^{SS}(\bar{q}_t)$  and solving for  $q^{SS}(E[\bar{q}_{t+1}|\vec{q}_t])$  yields an implicit formula for the “partial” steady state  $q^{SS}$  of  $q$  as a function of  $\bar{q}_{t+1}$ . This is an optimal policy target conditional on the neighborhood quality remaining at  $\bar{q}_{t+1}$  for each subsequent period.

$$c'(\delta q^{SS}) = \left( \frac{(1 - \omega)P_q(q^{SS} + \zeta, \bar{q}_{t+1} + \bar{\zeta}, d)}{r + \delta} \right) \quad (5.9)$$

In the case of  $\bar{\zeta} = 0$ , ex-ante symmetric neighborhoods with no amenities, the following equation defines a symmetric steady state, in which  $\bar{q}_{t+1} = q^{SS}$ :

$$c'(\delta q^{SS}) = \left( \frac{(1 - \omega)P_q(q^{SS}, q^{SS}, d)}{r + \delta} \right) \quad (5.10)$$

Existence of the symmetric steady state depends on the convexity of  $c(\cdot)$  and the weak concavity of  $P_q(q, \bar{q}_{t+1}, d)$  when  $\bar{q}_{t+1}$  is constrained such that  $q = \bar{q}_{t+1} = q^{SS}$ .

Equation (5.7) is the primary result of the solution to the Bellman equation. Specifically, the Euler equation summarizes the dynamic incentives building managers face in the model, and how these incentives are affected by changes in non-housing amenities, initial housing quality, and projected neighborhood quality.

#### 5.1.4. CREDIT CONSTRAINTS

Thus far, it has been assumed that building managers are free to borrow at will, and to run a deficit for any number of periods without repercussion. In reality, many building managers do not have this luxury, and are instead constrained in the extent to which they can maintain their building in each period.

There are many ways to impose a credit constraint within the model. One intuitive approach, that leans heavily on the existing model parameters, is a “self-sustaining investment” con-

straint. This imposes that building managers cannot invest more in the property in a given month than the rental income they accrue net of non-maintenance flow costs  $\omega$ . Since the model has no fixed costs, and costs are convex, this relatively strict condition will not prevent major maintenance, but will sometimes force substantial investments to be split across multiple periods. In practice, housing investors may save up over a few periods in order to do one larger investment. However, this entails some sort of non-convexity in the investment cost function that is not present in the model, and is reasonably approximated by a series of small investments. The constraint amounts to requiring that the choice  $q_{t+1}$  satisfy:

$$c(q_{t+1} - (1 - \delta)q_t) \leq (1 - \omega)P(q_t + \zeta, \bar{q}_t + \bar{\zeta}, d) \quad (5.11)$$

This constraint directly reflects the empirical finding that changes in flow income, caused by things like refinancing credit supply, have an observable effect on investment behavior. If investment decisions were made solely on the basis of savings, this would not be observable in the data. In practice, some housing investors do have savings from which to draw funds, but this constraint may still be relevant and binding if they are risk-averse and uncertain about their ability to capitalize any investments in the home into higher rental income by raising rents or finding a new tenant, or if they are homeowners and are unsure if the improvements will increase the market value of the home rather than personalize it to their tastes. Therefore while this constraint is intended to capture behavior associated with a complete lack of savings, it could also simply be the strategy of a frugal housing investor who opts not to invest more than her returns due to frictions that are otherwise slightly outside of the model.

In later sections, I show that this credit constraint can have an effect on the steady state distribution of housing quality in the city, depending on the initial housing quality distribution  $\bar{q}$ .

## 5.2. FUNCTIONAL FORM ASSUMPTIONS

To obtain a solution that captures the relevant complementarities between housing quality, neighborhood quality, and consumption, I assume that utility is CES in own quality and neighborhood quality with elasticity of substitution  $\sigma$ , nested in a Cobb-Douglas relationship between housing and consumption. In particular, workers have identical utility functions defined by:

$$u(Q, c) = Q^{1-\alpha_c} c^{\alpha_c}$$

Where

$$Q = \left( \alpha_q (q + \zeta)^{\frac{\sigma-1}{\sigma}} + \alpha_{\bar{q}} (\bar{q} + \bar{\zeta})^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} \quad (5.12)$$

$$\alpha_q + \alpha_{\bar{q}} = 1$$

$$\sigma > 0$$

Furthermore, the cost of maintenance  $c(m)$  is assumed to have the following simple functional form, where  $\rho = 1$  are linear maintenance costs, and  $\rho > 1$  are convex costs:<sup>36</sup>

$$c(m) = \frac{A_m}{\rho} m^\rho \quad \rho > 1 \quad (5.13)$$

---

<sup>36</sup> $\rho \geq 1$  is not strictly necessary - costs simply need to be more convex than prices to ensure that the model doesn't have an incentive for unbounded investment.

### 5.2.1. IMPOSING FUNCTIONAL FORMS

Combining the CES functional form with equation (5.1) yields an equation for  $P^s(q, \bar{q}, d)$ , the willingness to pay for a given location of individuals with skill  $s$ :<sup>37</sup>

$$P^s(Q, d) = A_l s - \tau d - \left( \frac{\bar{U}^s}{Q^{1-\alpha_c}} \right)^{\frac{1}{\alpha_c}} \quad (5.14)$$

Following the bid-rent procedure defined by the maximization problem in equation (5.2), the price of a particular location is determined by finding the worker type  $s$ , corresponding to the wage-utility pair  $(A_l s, \bar{U}^s)$ , that maximizes  $P^s(q, \bar{q}, d)$ .<sup>38</sup> This approach implicitly invokes the free mobility and small open city assumptions because it requires a national housing market, in which workers in other cities can move to the city and bid up prices for housing of their desired quality, and similarly that workers in the city who cannot obtain their desired housing quality at an acceptable price can leave. Importantly, the worker with the highest skill does not necessarily have the highest willingness to pay for all types of housing because high-skill workers have a higher-utility outside option in another city. Instead, higher skill workers demand higher quality houses, and the prices of their preferred houses are driven up by workers with slightly lower skill.

The outside option  $\bar{U}^s$  is defined as an increasing function of skill level  $s$ :  $\bar{U}^s = f(s)$ .  $P^s$  has a unique maximum as long as  $f'(s) > \alpha_c$ , since this ensures that the second derivative of  $P^s$  with respect to  $s$  is negative. As a result, plugging in  $\bar{U}^s = f(s)$  and taking a first order condition of  $P^s$  with respect to  $s$  yields an equation that implicitly defines the worker type  $s$  that “matches” to the given housing type:<sup>39</sup>

$$\frac{1}{\alpha_c} Q^{-\frac{(1-\alpha_c)}{\alpha_c}} f(s)^{\frac{1-\alpha_c}{\alpha_c}} f'(s) = A_l$$

<sup>37</sup>Intuitively, it can be shown that equation (5.14) can also be obtained by solving the differential equation that results from plugging the assumed utility function and budget constraint into the FOCs and solving the resulting differential equation.  $\bar{U}^s$  is the resulting constant of integration from that approach.

<sup>38</sup>See Duranton and Puga (2015) for more details on bid-rent procedures in cities.

<sup>39</sup>Here a “match” means that this worker type has the highest willingness to pay for that housing.

$$f(s)f'(s)^{\frac{\alpha_c}{1-\alpha_c}} = (A_l\alpha_c)^{\frac{\alpha_c}{1-\alpha_c}} Q$$

A natural simplification is to consider the case in which  $f(s) = a_u s$ , imposing that the outside option of moving to another city allows people with higher skill to translate that skill to utility linearly.<sup>40</sup> This yields the following matching condition:

$$s = \frac{1}{\alpha_u} \left( \frac{A_l\alpha_c}{\alpha_u} \right)^{\frac{\alpha_c}{1-\alpha_c}} Q$$

Plugging this in to equation (5.14) yields a closed form for the maximum willingness to pay of workers for housing as a function of quality and distance to the CBD:

$$\bar{P}(Q, d) = \left( \frac{A_l}{a_u} \right)^{\frac{1}{1-\alpha_c}} \alpha_c^{\frac{\alpha_c}{1-\alpha_c}} (1 - \alpha_c) Q - \tau d \quad (5.15)$$

This defines a quality and distance cutoff below which no workers are willing to live in a location for any positive price.  $R(Q, d)$  is an indicator for whether a particular location has high enough quality to be part of the city:

$$R(Q, d) = I\{P(Q, d) \geq 0\} \quad (5.16)$$

The price of housing is then:

$$P(Q, d) = \begin{cases} \bar{P}(Q, d) & : R(Q, d) = 1 \\ 0 & : R(Q, d) = 0 \end{cases} \quad (5.17)$$

To understand the building managers' investment incentives, the marginal value of increased quality is crucial. The corresponding derivative of the price function with respect to quality  $q$

---

<sup>40</sup>Note that a further simplifying assumption is that  $a_u = A_l$ , which implies that the utility returns to skill are equal to the wage returns to skill in the city.

holding  $\bar{q}$  fixed (plugging in for  $Q$  using equation (5.12)) is:

$$P_q(q, \bar{q}, d) = \begin{cases} \left(\frac{A_t}{\alpha_u}\right)^{\frac{1}{1-\alpha_c}} \alpha_c^{\frac{\alpha_c}{1-\alpha_c}} (1-\alpha_c) \alpha_q \left(\frac{Q}{q+\zeta}\right)^{\frac{1}{\sigma}} & : R(q, \bar{q}, d) = 1 \\ 0 & : R(q, \bar{q}, d) = 0 \end{cases} \quad (5.18)$$

Where

$$\left(\frac{Q}{q+\zeta}\right)^{\frac{1}{\sigma}} = \left(\alpha_q + \alpha_{\bar{q}} \left(\frac{\bar{q} + \bar{\zeta}}{q+\zeta}\right)^{\frac{\sigma-1}{\sigma}}\right)^{\frac{1}{\sigma-1}}$$

$P(\cdot)$  is concave in both forms of quality, and linear in distance. Concavity in  $q$  ensures that the value function in the subsequent section is concave, even in the case of linear maintenance costs. Furthermore,  $P$  is also continuous, differentiable, and strictly increasing. One can define an upper bound on  $q$  that makes the space in which  $q$  and  $q_{t+1}$  lies ( $Q \times Q_{t+1}$ ) compact, and  $c(m)$  is convex. Following Stokey et al. (2004) this implies that the value function is unique, bounded, continuous, differentiable, and strictly increasing and concave in quality. Therefore, the optimal policy function is single valued and continuous conditional on  $\bar{q}_{t+1}$ .<sup>41</sup>

Combining the general Euler equation equation (5.7) with the derivative of the price function equation (5.18) and the assumed functional form for  $c(m)$  equation (5.13) yields the following Euler equation:

$$(q_{t+1} - (1-\delta)q_t)^{\rho-1} = \beta T \left(\frac{E[Q_{t+1}|\bar{q}_t]}{q_{t+1} + \zeta}\right)^{\frac{1}{\sigma}} \quad (5.19)$$

$$+ (1-\delta)\beta(q_{t+2} - (1-\delta)q_{t+1})^{\rho-1} \quad (5.20)$$

<sup>41</sup>Since many building managers are solving this problem simultaneously,  $\bar{q}_{t+1}$  is not fixed. However,  $\bar{q}_{t+1}$  is treated as fixed when solving each building manager's problem, and handled in a two-step process by equation (5.6).

Where  $T$  is defined as the following combination of parameters for clarity:

$$T = (1 - \omega) \left( \frac{A_l}{a_u} \right)^{\frac{1}{1-\alpha_c}} \left( \frac{\alpha_q(1 - \alpha_c)}{A_m} \right) \alpha_c^{\frac{\alpha_c}{1-\alpha_c}} > 0$$

Comparative statics are discussed in detail in appendix D. Note that the optimal policy  $q_{t+1}$  does not depend on  $d$ , the distance to the CBD, except through the channel of determining whether the property is sufficiently close to the CBD to be habitable (see equation (5.16)), or similarly through credit constraints, which can vary with  $d$  depending on how they are specified.

Combining the functional form assumptions with equation (5.9) yields the following implicit equation for the partial steady state (SS) quality  $q^{SS}$  as a function of a fixed  $\bar{q}_{t+1}$ .

$$\left( q^{SS} \right)^{\rho-1} = Z \left( \alpha_q + \alpha_{\bar{q}} \left( \frac{\bar{q}_{t+1} + \bar{\zeta}}{q^{SS} + \zeta} \right)^{\frac{\sigma-1}{\sigma}} \right)^{\frac{1}{\sigma-1}} \quad (5.21)$$

Where  $Z$  is defined as the following combination of parameters:

$$Z = \frac{T}{(r + \delta)\delta^{\rho-1}} \quad (5.22)$$

Note that the LHS is increasing in  $q^{SS}$  and the RHS is decreasing in  $q^{SS}$ , so equation (5.21) implicitly defines a unique  $q^{SS}$  conditional on  $\bar{q}$ . This steady state level of quality is increasing in neighborhood quality for all acceptable parameter values, demonstrating the expected strategic complementarities. The symmetric steady state (SSS) can be defined explicitly when  $\bar{\zeta} = 0$  by combining equation (5.10) and functional form assumptions:

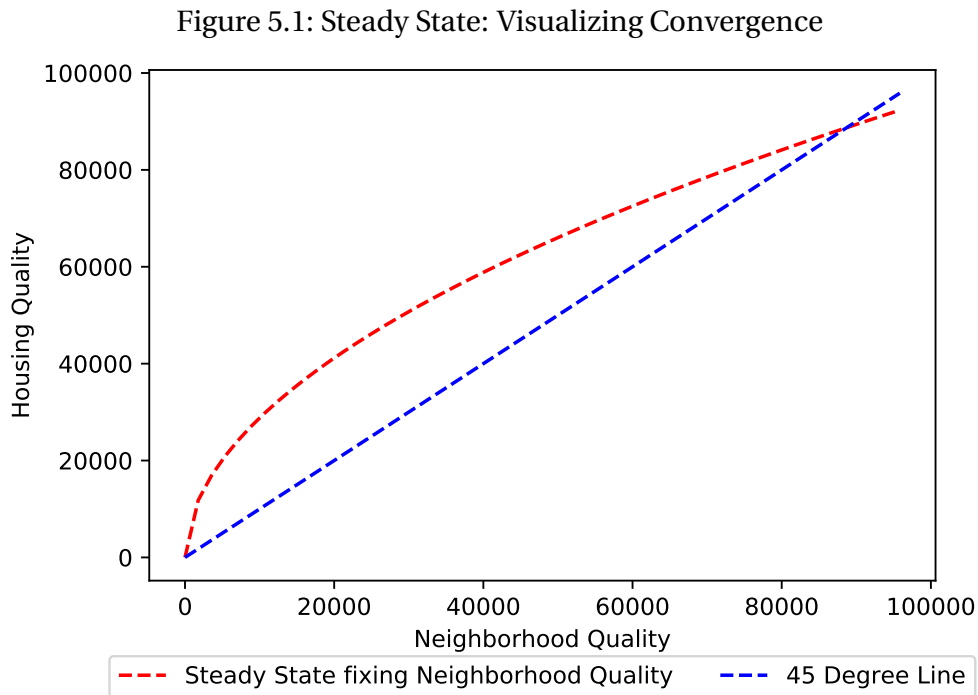
$$q^{SSS} = (Z)^{\frac{1}{\rho-1}} = \frac{1}{\delta} \left( (1 - \omega) \left( \frac{A_l}{a_u} \right)^{\frac{1}{1-\alpha_c}} \left( \frac{\alpha_q(1 - \alpha_c)}{A_m(r + \delta)} \right) \alpha_c^{\frac{\alpha_c}{1-\alpha_c}} \right)^{\frac{1}{\rho-1}} \quad (5.23)$$

The symmetric steady state does not depend on the elasticity of substitution  $\sigma$ ;  $\sigma$  only governs

the rate at which convergence occurs. Furthermore, as proven in appendix D, (i)  $q^{SS} > \bar{q}$  whenever  $\bar{q} < q^{SS}$ , and (ii)  $q^{SS} < \bar{q}$  whenever  $\bar{q} > q^{SS}$ . This implies that the symmetric steady state is the unique, long-run equilibrium when  $\vec{\zeta} = 0$  and in the absence of credit constraints. This steady state is stable, as neighborhoods below this quality level have an incentive to improve their quality, while neighborhoods above this level have incentives to allow their buildings to depreciate. To demonstrate this, note that in the Cobb-Douglas case discussed in appendix D.4 ( $\sigma = 1$ ),  $q^{SS}$  can be solved for explicitly as a function of  $\bar{q}_{t+1}$ :

$$q_{CD}^{SS} = Z^{\frac{1}{(\rho-1)+\alpha\bar{q}}} \bar{q}_{t+1}^{\frac{\alpha\bar{q}}{(\rho-1)+\alpha\bar{q}}}$$

Here whenever  $\rho > 1$ , the power on  $\bar{q}_{t+1}$  is less than 1. This is consistent with the symmetric steady state being the unique equilibrium when all locations are unconstrained and ex-ante symmetric ( $\vec{\zeta} = 0$ ), and is shown in figure 5.1. The symmetric steady state (SSS) is visually where the 45 degree line and  $q^{SS}$  meet.



Note: The parameters used in this plot match the calibration in table 6.1

### 5.2.2. IMPOSING CREDIT CONSTRAINTS

Combining the credit constraint equation (5.11) with the functional form for  $c(m)$  equation (5.13), allows for the definition of  $q_{t+1}^{MAX}$ , the largest  $q_{t+1}$  a household can afford under the constraint.

$$c(q_{t+1}^{MAX} - (1 - \delta)q_t) = (1 - \omega)P(Q, d) \quad (5.24)$$

$$q_{t+1}^{MAX} = (1 - \delta)q_t + \left( \frac{\rho(1 - \omega)}{A_m} P(Q, d) \right)^{\frac{1}{\rho}} \quad (5.25)$$

Furthermore, a location with characteristics  $q$ ,  $\bar{q}$ , and  $d$  is able to maintain its current housing quality if the following condition holds:

$$c(\delta q_t) \leq (1 - \omega)P(Q, d)$$

Where  $Q$  is again defined by equation (5.12). Plugging in functional forms and rearranging, I define a function  $S(q_t, \zeta_t, \bar{q}_t, \bar{\zeta}_t, d)$  such that when  $S(\cdot)$  is positive, the location can afford positive maintenance:

$$0 \leq \left( \frac{A_l}{a_u} \right)^{\frac{1}{1-\alpha_c}} \alpha_c^{\frac{\alpha_c}{1-\alpha_c}} (1 - \alpha_c) Q(q_t, \zeta_t, \bar{q}_t, \bar{\zeta}_t) - \tau d - \frac{A_m}{\rho(1 - \omega)} \delta^\rho q_t^\rho \quad (5.26)$$

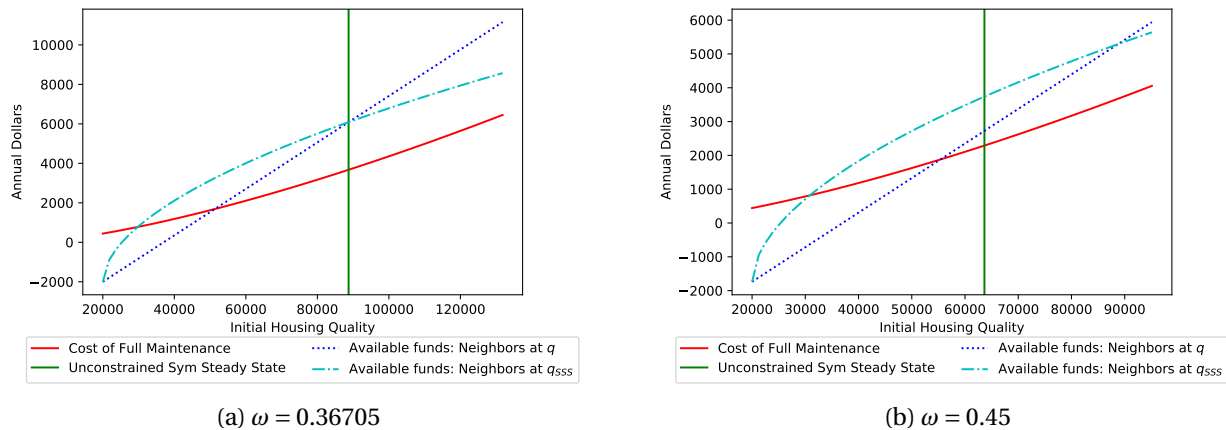
$$= S(q_t, \zeta_t, \bar{q}_t, \bar{\zeta}_t, d) \quad (5.27)$$

Intuitively, this “sustain” criterion is increasing in  $\bar{q}$ ,  $\zeta$ , and  $\bar{\zeta}$  through  $Q$ , because higher levels of external quality imply higher rents and a more slack constraint, but decreasing in  $d$ ,  $\tau$ , and  $A_m$ , which impose costs on the building manager by either increasing maintenance costs or decreasing the rental price. Note that there may be households who are constrained in that they cannot perform optimal positive maintenance, but are still able to retain or slightly improve their housing quality. These households can still converge to the long-run steady state, but will do so at a slower rate.

Figure 5.2 visualizes this criterion for two different values of  $\omega$ . The red line denotes the an-

nual cost of “full maintenance” ( $m = \delta q$ ), while the various dashed lines denote the available funds as a function of initial housing quality after paying mortgage and tax costs. In the blue dotted line, neighbors are assumed to be of equal quality ( $\bar{q} = q$ ), while in the cyan dashed line, neighbors are assumed to be at the symmetric steady state ( $\bar{q} = q^{SSS}$ ). In the right panel, a higher value of  $\omega$  leads to both a lower symmetric steady state, since investment is disincen-  
 tivated, and a higher threshold for households to have enough income to afford full maintenance. From both plots, it’s clear that both the mortgage interest and tax rate  $\omega$  and the quality of one’s neighbors  $\bar{q}$  are critically important for obtaining sufficient rent to avoid being con-  
 strained, particularly in lower quality areas. Though not in the visualization, varying  $\zeta$  is also critically important - locations with higher  $\zeta$  command higher rents, and are more easily able to afford maintenance.

Figure 5.2: Maintenance Funds relative to the cost of full maintenance



Note: The parameters in these plots match the calibration in table 6.1, with  $d = 9.32mi$  and amenities  $\zeta = -20000$ . Neighbors at  $q$  means that the location’s neighbors’ quality tracks the location’s quality. The maximum distance from the CBD  $d$  is 18.7mi for the city of Chicago proper.

These credit constraints can affect the long-run steady state. Specifically, if the model is ini-  
 tialized at quality levels that leave some locations “sustain” credit constrained, these locations  
 will steadily depreciate, and their neighbors will no longer converge to the symmetric steady  
 state because their neighborhood contains locations that are steadily depreciating, rather than  
 converging to the steady state. Instead, these neighbors will converge to an equilibrium below

the symmetric steady state. This effect carries over in turn to their neighbors, meaning that credit constraints that bind at some locations can generate substantial heterogeneity in equilibrium quality across the city. In the long run, the persistence of this constrained equilibrium requires a critical mass of spatially concentrated locations to have initially low quality, leading to a self-perpetuating downwards spiral.

Credit constraints can play a further dynamic role in the feedback channel. Specifically, constrained neighbors are limited in the extent to which they can adjust their maintenance in response to neighborhood improvements - they are only able to adjust insofar as increasing neighborhood quality increases their rents and slackens their constraint. This leads to sluggish development even in areas where rents are high enough to incentivize maintenance. This effect reinforces many of the previous conclusions - lower quality areas have additional barriers to improvement through the limited extent to which spillovers accrue to constrained neighbors.

Overall, the model provides a simple way to think about the dynamic problem that housing investors solve on a yearly basis, and relies on only a handful of parameters to capture many of the primary features of the market. In the subsequent sections, I calibrate the model using a combination of historical data and model equations and apply the calibrated model to simulate the impact of Chicago's opportunity zones from 2018-2027. In appendix G, I also consider a number of other simulations that can be thought of also as robustness, testing the impact of varying a many model parameters relative to the baseline, and what effect these changes have in the short, medium, and long-run.

## 6. MODEL ESTIMATION AND CALIBRATION

From the model, there are three equations that are relevant for estimation: the pricing (inverse demand) equation (5.15), the dynamic Euler equation (5.19), and the credit constraint equation (5.24) at each location  $i$  and time  $t$ :

$$\begin{aligned}
 \bar{P}_t(Q_t, d) &= \left(\frac{A_l}{a_u}\right)^{\frac{1}{1-\alpha_c}} \alpha_c^{\frac{\alpha_c}{1-\alpha_c}} (1-\alpha_c) Q_t - \tau d && \text{Pricing} \\
 (q_{t+1} - (1-\delta)q_t)^{\rho-1} &= \beta T \left(\frac{E[Q_{t+1}|\vec{q}_t]}{q_{t+1} + \zeta}\right)^{\frac{1}{\sigma}} + (1-\delta)\beta(q_{t+2} - (1-\delta)q_{t+1})^{\rho-1} && \text{Euler} \\
 q_{t+1} &\leq (1-\delta)q_t + \left(\frac{\rho(1-\omega)}{A_m} \bar{P}_t(Q_t, d)\right)^{\frac{1}{\rho}} && \text{Credit Constraint}
 \end{aligned}$$

Where:

$$\begin{aligned}
 Q_t &= \left(\alpha_q(q_t + \zeta)^{\frac{\sigma-1}{\sigma}} + \alpha_{\bar{q}}(\bar{q}_t + \bar{\zeta})^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}} \\
 T &= (1-\omega) \left(\frac{A_l}{a_u}\right)^{\frac{1}{1-\alpha_c}} \left(\frac{\alpha_q(1-\alpha_c)}{A_m}\right) \alpha_c^{\frac{\alpha_c}{1-\alpha_c}} > 0 \\
 1 &= \alpha_q + \alpha_{\bar{q}}
 \end{aligned}$$

Estimating the model provides structural estimates for  $\alpha_{\bar{q}}$  and  $\alpha_q$ , summarized by  $\alpha_{\bar{q}}/\alpha_q$ , the relative preference of households for the quality of their neighborhood relative to the quality of their house. This is the main parameter for understanding the strength of investment spillover effects. In the extreme case, if households exhibit no willingness to pay for neighborhood quality ( $\alpha_{\bar{q}}/\alpha_q = 0$ ), then the extent to which neighborhood rejuvenation is reflected in housing prices is limited, and there is no incentive for building managers to invest in housing quality when their neighborhood improves. Though the model could also fit  $\sigma$ , I assume  $\sigma = 1$  for simplicity. In practice, varying  $\sigma$  has no effect on the long-run steady state and only a minor effect on the speed of convergence.

Estimation of the model requires:

- Data:  $(G, P, d, \zeta)$
- Parameters:  $\theta = \{A_l, A_m, \alpha_c, \alpha_u, \tau, \rho, \delta, \beta, \sigma\}$

Which are used to estimate:

- Housing quality:  $\vec{q}_t$
- Relative preference:  $\alpha_{\vec{q}}/\alpha_q$

Importantly, data must be available at a very granular geographic level, to minimize the mismatch between the micro model, in which the unit of observation is the building manager, and the estimation, in which the unit of observation is an average of building managers in an areal unit. In the subsequent section, I describe the tract-level data used for estimation and how the necessary parameters are calibrated, both using exercises outside of the model or in the process of estimating  $\vec{q}_t$  and  $\alpha_{\vec{q}}/\alpha_q$ .

## 6.1. CALIBRATION

First, recall from the empirical section that the spatial weights matrix  $G$  is  $N \times N$ , where  $N$  is the number of locations. I assume that households' willingness to pay for housing depends on their neighborhood following a weighted inverse squared distance metric, with range  $\bar{d}$  miles, defined in equation (4.1). In the baseline, tracts  $j$  above  $\bar{d} = 3$  miles away from tract  $i$  have spatial weight  $g_{ij} = 0$ . Results are not qualitatively affected by varying  $\bar{d}$  between 2.5 and 5 miles.<sup>42</sup>

Granular housing cost data are reported annually in the American Community Survey (ACS) at the tract and block group levels as 5-year averages reported annually, starting with 2006-2010 and continuing annually to 2013-2017 (Manson et al. (2018)).<sup>43</sup> Assessing year-on-year

<sup>42</sup>Below  $\bar{d} = 2.5$ , issues involving tracts having no neighbors arise, as distances are computed from tract centroids.

<sup>43</sup>Block group level prices have extremely large margins of error, so tract-level data is used. Furthermore, at the tract level the ACS constructs a "Median Housing Cost" variable, while at the block group level, this variable must be constructed manually from renter and owner costs, which is possible, but difficult for block groups in which there are few renters or owners. The Census Bureau notes that housing cost data is relatively well-

changes in these prices is imperfect because they include overlapping samples.<sup>44</sup> In particular, the difference between the 2006-2010 and 2007-2011 house price estimates, for instance, is  $(\text{Price}_{2011} - \text{Price}_{2006})/5$ , or the annualized increase in prices from 2006 to 2011, rather than the desired  $\text{Price}_{2011} - \text{Price}_{2010}$ . However, if 2011 prices reflect a departure from prices in the previous five years, both metrics capture this change in comparable units. Prices are adjusted for inflation and city-level house price index (HPI), so that residual changes better reflect changes in housing and neighborhood quality, rather than city-level or national demand and financial forces. ACS housing cost data at the tract-year level can be noisy and imply some unrealistic year-on-year changes that are reverted in the subsequent year. To address these reversions, I employ a light data smoothing procedure described in appendix E, which has almost no impact on the data series (retaining a correlation of over 0.99 with the original series), but a substantial effect on the application of the data to the model, which otherwise struggles to rationalize large price fluctuations.

Distances  $d$  are obtained by calculating the geographic distance from the tract centroid to the center of the city. Estimating the vector  $\zeta$  of non-housing amenities in each tract draws on a variety of data sources. Yelp provides information on local businesses by general and specific categories, including information on the number of local bars, restaurants, shopping, services, and other amenities.<sup>45</sup> Couture and Handbury (2017) construct a measure of transit amenities at the 2010 tract level by estimating the time it takes according to Google Maps to travel to random points at a variety of distances from the initial location. School quality is measured by a variety of school-level surveys done between 2011-2014 and reported by the Chicago Open Data Portal. Data on crime, building code violations, and abandoned buildings are also proxy measures of local street conditions, similar to walkscore measures.

---

estimated, given that the ACS is a mandatory survey, and monthly housing costs are constructed using data from a wide variety of questions on homeowner and renter costs.

<sup>44</sup>See Appendix 4 here. However, the Census Bureau notes that one of the primary goals of this data is to understand changing costs over time.

<sup>45</sup>Yelp data has been found by Folch et al. (2017) to be a relatively good dataset of local amenities, though with coverage bias towards areas with more college-educated workers relative to administrative data.

After compiling these measures of amenities, using them to estimate  $\vec{\zeta}$  requires converting them to units that match the units that define housing quality, effectively assigning an annual dollar value to amenities. As will be described in the subsequent section, the pricing equation provides a way to estimate the sum  $q_{it} + \zeta_i$  given data on prices and values of some model parameters. The values of  $\zeta$  can then be constructed from the extent to which observed amenities explain the time average of the vector  $q_{it} + \zeta_i$ , for instance using the following linear regression:

$$\left( q_i^m + \zeta_i \right) = q_i^m + \sum_{v=1}^V \beta_v a_v \quad (6.1)$$

Where  $q_i^m$  is the average  $q_i$  across years:

$$q_i^m = \frac{1}{T} \sum_{t=1}^T q_{it}$$

$\hat{q}_i^m$  is the residual variation in  $q_i^m + \zeta_i$  after controlling for  $V$  observable amenities  $a_1, \dots, a_V$ . Using this specification,  $\zeta_i$  is defined as the predicted value renormalized to have weighted mean zero:

$$\zeta_i = \sum_{v=1}^V \hat{\beta}_v a_v - \frac{\sum_i \text{Households}_i \zeta_i}{\sum_i \text{Households}_i}$$

In practice, the relationship in equation (6.1) is unlikely to be exactly linear. Results are similar whether this is estimated as OLS following the procedure above, or using a more general random forest. For this calibration, I use the random forest because it both provides a better fit and better handles overfitting when many observable amenities are used.

The primary concern with this two-step approach is that  $q$  and  $\zeta$  are likely to be correlated, as housing and non-housing quality evolve together. If  $q$  and  $\zeta$  are positively correlated, then any specification that explains  $q + \zeta$  using observable components of  $\zeta$  will overestimate the importance of amenities relative to housing quality in explaining total location quality  $q + \zeta$ : a standard omitted variable bias problem. This concern is common in hedonic models that are interested in estimating a residual, and cannot be addressed in the specification without more

explicit data on the quality of housing.

The second concern is that there is no way to identify the extent to which housing or non-housing amenities explain the constant. I assume and normalize  $\zeta_i$  to have weighted mean zero across space, attributing all of the constant value of housing to the housing itself rather than the local amenities. In words, this amounts to normalizing that being homeless in an average part of the city has value 0, and has negative value in lower quality parts of the city (where, for instance, crime may be more prevalent).

To avoid excessive overfitting in this procedure, rather than using the amenities themselves as the explanatory variables  $a_v$ , I instead use principal components of the various measures of local amenities. I find that the first principal component explains 51% of the variation in amenity variables, and that the first six explain 80%. I use these six principal components when estimating  $\zeta_i$ , rather than the full set of 30 local amenity measures. The results are almost identical when only the first two principal components are used. Figure 6.1 shows the first two unrotated principal components.

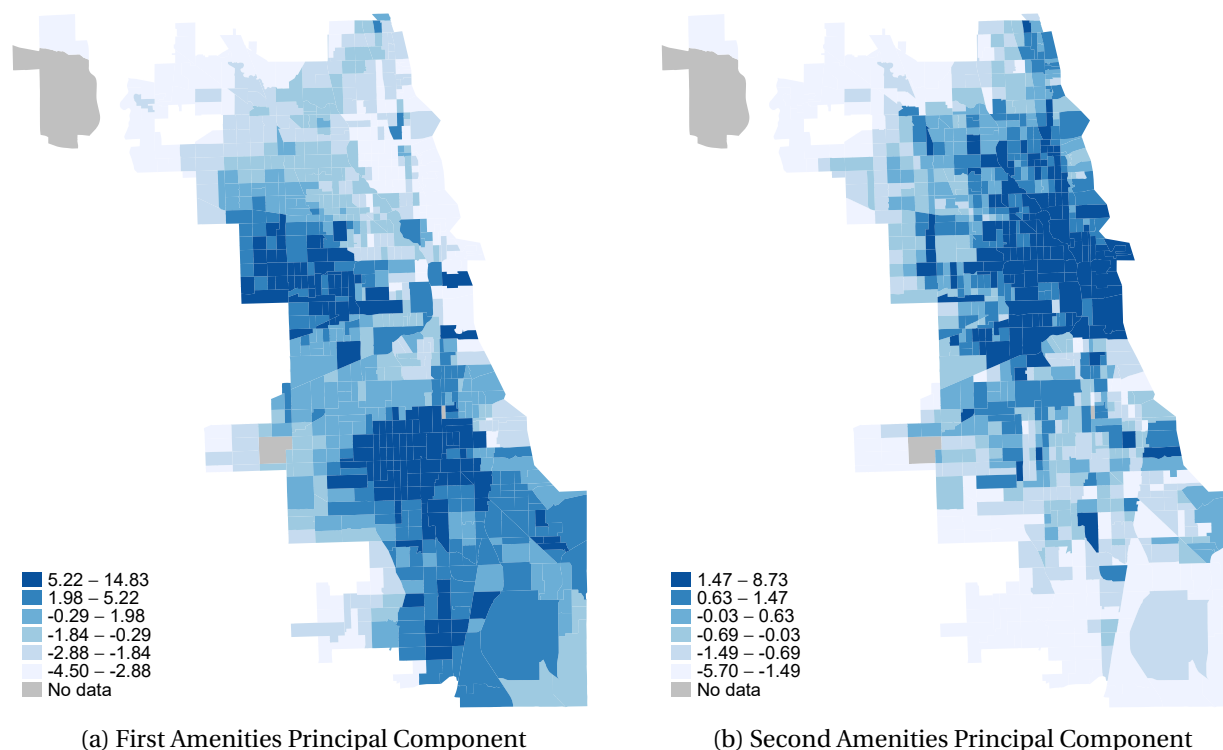
Because of the various difficulties associated with estimating  $\vec{\zeta}$ , I considered robustness exercises in which  $\vec{\zeta} = 0$ , and estimate the impact of varying  $\vec{\zeta}$  in appendix G. Estimating  $\zeta$  with this amenity data and setting  $\zeta = 0$  represent two extremes - the former potentially overestimates the importance of non-housing amenities, insofar as it is correlated with housing quality, while the latter attributes no location quality to non-housing sources, or assumes that non-housing amenities track housing quality exactly, through commercial building managers responding to similar incentives as residential building managers.<sup>46</sup> In practice, although the estimation of  $\vec{\zeta}$  is imprecise, the resulting spatial distribution of non-housing amenities matches the distribution of Chicago's attractions, parks, transit accessibility, and quality schools and restaurants quite closely.

$A_l/a_u$  reflects the productivity of the city relative to the outside option, and is assumed to be

---

<sup>46</sup>Appendix E presents additional information from residential assessments on the spatial distribution of land value relative to assessed building value as it relates to estimating  $\zeta$ .

Figure 6.1: First two amenity principal components, unrotated



one.  $\alpha_c$ , the observable fraction of income households spend on non-housing goods, is set to 0.6 to match the 2015-2016 CEX's fraction of income spent on housing for the average household.<sup>47</sup>  $\beta$ , the discount factor, is constructed as  $1/(1+r)$ , where  $r$  is the annual rate of return on similarly risky investments.<sup>48</sup>  $\delta$ , the rate of depreciation, is partially informed by IRS rules on calculating rental property depreciation. However, these rates take into account basic maintenance: actual rates of depreciation are much higher. From housing cost changes in the ACS and ZTRAX, an annual depreciation rate of 0.3 is required to reasonably explain residual year-on-year changes in prices controlling for inflation and house price trends. This is discussed in more detail in the subsequent estimation of maintenance costs using Zillow ZTRAX data.

<sup>47</sup>In practice, this varies depending on households' wealth, with richer households spending relatively less on housing. The model does not currently accommodate this. Changes in  $\alpha_c$  do not substantially affect results - increases in  $\alpha_c$  scales implied qualities up and slightly weakens the relationship between prices and housing quality.

<sup>48</sup>There is some preliminary macro evidence that the return on housing is high relative to other rates of return in the economy, though that may be due to selection of investment into high return areas.  $\beta$  here is targeted at the general interest rate.

$A_m$  and  $\rho$  define the cost of performing maintenance in each period. Estimating these is both critically important and difficult, since they reflect the cost of obtaining units of quality which are unobserved model objects. These are roughly estimated using data from Zillow ZTRAX on repeat sales of single family homes matched to observed permit expenditures of households. Specifically, I construct sales price differences controlling for inflation and city-level HPI changes. Repeat sales are subset to:

- Sales that occur between 0.5 and 8 years of one another, as a trade-off between having a sufficiently large sample size and minimizing the role of other changes, such as changes in neighborhood amenities or tax rates, that affect sales prices.
- Houses for which implied maintenance is between the 10th and 90th percentile, removing houses with abnormally large sales price differences.
- Houses which are the only unit at their address, to omit apartment buildings or other multi-unit dwellings for which costs and benefits are shared across units.
- Houses for which investment is below the 99th percentile, to minimize the potential for extreme outliers to contaminate the sample, such as houses rebuilt after a fire.

As noted previously, depreciation is estimated as  $\delta = 0.3$ . This value is required to rationalize 90% of low and negative HPI and inflation-adjusted sales prices differences for non-distressed sales. I allow for 10% of negative sales differences to fall below this threshold because some houses may sell for less than their depreciated market value if there is unobserved distress on the part of the seller, or unobserved changes in their neighborhood, which may lead to a decline in price that is larger than would be warranted by a lack of maintenance. Since there is no fixed cost of maintenance, the exact value of delta has very limited impact on the results, and erring on the side of a large delta ensures that almost all households invest a nonzero amount in maintenance, at initially very high marginal returns.<sup>49</sup>

---

<sup>49</sup>Note that changing  $\delta$  affects the calibration of other parameters. These subsequent changes adjust investment costs through parameters  $A_m$  and  $\rho$  in such a way that changing  $\delta$  has a minor effect.

Using the sales price differences and the estimated  $\delta$ , I construct a measure of the implied maintenance associated with the sales derived from equation (5.15), assuming that for each house,  $\bar{q}'$  moves with  $q$ , so the equation does not depend on  $\alpha_q$  and  $\bar{\alpha}_q$ .<sup>50</sup>

$$\text{ImpMaint} = \frac{RC * (\text{Price}_{t+k} - (1 - \delta)^k * \text{Price}_t)}{(\alpha_c)^{\frac{\alpha_c}{(1-\alpha_c)}} (1 - \alpha_c)} \quad (6.2)$$

Where  $RC = 12 * 0.01$  converts sale prices roughly to the model's annual rental prices using the 1% rule.<sup>51</sup>

Comparing this with the observed level of investment (zero dollars for addresses with no permit, the estimated cost of improvements for address with a permit) through nonlinear estimation of equation (5.13) estimates the maintenance cost scale-parameter  $A_m = 0.0001934$  and the slope parameter  $\rho = 1.422794$ . The level of  $A_m$  depends on properly matching sales price differences to implied maintenance in the same units as  $q$  from other exercises. As noted in the empirical section, permits do not capture the universe of home investment, but rather between 6-14%. Assuming that true investment is roughly proportional to observed investment, the estimated  $A_m$  from this nonlinear estimation must therefore be grossed up by a factor between  $100/14 = 7.14$  and  $100/6 = 16.67$ .

In practice, this estimation is used primarily to calibrate  $\rho$ , the convexity of maintenance costs, and to provide a suggestive estimate for  $A_m$ . To confirm the value of the scaling factor  $A_m$ , once the convexity of maintenance  $\rho$  is estimated using ZTRAX,  $A_m$  can be obtained by targeting the 2013 AHS moment that Chicago households report spending \$3,750 per year on average on maintenance and repairs. Assuming that this is a statement about households that are at the

<sup>50</sup>Equivalently, one can think of this as treating neighborhood factors as unobserved error when constructing implied maintenance, as changes in house prices can reflect both positive and negative changes in local amenities. Throughout the modeling exercise, I treat non-housing amenities as stationary in the short term.

<sup>51</sup>As a rule of thumb, rental costs are 1% of home values. This is not a strict rule, but is generally a good approximation. See Investopedia (2018)

implied mean quality  $q^m$ ,  $A_m$  can then be obtained by solving:

$$\frac{A_m}{\rho} (\delta q^m)^\rho = 3750$$

$A_m$  must be targeted within the estimation, as it depends on  $q^m$ , which itself depends on the fraction of quality  $q + \zeta$  is attributed to non-housing amenities  $\zeta$ . In the baseline result,  $A_m = 0.002643319$ , which is 13.7 times the value implied by the ZTRAX approach. This is consistent with the projected scaling factor between 7.14 and 16.67.

Finally,  $\omega$ , the fraction of rent paid in taxes and mortgage payments is calibrated as:

$$\omega = \frac{\text{Monthly Mortgage Interest Payment} + \text{Monthly Tax Payment}}{\text{Monthly Rental Income}}$$

Many of the values in this equation vary in practice from household to household. In particular the mortgage interest payment varies depending on the term of the mortgage and how much of the principal has already been paid. As a result,  $\omega$  will necessarily be an approximation.

Using as a baseline a 30 year, fixed rate mortgage with a 20% down payment and a 4% mortgage interest rate, and Illinois' property tax rate of 2.009%, for a \$160,000 house that by the 1% rule would command roughly \$1,600 in rental value, this calculation yields:<sup>52</sup>

$$\omega = \frac{763.86 * (114,993/274,993) + (160000 * .02009)/12}{160000 * .01} = 0.36705$$

In practice,  $\omega$  can often be different from this value. Taxes are based on assessed value, which are generally lower than actual home sales values because they are often contested (Zorn (2017)). Property taxes and mortgage payments can also be partially or fully deducted by building managers in income tax calculations, further offsetting annual costs. Finally, not all building managers have a mortgage if they have paid off the mortgage in the past, or had the ability to make

---

<sup>52</sup>References: Property Tax Rate and Mortgage Rate. (114,993/274,993) reflects the average amount of each payment that goes to interest relative to principal, the latter of which is not included in monthly costs because it is capitalized into the house.

a different down payment.

Because it incorporates property tax and mortgage rates,  $\omega$  is a central parameter in the model - decreases in the mortgage interest rate that allow households to refinance to lower monthly payments, or decreases in the property tax rate that afford households more funds directly, decrease  $\omega$  and the range of households for which the credit constraint binds.

Finally,  $\tau$  is estimated simply using a regression of observed costs on distances based on the pricing equation, and treating unobserved quality as part of the error:

$$P_{it} = \alpha - \tau d_{it} + \epsilon_{it}$$

This approach attributes all changes in changes in prices to distance rather than quality, and risks overestimating  $\tau$  if  $Q$  is negatively correlated with  $d$ . An alternative approach involves estimating  $\tau$  simultaneously with  $q + \zeta$ . However, the Alonso-Muth condition from the worker's problem (see appendix D) suggests attributing distance-related differences in prices to  $\tau$ . Figure 6.2 depicts the implied dollar value of distance disamenity in annual rental values associated with the estimated value of  $\tau = 0.3161062$ , and suggests a reasonable dropoff in annual housing value further from the city center.

Following the above methodology, table 6.1 summarizes the calibration targets, and the value to which each parameter is calibrated in the baseline.

Figure 6.2: Distance disamenity in dollars per year

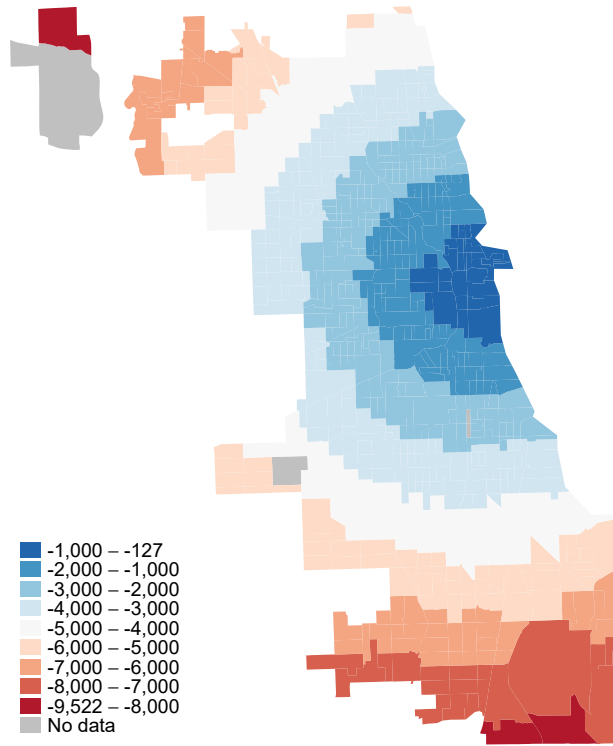


Table 6.1: Baseline Parameter Values

Parameter	Explanation	Value	Target/Source
<b>Workers' Problem</b>			
$\alpha_c$	Consumption Preference	0.6	CEX 2015-2016
$\sigma$	Housing Elasticity	1	Cobb-Douglas
$A_l/\alpha_u$	City vs outside option	1	
$\tau$	Commuting cost	0.3161062	ACS Prices, Alonso-Muth Condition
$\bar{d}$	Neighborhood Preference Range	3mi	
<b>Maintenance Costs</b>			
$A_m$	Scale of Maintenance Costs	0.002643319	ZTRAX targeting, AHS 2013 moment
$\rho$	Convexity of Maintenance Costs	1.422794	ZTRAX targeting, AHS 2013 moment
<b>Managers' Problem</b>			
$\beta$	Discount Factor	0.9713	Interest rate $r = 2.5\%$
$\delta$	Annual depreciation with 0 invest	0.3	ACS and ZTRAX Price Changes
$\omega$	Flow costs	0.36705	Property tax and mortgage interest rates

## 6.2. ESTIMATION STEPS

Estimation of  $\alpha_q$  and  $\alpha_{\bar{q}}$  given the calibrated parameters follows a few steps. First, data on rental prices (housing costs) from the ACS are used to infer quality, using the pricing equation at each location  $i$  and time  $t$ , conditional on a grid of values of  $\alpha_q \in (0, 1)$ :

$$P_{it} = \left( \frac{A_t}{a_u} \right)^{\frac{1}{1-\alpha_c}} \alpha_c^{\frac{\alpha_c}{1-\alpha_c}} (1-\alpha_c) Q_{it} - \tau d_{it}$$

Where

$$Q_{it} = \left( \alpha_q (q_{it} + \zeta_i)^{\frac{\sigma-1}{\sigma}} + \alpha_{\bar{q}} (\bar{q}_{it} + \bar{\zeta}_i)^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}$$

Appendix F discusses the construction of an analytic gradient of the price function used for this inversion, as well as the steps necessary for ensuring convergence. This step yields estimates  $(q_{it} + \zeta_i)$  for each tract  $i$  and year  $t$ . Following equation (6.1), I then estimate  $\zeta_i$  at each location using data on the first six principal components of amenities. Figure 6.3 depicts the implied annual values of local amenities  $\zeta$  from this estimation.

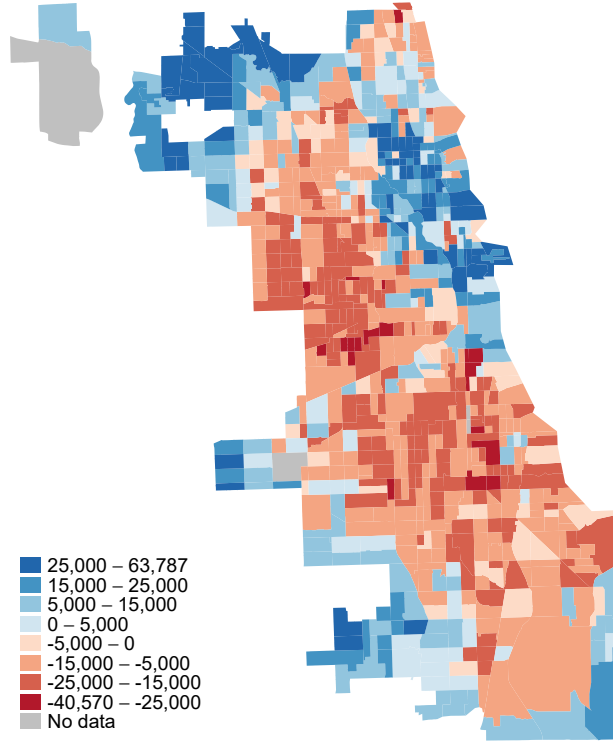
These  $N \times T$  values of  $q_{it}$  and  $B$  values  $\zeta_i$  associated with each grid value  $\alpha_q$  are used to estimate the weighted Euler loss, which indicates how accurately the implied qualities match the intertemporal incentives implied by the Euler equation. This Euler loss is constructed as the following weighted sum of squares:

$$EL(\alpha_q) = \sum_{t=1}^{T-2} \sum_{i=1}^N w_i \left( (q_{i,t+1} - (1-\delta)q_{i,t})^{\rho-1} - \beta T \left( \frac{Q_{i,t+1}}{q_{i,t+1} + \zeta_i} \right)^{\frac{1}{\sigma}} - (1-\delta)\beta(q_{i,t+2} - (1-\delta)q_{i,t+1})^{\rho-1} \right)^2$$

Where weights  $w_i$  are 2010 housing units in location  $i$ , normalized to have mean 1. The estimated values of  $\alpha_q$  and  $\alpha_{\bar{q}}$  are the values on the grid that minimize this loss:

$$\alpha_q^E = \operatorname{argmin}\{EL(\alpha_q)\}$$

Figure 6.3: Model implied amenities  $\zeta$  in quality units



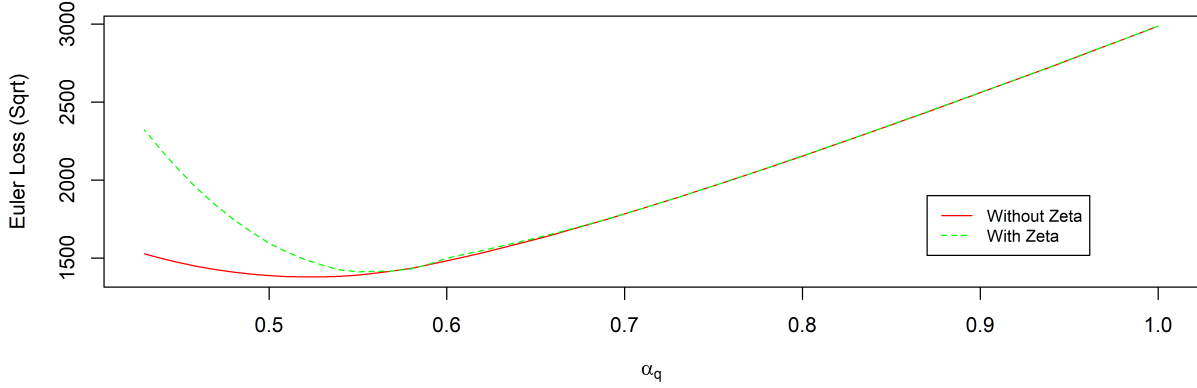
$$\alpha_{\bar{q}}^E = 1 - \alpha_q^E$$

Figure 6.4 plots the smoothed and interpolated Euler losses associated with each value of  $\alpha_q$ , which is minimized in the case of  $\bar{\zeta} \neq 0$  at  $\alpha_q = 0.55$  and  $\alpha_{\bar{q}} = 0.45$ .

### 6.3. SIMULATION

Using the calibrated parameters, as well as the estimated values of  $\alpha_q$  and  $\alpha_{\bar{q}}$  from the previous exercise, it is possible to simulate the model by solving for the optimal policy function  $q_{t+1}(q_t | E[\bar{q}_{t+1}] + \bar{\zeta}, \zeta, \theta)$ , and applying it to building manager's decisions given current conditions. This approach - taking  $E[\bar{q}_{t+1}]$  as given - allows for subsequent exploration of the impact of the rational expectations condition equation (5.6). Myopic building managers will choose the optimal policy assuming  $E[\bar{q}_{t+1}] = \bar{q}_t$ , while perfect-foresight (rational) building managers will iteratively update this expectation until it is correct.

Figure 6.4: Euler loss conditional on  $\alpha_q$



I estimate the optimal policy function using Chebychev polynomials, solving for the following unique optimal policy function over a grid of values of  $(E[\bar{q}_{t+1}] + \bar{\zeta}, \zeta, \theta)$ :

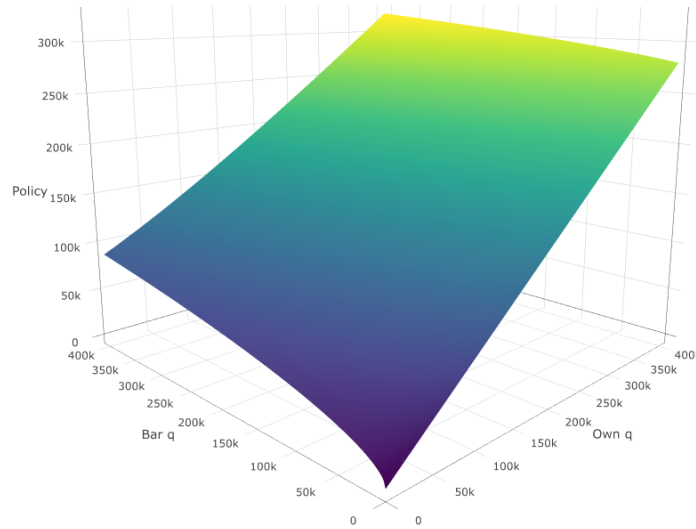
$$q_{t+1}(q_t | E[\bar{q}_{t+1}] + \bar{\zeta}, \zeta, \theta)$$

Figure 6.5 depicts the continuous optimal policy function when  $\bar{\zeta} = 0$ , before the constraint that  $q_{t+1} \geq q_t(1 - \delta)$  is applied. The optimal policy function when  $\zeta \neq 0$  is also estimated with this method, though adding the  $\zeta$  dimension makes it 4-dimensional, and therefore difficult to visualize. However, there is a corresponding 3D plot of the optimal policy function for each value of  $\zeta$  on the grid. Adding a fifth dimension of the function over  $\bar{\zeta}$  separate from  $E[\bar{q}_{t+1}]$  is not necessary due to the model assumptions - the optimal policy depends only on the sum  $E[\bar{q}_{t+1}] + \bar{\zeta}$ .

Simulation proceeds as follows. Using the initial values of  $q_{it}$  for  $t = 1$  and  $\zeta_i$ , optimal policies for all subsequent periods can be calculated using the optimal policy function. In each period, this requires the following steps:

1. Construct  $\bar{q}_{it}$ .
2. Obtain the optimal policy assuming that next period's neighborhood quality is the same

Figure 6.5: Optimal Policy function with  $\zeta = 0$



as the current period's:  $\bar{q}_{i,t+1} = \bar{q}_{i,t}$ .

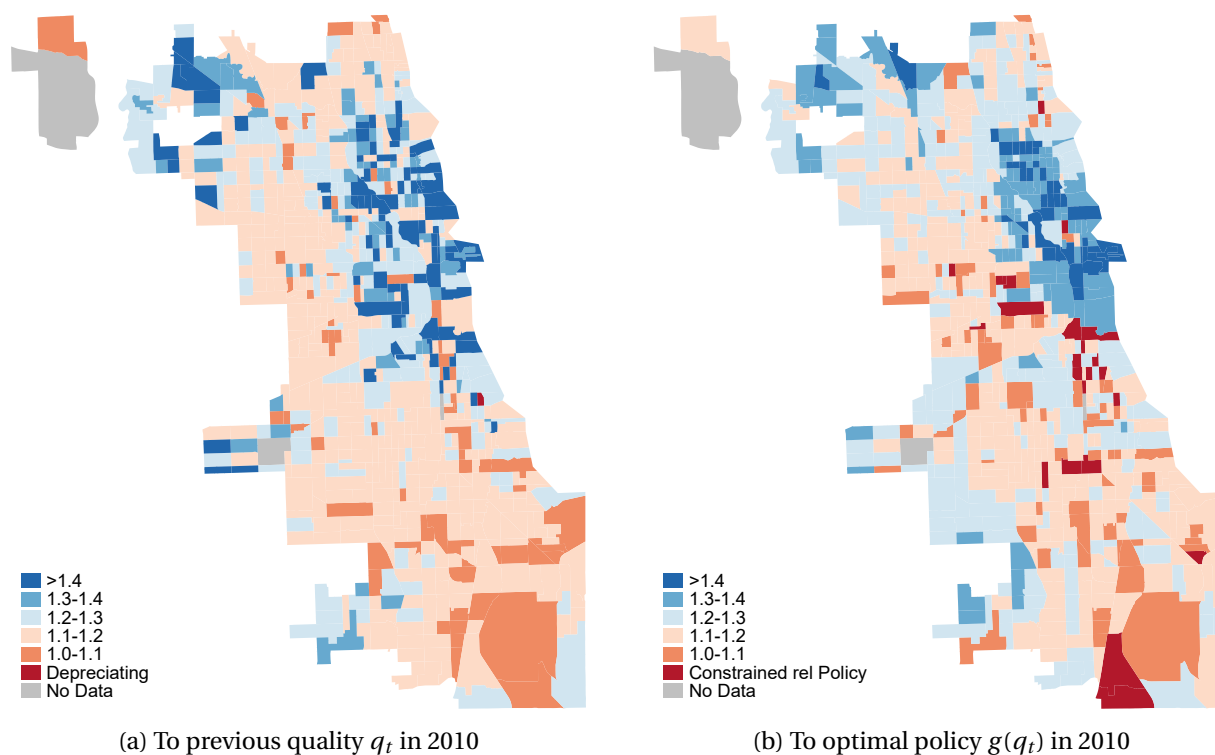
3. Optionally: Apply credit constraints, imposing that households can only invest what they received in the previous period in rents, net of tax and mortgage interest costs  $\omega$ .
4. Optionally: Iterate until rational expectations are met, updating the value of  $\bar{q}_{i,t+1}$  to match the selected optimal policies, and stop once the expectations mismatch is small.
5. Repeat (1)-(4) for subsequent time periods.

These steps yield model-implied values for  $q_{it}$  for  $t = 2, \dots, T$ . Using the pricing equation, these can be converted into model implied prices.

Figure 6.6 shows the model-imposed credit constraints as applied in the first period when choosing quality in the second period  $t = 2011$ . The left panel shows the ratio of the constrained upper bound of  $q_{t+1}$ , divided by  $q_t$ . A value greater than or equal to one implies that maintenance that offsets depreciation is possible, given the constraint. The second plot depicts the ratio of the constrained upper bound to the unconstrained optimal policy. Values below one imply that the building manager is credit constrained, in that she would optimally choose to spend a larger sum on improvements were she not constrained. In later periods, credit constraints

bind in fewer tracts, as previously-constrained tracts increase their housing quality and have a higher rental income with which to afford investment.

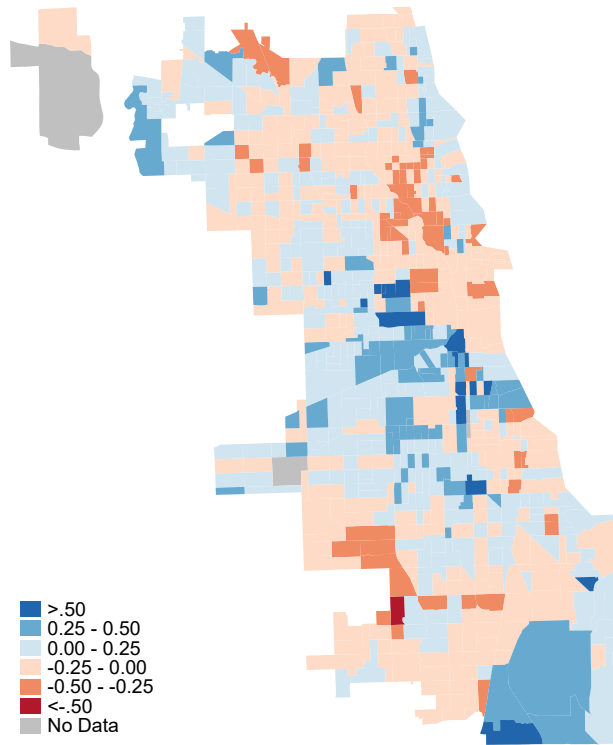
Figure 6.6: Ratios of constrained  $q_{t+1}^c$



The in-sample difference between the growth rate in prices implied by the model and data is presented in figure 6.7. The housing unit weighted correlation in these growth rates is 0.48, with the model implying more investment in the South and West Loop than is observed in prices. In practice, unmodeled menu cost frictions could explain this mismatch, such as practical difficulties for landlords changing prices or finding new tenants to justify this investment, or any other means by which the market for housing is not as frictionless as the model assumes. Within the model, larger values of  $\sigma$  could explain this mismatch as well. As the model is not initialized at or near a steady state, initial periods of the simulation involve substantial investments by building managers as they strive to better match the quality of surrounding tracts  $\bar{q} + \bar{\zeta}$ . Larger values of  $\sigma$  increase the elasticity of substitution between housing quality and neighborhood quality, which would impose weaker incentives on building managers and lead to slower con-

vergence.

Figure 6.7: Difference between model and data growth rate in annual rental cost



## 7. MODEL APPLICATIONS

Using this calibrated model, it's possible to explore the city-wide effects of a variety of policies that are both local (tract-specific) or at the city or national level (parameter-specific). Specifically, there are a few model inputs that can be altered by policies directly, rather than being the result of market forces.  $\omega$  is the primary policy instrument in the model, and is affected by policies or financial forces that affect property taxes paid, the home mortgage deduction, mortgage rates, or any other flow-cost that detract from rents as a percentage. Changes in the tract-specific  $\vec{\zeta}$  can be used to interpret the effects of local policies that improve amenities in parts of the city, such as the building of new parks or attractions, the demolition of abandoned buildings, and other quality-of-life improvements. Finally, changes in the cost of investment or the ease with which investments costs are realized, such as laws about eviction (menu cost), investment credits, permit costs, or building technology improvements can be modeled as changes in  $A_m$  and  $\rho$ .

All of these parameters can be adjusted at the tract level if these changes are not uniform across the city. In the upcoming section, I simulate the effects from 2017-2026 of Chicago's "opportunity zones" relative to the baseline model's predictions. In appendix G, I further consider the effect of building of the Obama Presidential Center, as well as additional hypothetical policies that affect  $\omega$ ,  $\vec{\zeta}$ , and  $A_m$ . In all simulations, I focus on visualizing the effects of these policies in terms of the spatial distribution and magnitude of their effects on short, medium, and long-run house prices.<sup>53</sup>

---

<sup>53</sup>Results are similar for housing quality, but prices are a cleaner market outcome that both reflect a future observable outcome and internalize the full spillover effects.

## 7.1. CHICAGO OPPORTUNITY ZONES

### 7.1.1. BACKGROUND INFORMATION AND MODEL RELATIONSHIP

Among the most direct and relevant applications of the model to policy is the December 2017 bipartisan Congressional legislation to create opportunity zones. Opportunity zones are census tracts selected by state governments and approved by the US Treasury, subject to a per-state quota, where individuals and corporate investors can defer capital gains on investments for a fixed number of years to incentivize investment (US Treasury (2018)). While some city and state governments, such as in Cleveland, use these zones to promote development in areas that are already gentrifying without incentives, Chicago has selected primarily poorer neighborhoods as opportunity zones.<sup>54</sup> In the city of Chicago, opportunity zones were selected to have (i) an unemployment rate of 20% or more, (ii) a median family income of less than \$38,000, and (iii) a poverty rate of 30% or more according to the 2011-2015 ACS (City of Chicago (2018)).<sup>55</sup> Figure 7.1 shows the opportunity zones designated for the city of Chicago, which are concentrated primarily in the poorer west and south sides of the city, as well as updated figures for the selection criteria as of the 2013-2017 ACS. Selected opportunity zones in the greater Chicago area on average have a 41% poverty rate, a 27% unemployment rate, and a median home price of just \$140,000 (CityLab (2019)).

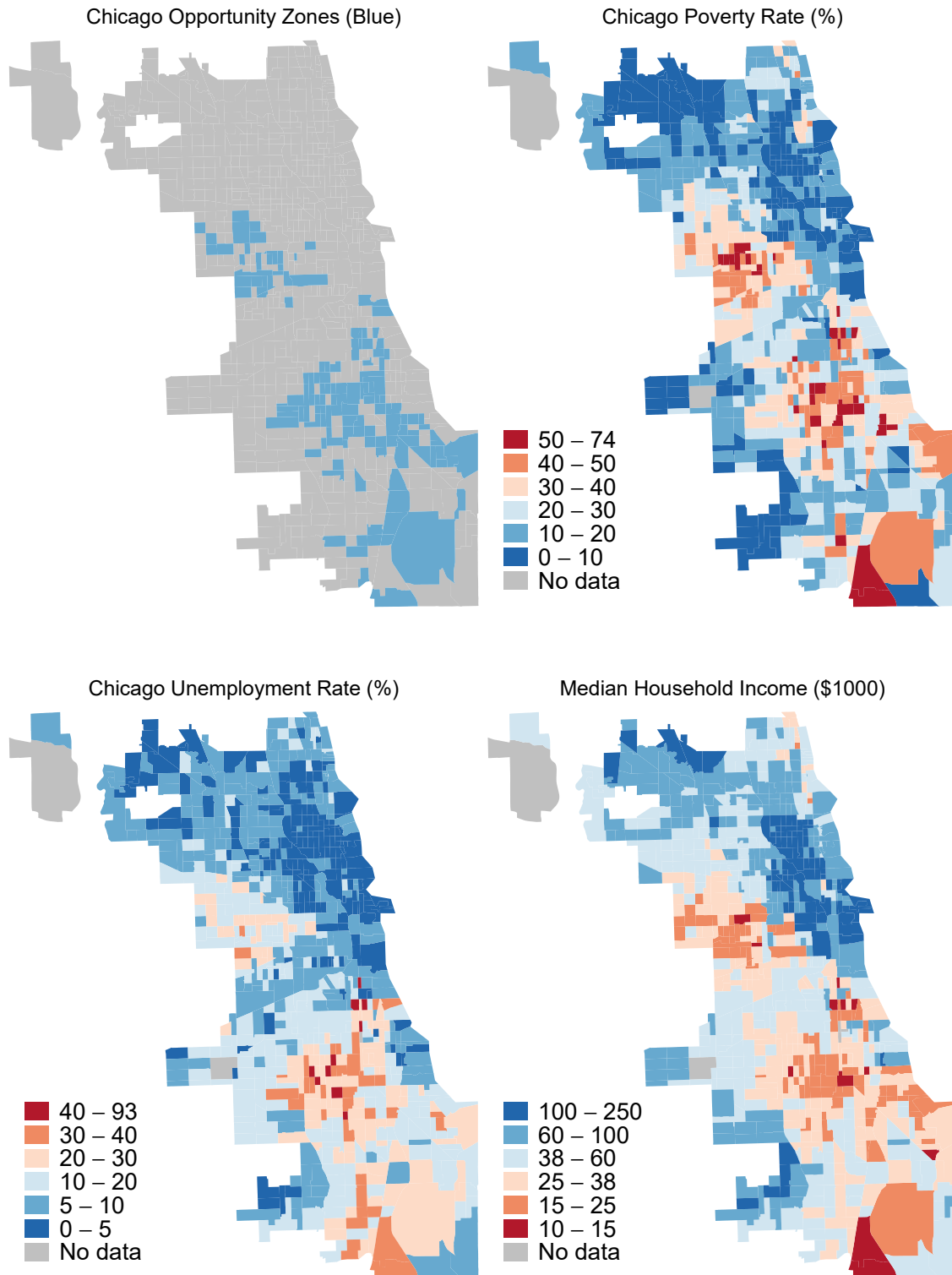
In the context of the model, opportunity zones can be modeled as a decrease in either  $\omega$ , the policy parameter governing property taxes and other flow costs, or  $A_m$ , the scale parameter on the cost of investment. Because they specifically lower taxes on new investment through capital gains tax writeoffs,  $A_m$  is a closer fit for the policy. Existing homeowners who don't invest will see no benefit from opportunity zones in the same way that they would if property taxes in  $\omega$

---

<sup>54</sup>See Cleveland.com (2018).

<sup>55</sup>In practice, these rules do not fully guarantee Opportunity Zone status. The city says these factors were used to select the tracts, but also notes that “The City consulted with aldermen to confirm that the identified tracts within communities with these three qualifying factors were also the tracts that had the most investment potential. In certain limited instances, eligible tracts from within the same community in which economic development activity is underway or which eligible tracts are adjacent to a tract in which development is underway were strategically exchanged.”

Figure 7.1: Opportunity Zones in Chicago with Selection Criteria



Note: Cutpoints selected so that red hues reflect eligible tracts according to selection criteria. City data uses the 2013-2017 ACS. Opportunity Zones were selected according to the 2011-2015 ACS and some level of political discretion.

were reduced.

Capital gains rates on real estate vary depending on the financial and marriage status of the seller, whether or not they live in the home in advance of the sale, and whether they have recently claimed capital gains exemptions on other properties. Many of these factors are outside the scope of the model. However, capital gains taxes and other costs of investment are already a part of the calibrated value  $A_m$  used in the baseline model. Understanding the effect of the policy in the context of the model requires estimating the extent to which capital gains write-offs decrease  $A_m$ . To estimate the largest possible improvement, I note first that capital gains taxes are capped at 20% of the difference between the property value and the cost basis of the property (TurboTax (2018)). Initially, the cost basis of a property is the previous sale price.<sup>56</sup> The cost basis increases if IRS Publication 523 eligible improvements are undertaken between sales. Generally, this requires renovation that is obvious at the time of sale, such as the addition of a bedroom, bathroom, or new kitchen, rather than maintenance, such as cleaning or refurbishing of existing rooms or edifices.

Given this information, one can consider a hypothetical investor whom this policy helps the most: an outside investor who purchases a property, invests in it in a way that does not improve the cost basis, never lives in the home to benefit from tax exemptions, and immediately sells the property to either a landlord who rents it out or an owner-occupier who lives in it. This investor, for example with an initial purchase of \$100,000, a \$80,000 dollar investment, and a \$200,000 sale, would have exactly broken even, paying \$20,000 to the federal government in capital gains and effectively paying \$100,000 for the investment overall. Therefore  $A_m$  would decline for this investor by roughly a factor of 0.2.<sup>57</sup> In reality, most investors will not see this

---

<sup>56</sup>The cost basis can decrease if depreciation is claimed on the property. However, this just results in an intertemporal substitution - losses due to depreciation offset taxes in years where the property is held, but ultimately their reduction of the cost basis leads to a larger capital gains tax paid when the property is sold. Because of this, the tax effects of claiming depreciation are considered a wash in this analysis, and are ignored in the simulation.

<sup>57</sup>In reality, this policy likely affects both  $\rho$  and  $A_m$ , if the extent of investment is correlated with how easy it is to claim changes to the cost basis when filing taxes. However, estimating the impact of the policy on the convexity of the cost curve is difficult, and simply scaling the curve up or down using  $A_m$  provides a reasonable first-order approximation of the cost changes that result from the policy.

large of a benefit. Living in the house, performing renovations that increase the cost basis, or renting the property oneself - and therefore forgoing capital gains tax indefinitely - are all ways to mitigate the initial tax costs and lessens the extent to which this policy decreases investor costs. Therefore, to simulate the policy I consider a range of possible multipliers for  $A_m$  ranging from  $0.85 * A_m$  to  $0.95 * A_m$ .  $0.9 * A_m$  is used as a baseline.

Next, when simulating this policy, it's crucial that  $A_m$  changes at the tract, rather than the city, level. This results in tracts adhering to different optimal policy functions, based on different values of  $A_m$ . Rational expectations equation (5.6) accords these heterogeneous policies period-by-period - households make a decision given their particular parameters and the associated optimal policy function, and the model's simulation iterates until their expectations about their neighbors match reality.

#### 7.1.2. BASELINE RESULTS

In this section, I simulate the impact of the policy on population-weighted housing prices relative to the baseline model projections one, five, and nine years into the policy, and using multipliers for  $A_m$  of 0.85, 0.9, and 0.95 for the opportunity zone tracts. Impacts are calculated as price changes relative to the baseline. At the tract level, this is simply:

$$\text{Impact}_{i,t} = 100\% * \left( \frac{P_{i,t}^{\text{Policy}}}{P_{i,t}^{\text{Baseline}}} - 1 \right)$$

At the city level, this can be calculated as either asset-weighted (AW) or population-weighted (PW). The former uses the total value of housing in the city in the policy and the baseline, and calculates the percentage change from that, while the latter calculates the average change in home value that a random homeowner experiences between the two simulations:

$$\text{Impact}_t^{AW} = 100\% * \left( \frac{\left( \sum_i P_{i,t}^{\text{Policy}} * \text{Households}_i \right)}{\left( \sum_i P_{i,t}^{\text{Baseline}} * \text{Households}_i \right)} - 1 \right)$$

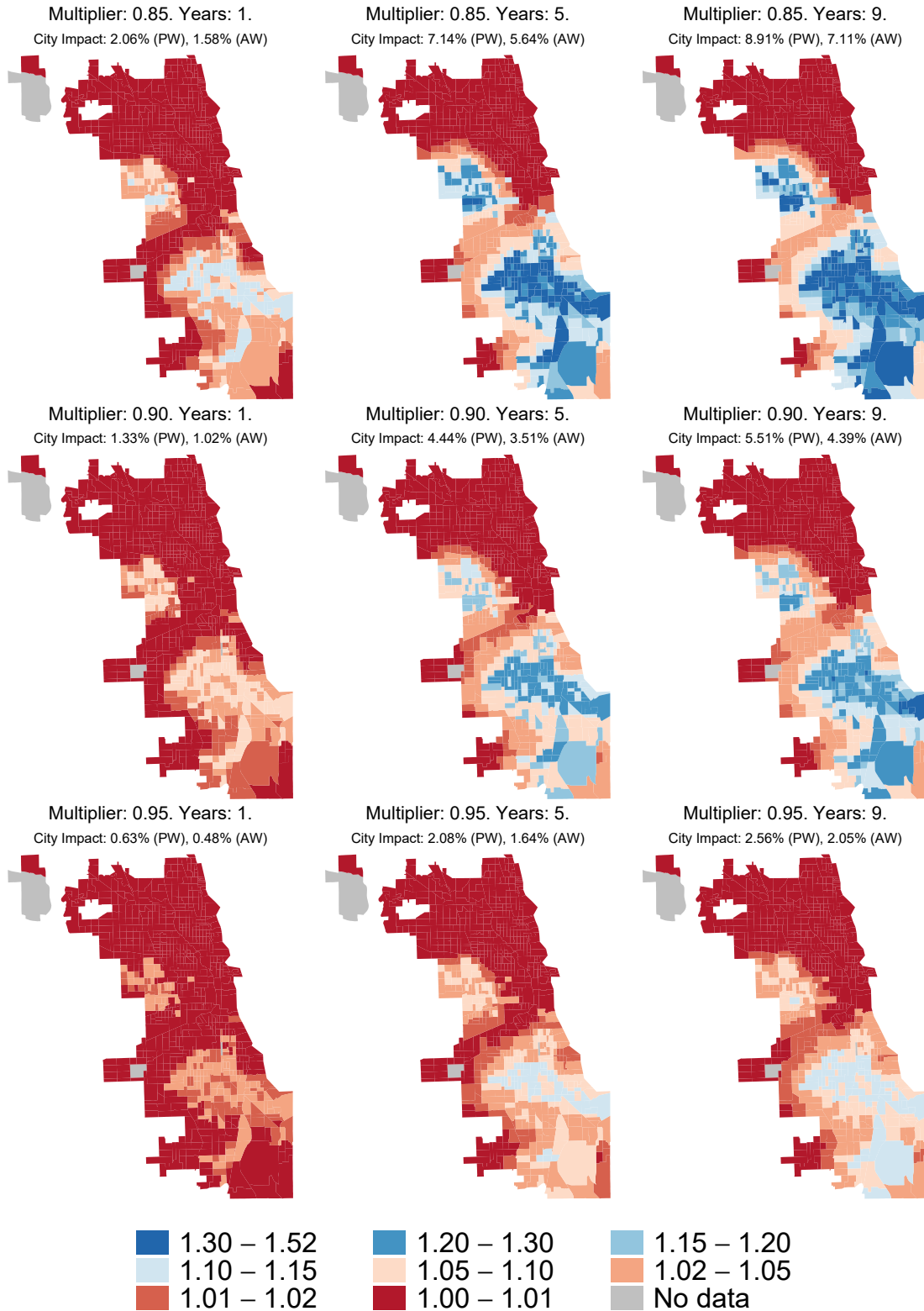
$$\text{Impact}_t^{PW} = 100\% * \left( \sum_i \frac{P_{i,t}^{\text{Policy}}}{P_{i,t}^{\text{Baseline}}} * \text{Households}_i - 1 \right)$$

In the text, I focus primarily on population-weighted outcomes, as these more directly relate to the household-level impact of the policy, and are not biased towards policies that have a large impact on initially high value properties. However, I provide both measures for comparison in visualizations and tables, as the outcome of interest depends on the policy goals.

Figure 7.2 shows that the impact of the policy is substantially larger for smaller  $A_m$  multipliers, with 9-year impacts varying from a city-wide increase of 8.4% in prices for  $A_m = 0.85$  to as low as 2.4% for  $A_m = 0.95$ . For all multipliers, the policy impact is primarily seen in the first 5 years, after which tracts continue to make relatively minor adjustments along the new equilibrium path. One of the highlights of the simulation is the prevalence of spatial smoothing caused by the model's spillover effects. Comparing to figure 7.1, tracts on the south and west side of Chicago that were not selected for the policy experience similar effects to tracts that were selected due to neighborhood improvement and the resulting effect on prices, constraints, and incentives. The entire nine year path, highlighting the exact year-to-year transitions, is shown for  $A_m = 0.90$  in figure H.1.

In addition to affecting  $A_m$ , this policy may also be interpreted as affecting the credit constraint. Recall that the policy primarily provides incentives for outside, non-homeowner, investors to buy, invest, and sell on the property to a local landlord without having to pay extensive capital gains tax. By contrast, owner-occupiers who either plan to live in the home themselves (be the landlord and the renter) or who plan to own the home indefinitely, never paying capital gains on a sale, see almost no benefits. In this sense, the policy “opens up” locations to a wider pool of investors who can invest at tax rates similar to what owner-occupiers may experience, increasing the odds of finding an unconstrained investor. To roughly estimate this effect in the model, in figure H.2 I additionally impose that the credit constraint is not applied in the opportunity zone tracts. Generally, the long-run result is roughly the same, as the constraint does

Figure 7.2: Opportunity Zone Simulation: Price Impact Relative to Baseline



Note: PW is population-weighted price impact

not affect the long-run equilibrium. However, the on-impact effect of the policy is larger, with extremely large gains seen in some areas that were previously constrained in the simulation. In practice, when there are more feedback effects internalized, such as the impact of residential development on commercial investment or other factors that improve  $\vec{\zeta}$ , this stronger initial policy response could have long-term consequences because it may spur concurrent business development.

### 7.1.3. SELECTING OPPORTUNITY ZONES: “RANDOM” AND “SINGLE TRACT” APPROACHES

In addition to simulating the impact of the current policy, the model can also be used for policy evaluation. The primary question that I consider is *“did the city of Chicago select the set of tracts that have the largest aggregate impact on Chicago, or on struggling neighborhoods in the city, subject to political constraints?”* In this section, I consider this question by simulating counterfactuals using a “random” and a “single tract” approach.

First, it's worth noting that the Chicago city government had a large and non-trivial task in selecting a large swath of the city's tracts for this policy. Specifically, if the problem is how to select 136 tracts from 806 eligible tracts - i.e. that there is no leeway in the number of tracts selected in Chicago relative to the rest of Illinois, even though the policy tracts are ultimately selected at the state level - that yields  $2.8e157$  possible choices. Analyzing which of these combinations would have the largest impact, even with a model simulating the effects of each combination of tracts, is computationally infeasible.

A “random” approach to solving this problem proceeds as follows. First, I select 136 tracts randomly out of a set of eligible tracts. Next, I simulate the effect of a policy using those 136 tracts as opportunity zones and construct a summary metric of the impact of the policy. This procedure is repeated  $S$  times to (i) estimate the distribution of possible impacts, and (ii) select the set of tracts that have the largest impact out of the  $S$  random attempts.

I implement a few variations on this general methodology. For the set of opportunity-zone-

eligible tracts (the “selection limits”), I use two approaches: (i) all 806 tracts being eligible, and (ii) just the 278 that are below the median (BM) in the 2017 ACS for the three major selection criteria used by the city of Chicago in their opportunity zone selections (inverse poverty rate, inverse unemployment rate, and household income). For the policy impact (the “targeted tracts”), I focus on medium-run (after 5 years of the policy) population-weighted price impacts, either (i) for the city of Chicago as a whole, or (ii) on the same 278 “below median” tracts. The simulations that limit the selection criteria or targeting to below median tracts reflect political or social pressures faced by the city government. The city may have stronger political incentives and public support to implement the policy that has the largest effect on struggling neighborhoods rather than the city as a whole, or may need to select tracts in these regions for political reasons but ultimately care about the aggregate effect that these selections have on the city. Because these political factors make the city’s utility function and selection criteria uncertain, I report results for all four combinations of the above variations.

For the “single tract” approach, I instead consider the counterfactual impact of using each specific tract in the city as an opportunity zone. Once this is estimated using either impact metric, I select the 136 tracts with the largest individual impacts, and separately simulate the impact of these tracts together as a counterfactual opportunity zone selection. Ultimately, this also has four variations that match the “random” approach, depending on the selection limits and targeted tracts. The single tract approach is more straightforward than the random approach, but provides less distributional information and doesn’t internalize complementary or conflicting effects. It’s almost certain that neither approach will find the “true maximum,” because the computational scale of the problem makes a precise answer unlikely to be found by either approach. However, both provide a feasible way to use the model to inform policy-makers about high-impact tracts, visualize the spatial distribution of opportunity zone selections that produce the highest medium-run effects, and compare a set of selected tracts’ impacts with a distribution of possible effects.

Figure 7.3 and figure 7.4 present the distribution of impacts for the “random” method for the City-City and Below Median-Below Median cases respectively. In the first figure, the opportunity zones selected by the city have almost exactly an average effect on the city as a whole as a random selection of tracts. In the latter figure, when constrained to select from only below median tracts and considering only the impact on these tracts, the city’s selected opportunity zones fall above the average impact, suggesting that for this policy goal, the selection was somewhat effective.<sup>58</sup> In addition to the impact of the city’s choice, these figures give the sense that the distribution of impacts is relatively small - between 4 and 5% in the first figure for city-level impact, and between 10 and 13% for the latter. There are two important things to note. First,  $S = 2000$  and  $S = 1000$  simulations does not come close to representing the full distribution, particularly in the tails, for the  $2.8e157$  possible choices. Second, this highlights the impact of spatial nature of the model, which smooths the impact of investment in individual tracts. Specifically, because the aggregate impact of different tract selections is roughly similar (136 tracts are selected in all simulations), the model translates those into similar outcomes, with heterogeneity driven by each tract’s neighbors and the multiplier effect that investment in the tract would have on surrounding neighborhoods’ investment. This can lead to a tight distribution when averaging across random draws, but does not rule out possible opportunity zone selections in the right tail that are far better than the mean.

Table 7.1 presents the overall results of the nine simulations. The first row are the city of Chicago’s actual selections, as well as their simulated impact on population-weighted city-level prices and prices in below median tracts. The next four rows use the random methodology, with the two selection limits for eligible tracts and two variants for targeted outcomes. The final four rows present the same results using the single tract methodology.

The first result from this table is that the actual tracts selected were outperformed by the best

---

<sup>58</sup>Note that the city’s methodology used the 2011-2015 ACS, while below median for the simulations uses the 2013-2017 ACS. The policy was not implemented until after 2018, which makes the latter a more relevant benchmark, and some of the tracts the city selected were no longer eligible according to the more recent data. This is visualized in the top left panel of figure 7.5

Figure 7.3: “Random” Approach distribution: selecting and targeting all tracts

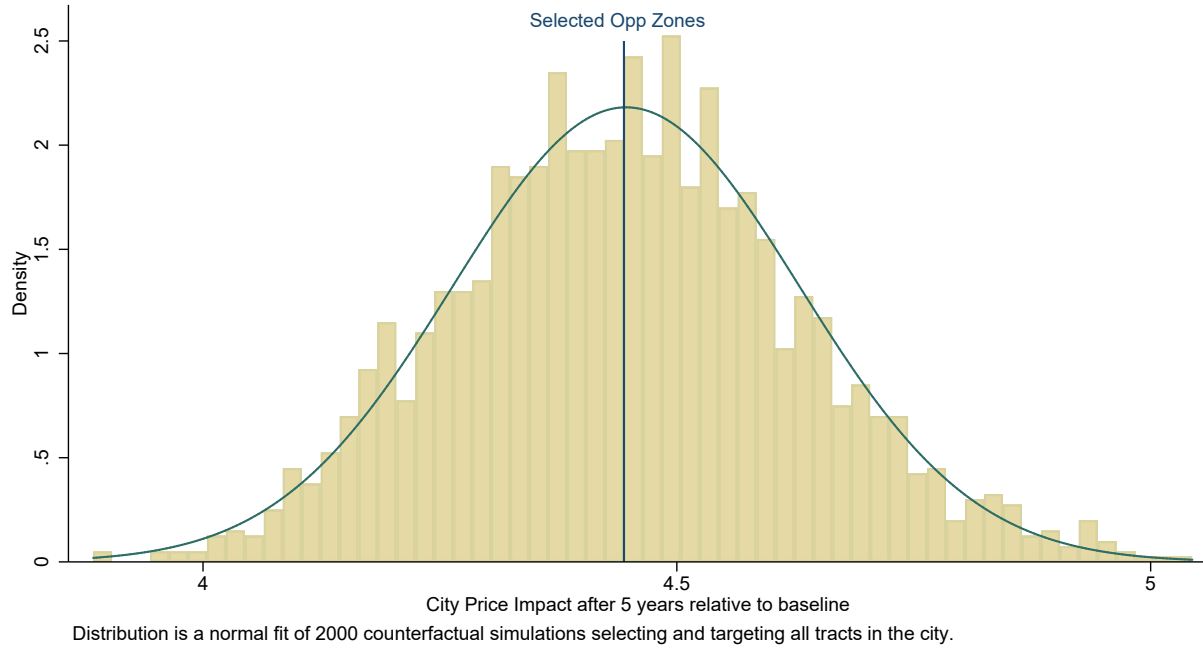


Figure 7.4: “Random” Approach distribution: selecting and targeting below median tracts

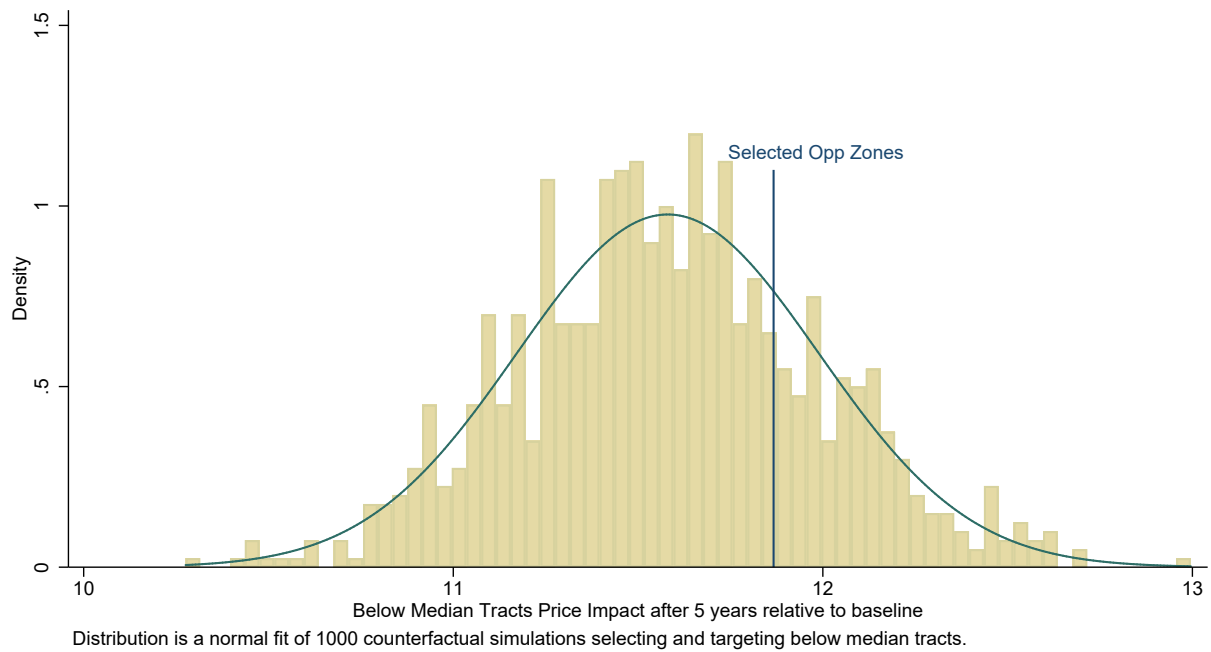


Table 7.1: Price impact of Chicago's selections compared to best counterfactuals

Method	Selection	Target	City PW Imp	City AW Imp	BM PW Imp	BM AW Imp
Actual			4.44	3.51	11.87	11.39
Random	BM	BM	4.99	4.01	12.98	12.58
Random	BM	City	4.99	4.01	12.98	12.58
Random	City	BM	4.63	4.3	6.6	6.48
Random	City	City	5.03	4.92	5.57	5.49
Single	BM	BM	6.38	5.05	17.04	16.38
Single	BM	City	6.52	5.24	16.58	16.07
Single	City	BM	6.62	5.22	17.27	16.55
Single	City	City	8.44	8.21	10.18	9.97

BM refers to tracts that are Below Median according to ACS metrics. PW Imp is population-weighted price impact, AW Imp is asset-weighted price impact.

random simulation when both limited their selection roughly to the below median tracts in terms of their impact on these same tracts (11.87 vs 12.98), and were significantly outperformed by the single tract approach (11.87 vs 17.04 or 16.58). Furthermore, the actual tracts selected were outperformed by all other “best” simulations in terms of city-level impact (4.44 vs 4.99-8.44).

Next, in all cases the single tract method outperforms the random method and highlights an effective combination of tracts far off in the right tail of the distribution. This suggests that the random method is unable to overcome the computational scope of the problem to provide a competitive suggestion. Instead, the best set of tracts is likely found by starting with the “single” tract suggestion, and attempting minor perturbations around this selection, focusing on the top tracts in the single tract opportunity zone simulations that did not make the top 136, until a maximum that internalizes complementarities between initially high-impact tracts is reached.

Finally, both the random and the single tract method find counterfactual opportunity zone selections that outperform the city's choice for the city as a whole and below median tracts simultaneously. The last two rows of the table, which do not limit the selection criteria, but consider different targets, represent the two best counterfactuals, depending on the goal of the policy.

Figure 7.5 and figure 7.6 show the selected tracts and the resulting 5-year impacts for the nine rows of table 7.1. From these figures, there are a wide variety of crucial results.

First, city-level impacts require city-level selections. The best selections that target city-level effects - the middle panel and the bottom right panel - choose a wide variety of tracts outside of the below median tracts, and comfortably outperform policies that target city level effects but limit selection to below median tracts.

Second, the bottom middle panel of the first figure, which has the largest impact on below median tracts, shows that selecting a few high-impact tracts outside of the targeted tracts can be important for achieving the maximum impact even if the policy goal focuses on a selected subset of the city. This result is a direct consequence of the model's spillover effects, and suggests that the counterfactual exercise does a good job of identifying tracts for which the spillover effects are strong enough to outweigh direct effects available through selecting tracts in the target area.

Third, the distribution of impacts in the bottom right panel - which uses the 136 highest impact tracts overall - suggests that certain parts of the city have a higher multiplier than others, likely due to nearby population density. This highlights a weakness of the model in that, by focusing on renovation, it does not internalize new development and population growth that could result from this kind of policy. As a result, it may under-promote tracts that initially have low population in and around them, and therefore low initial impact, even if these tracts have the potential to attract new residents.

Fourth, the random exercise results, particularly in the middle panel, suggest that apart from targeting specifically high impact tracts (as the single approach does), the best first-order approach is to choose a wide variety of spatially distributed tracts in the targeted area so as to maximize the spatial coverage of the impact. This suggests that outside of fundamentally high-impact tracts, there are diminishing returns associated with selecting multiple tracts near one another. Instead, selecting the highest impact tracts in each neighborhood, and relying on

spillover effects can be more effective, if less politically feasible.

Finally, the stark contrast between both selections and impacts depending on the targeting set suggest that it is difficult to achieve multiple goals simultaneously. Though the bottom middle panel does a reasonable job of both having a large city and below median impact, as shown in table 7.1, its effects are still relatively concentrated in the targeted areas. This highlights the importance of having a clear initial policy target - whether that be one of the two used in this exercise, a combination of both, or another target entirely - in order to select the highest impact policy.

Figure 7.5: Comparison of selected tracts to best counterfactuals

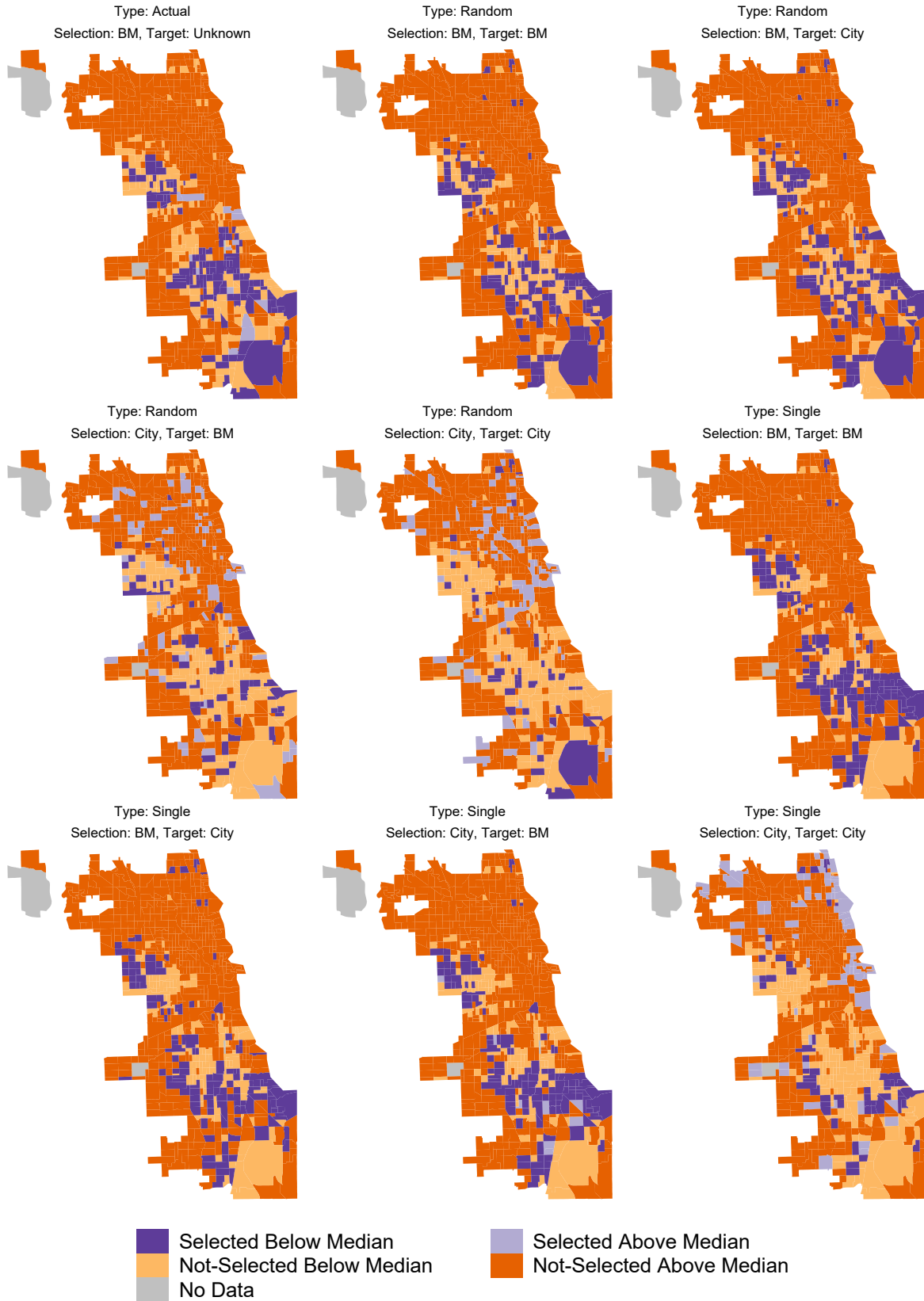
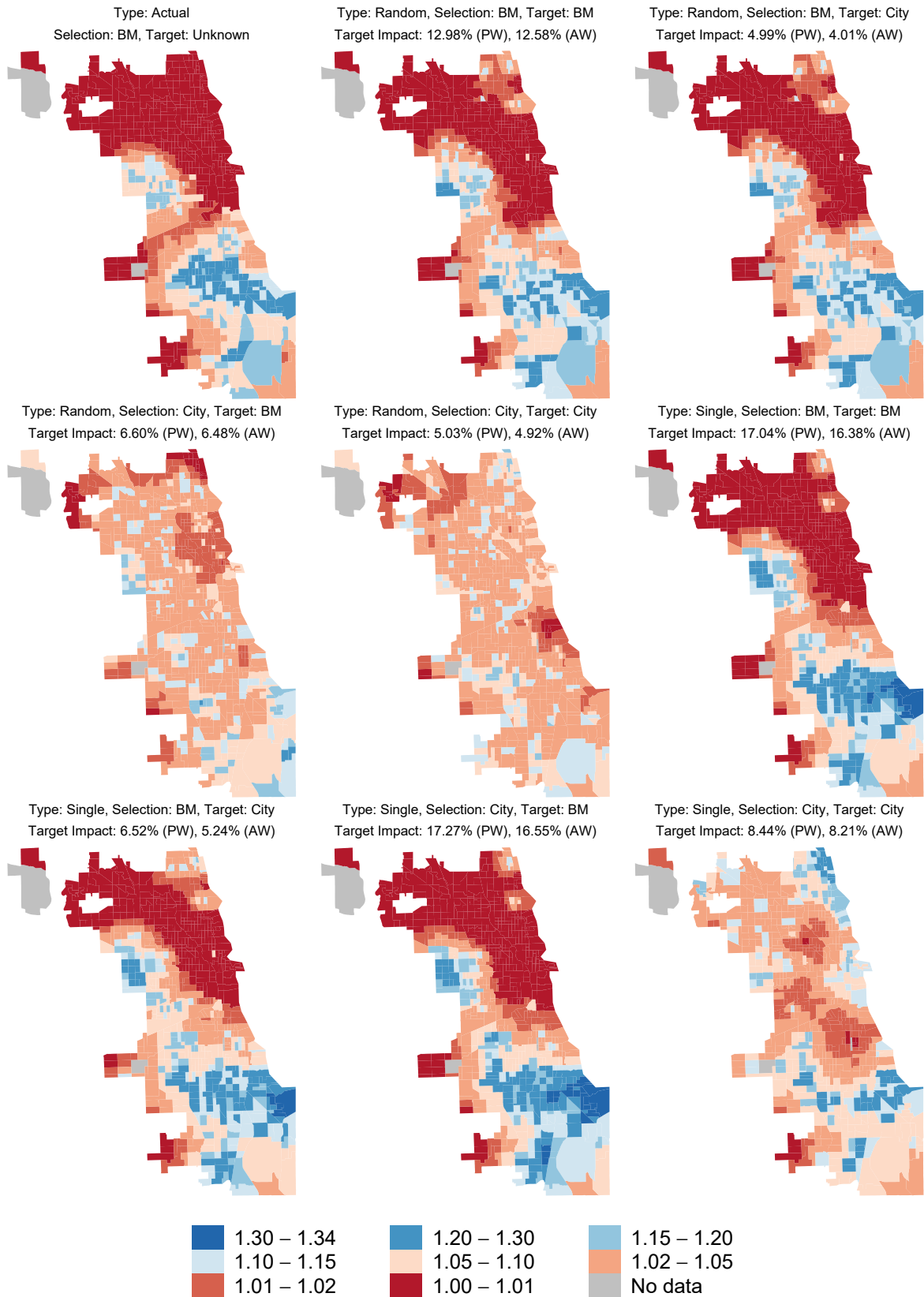


Figure 7.6: Comparison of simulated impact of selected tracts to best counterfactuals



## 8. CONCLUSION

In this paper, I explore the effect of credit supply and neighborhood quality factors on households' and landlords' decisions to invest in the quality of residential buildings. I use data on national lending behavior and housing investment in the city of Chicago to show that changes in the supply of credit affect the level of household investment at the census tract level. I further show that investment decisions are strongly predicted by investment in neighboring tracts, and that these spillovers can amplify the effects of credit supply shocks. Using these insights, I construct a model of dynamic investment, in which incentives to invest in housing quality are driven by households' preferences. I estimate the model, and find that the model fits best when households have roughly an equal preference for the quality of their tract and the quality of surrounding tracts. Simulating the model and comparing the in-sample predictions, I find a correlation of between 0.48 and 0.53 between the model and data, depending on the comparison metric. I then apply the model to simulate the impact of Chicago's opportunity zones on medium-run price growth in the city, and apply a few methods for constructing counterfactual policies subject to a number of constraints that outperform the city's selections. Overall, I find that spillover effects are important enough to be actively considered in housing policy, and that the long term development of individual neighborhoods, and the city as a whole, depend in part on these effects. Finally, I introduce a relatively simple and tractable model that can be used to analyze the impacts of a wide variety of policies that affect local amenities, taxes, mortgage rates, renovation costs, interest rates, and transit effectiveness. Simulations with this model are a great basis for public policy, and can be used to hone and quantify policy ideas at granular spatial levels. To demonstrate this, I apply the model to an analysis of Chicago's opportunity zones, and show how the model can be used to target a variety of policy goals subject to any specified constraints.

## REFERENCES

- ACEMOGLU, D. AND M. K. JENSEN (2013): "Aggregate comparative statics," *Games and Economic Behavior*, 81, 27–49.
- ADELINO, M., A. SCHOAR, AND F. SEVERINO (2015): "Loan Originations and Defaults in the Mortgage Crisis : Response to Mian and Sufi ( 2015a , 2015b )," .
- (2016): "Loan Originations and Defaults in the Mortgage Crisis: The role of the middle class," *Review of Financial Studies*, 29, 1635–1670.
- AHLFELDT, G. M., S. J. REDDING, D. M. STURM, AND N. WOLF (2015): "The Economics of Density: Evidence From the Berlin Wall," *Econometrica*, 83, 2127.
- ALBANESI, S., G. D. GIORGI, AND J. NOSAL (2017): "Credit Growth and the Financial Crisis: A New Narrative," .
- ANAS, A. (1978): "Dynamics of Urban Residential Growth," *Journal of Urban Economics*, 5, 66.
- ANAS, A. AND L. N. MOSES (1979): "Mode Choice, Transport Structure and Urban Land Use," *Journal of Urban Economics*, 6, 228–246.
- ANSELIN, L. (1988): *Spatial econometrics : methods and models.*, Springer economics.
- (2009): "Spatial Regression," *Handbook of Spatial Analysis*, 255–275.
- ANSELIN, L., J. L. GALLO, AND H. JAYET (2008): "Chapter 19: Spatial Panel Econometrics," *The Econometrics of Panel Data*, 625–660.
- ARELLANO, M. AND S. BOND (1991): "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *The Review of Economic Studies*, 277.
- ASIM IJAZ, K. AND M. ATIF (2008): "Tracing the Impact of Bank Liquidity Shocks: Evidence from an Emerging Market," *The American Economic Review*, 1413.

- AUTOR, D. H. (2003): “Outsourcing at Will: The Contribution of Unjust Dismissal Doctrine to the Growth of Employment Outsourcing.” *Journal of Labor Economics*, 21, 1 – 42.
- AUTOR, D. H., C. J. PALMER, AND P. A. PATHAK (2014): “Housing Market Spillovers: Evidence from the End of Rent Control in Cambridge, Massachusetts,” *Journal of Political Economy*, 122, 661–717.
- AVDJIEV, S., U. AYSUN, AND R. HEPP (2017): “What drives local lending by global banks?” .
- BAUM-SNOW, N. (2007): “Suburbanization and transportation in the monocentric model,” *Journal of Urban Economics*, 62, 405–423.
- BAYER, P., K. MANGUM, AND J. W. ROBERTS (2016): “Speculative Fever: Investor Contagion in the Housing Bubble.” .
- BEHRENS, K., G. DURANTON, AND F. ROBERT-NICOUD (2015): “Household sorting in the city,” *Preliminary*.
- BENMELECH, E., A. GUREN, AND B. MELZER (2017): “Making the House a Home: The Stimulative Effect of Home Purchases on Consumption and Investment,” .
- BONACICH, P. (1987): “Power and centrality: a family of measures,” *American Journal of Sociology*, 92, 1170–1182.
- BRAID, R. M. (1981): “The Short-Run Comparative Statics of a Rental Housing Market,” *Journal of Urban Economics*, 10, 286.
- (2001): “Spatial Growth and Redevelopment with Perfect Foresight and Durable Housing,” *Journal of Urban Economics*, 49, 425–452.
- BRUECKNER, J. K. AND S. S. ROSENTHAL (2009): “Gentrification and Neighborhood Housing Cycles: Will America’s Future Downtowns Be Rich?” *Review of Economics and Statistics*, 91, 725–743.
- BUILDFAX (2018): “BuildFAX Permit Records.” <https://www.buildfax.com/>.

- CAPOZZA, D. R. AND R. W. HELSLEY (1989): “The Fundamentals of Land Prices and Urban Growth,” *Journal of Urban Economics*, 26, 295.
- (1990): “The Stochastic City,” *Journal of Urban Economics*, 28, 187.
- CITYLAB (2019): “How to Make Opportunity Zones Work in Chicago,” <https://www.citylab.com/solutions/2019/01/opportunity-zones-tax-benefits-economic-development/579868/>.
- COMBES, P.-P., G. DURANTON, AND L. GOBILLON (2012): “The costs of agglomeration: Land prices in French cities,” .
- COUTURE, V. AND J. HANDBURY (2017): “Urban Revival In America, 2000 to 2010,” *NBER Working Paper*.
- DAVIS, S. J., J. C. HALTIWANGER, AND S. SCHUH (1996): *Job creation and destruction*, The MIT Press.
- DE BARTOLOME, C. A. M. AND S. L. ROSS (2003): “Equilibria with local governments and commuting: income sorting vs income mixing,” *Journal of Urban Economics*, 54, 1–20.
- DEGRYSE, H., O. DE JONGHE, S. JAKOVLEJEVIC, K. MULIER, AND G. SCHEPENS (2016): “The Impact of Bank Shocks on Firm-Level Outcomes and Bank Risk-Taking,” .
- DURANTON, G. AND D. PUGA (2014): “Chapter 5: The Growth of Cities,” *Handbook of Economic Growth*, 2, 781–853.
- (2015): “Chapter 8: Urban Land Use,” *Handbook of Regional and Urban Economics*, 5, 467–560.
- ESTEBAN, R., S. PIERRE-DANIEL, AND I. RAYMOND OWENS (2010): “Housing Externalities,” *Journal of Political Economy*, 118, 485–535.
- FAVARA, G. AND J. IMBS (2015): “Credit Supply and the Price of Housing,” *American Economic Review*, 105, 958–992.

- FEI LEE, L. AND J. YU (2010): “Some recent developments in spatial panel data models,” *Regional Science and Urban Economics*, 40, 255–271.
- FISHER, L. M., L. LAMBIE-HANSON, AND P. WILLEN (2015): “The role of proximity in foreclosure externalities: Evidence from condominiums,” *American Economic Journal: Economic Policy*, 7, 119–140.
- FOLCH, D. C., S. E. SPIELMAN, AND R. MANDUCA (2017): “Fast Food Data: Where User-Generated Content Works and Where It Does Not,” *Geographical Analysis*, 1–16.
- FUJITA, M. (1982): “Spatial Patterns of Residential Development,” *Journal of Urban Economics*, 12, 22.
- FUJITA, M. AND H. OGAWA (1982): “Multiple equilibria and structural transition of non-monocentric urban configurations,” *Regional Science & Urban Economics*, 12, 161–196.
- GABRIEL, J., O. STEVEN, P. JOSÉ-LUIS, AND S. JESÚS (2012): “Credit Supply and Monetary Policy: Identifying the Bank Balance-Sheet Channel with Loan Applications,” *The American Economic Review*, 2301.
- (2014): “Hazardous Times for Monetary Policy: What do twenty-three million bank loans say about the effects of monetary policy on credit risk-taking?” *Econometrica*, 463.
- GERARDI, K., E. ROSENBLATT, P. WILLEN, AND V. YAO (2015): “Foreclosure externalities: New evidence,” *Journal of Urban Economics*, 87, 42–56.
- GIBBONS, S., H. G. OVERMAN, AND E. PATACCHINI (2015): *Spatial Methods*, vol. 5, Elsevier B.V., 1 ed.
- GLAESER, E. L. AND J. GYOURKO (2005): “Urban Decline and Durable Housing,” *Journal of Political Economy*, 113, 345–375.
- GLAESER, E. L., H. KIM, AND M. LUCA (2017): “Nowcasting the Local Economy: Using Yelp Data to Measure Economic Activity,” *NBER Working Paper No. 24010*.

- GLAESER, E. L., S. D. KOMINERS, M. LUCA, AND N. NAIK (2015): “Big Data and Big Cities: The Promises and Limitations of Improved Measures for Urban Life.” .
- GREENSTONE, M., A. MAS, AND H.-L. NGUYEN (2014): “Do Credit Market Shocks affect the Real Economy? Quasi-Experimental Evidence from the Great Recession and ‘Normal’ Economic Times,” .
- GUERRIERI, V., D. HARTLEY, AND E. HURST (2013): “Endogenous gentrification and housing price dynamics,” *Journal of Public Economics*, 100, 45–60.
- GURUN, U. G., G. MATVOS, AND A. SERU (2016): “Advertising Expensive Mortgages,” *Journal of Finance*, 71, 2371–2416.
- GYOURKO, J. AND J. TRACY (2006): “Using Home Maintenance and Repairs to Smooth Variable Earnings,” *The Review of Economics and Statistics*, 88, 736–747.
- HALTENHOF, S., S. J. LEE, V. STEBUNOV, S. JUNG LEE, AND V. STEBUNOV (2014): “The credit crunch and fall in employment during the Great Recession,” *Journal of Economic Dynamics & Control*, 43, 31–57.
- HAUGHWOUT, A., S. SUTHERLAND, AND J. TRACY (2013): “Negative equity and housing investment,” .
- HORNBECK, R. AND D. KENISTON (2017): “Creative destruction: Barriers to urban growth and the Great Boston Fire of 1872.” *American Economic Review*, 107(6), 1365–1398.
- INVESTOPEDIA (2018): “One Percent Rule,” <https://www.investopedia.com/terms/o/one-percent-rule.asp>.
- KATZ, L. (1953): “A new status index derived from sociometric analysis,” *Psychometrika*, 18, 39.
- KELEJIAN, H. H. AND D. P. ROBINSON (1993): “A suggested method of estimation for spatial interdependent models with autocorrelated errors, and an application to a county expenditure model.” *Papers in regional science*, 297–312.

- KHUDABUKHSH, A. (2006): “A Survey on Supermodular Games,” 1–7.
- LANDVOIGT, T., M. PIAZZESI, AND M. SCHNEIDER (2015): “The Housing Market(s) of San Diego,” *American Economic Review*, 105, 1371–1407.
- LEE, S. AND J. LIN (2015): “Natural Amenities, Neighborhood Dynamics, and Persistence in the Spatial Distribution of Income,” *The Review of Economic Studies*.
- LUCAS, R. E. AND E. ROSSI-HANSBERG (2002): “On the Internal Structure of Cities,” *Econometrica*, 70, 1445–1476.
- MÄÄTTÄNEN, N. AND M. TERVIÖ (2014): “Income distribution and housing prices: An assignment model approach,” *Journal of Economic Theory*.
- MALDONADO, W. L. AND H. MOREIRA (2006): “Solving Euler equations: classical methods and the  $C^1$  contraction mapping method revisited,” *Revista Brasileira de Economia*, 60, 167–178.
- MANSKI, C. F. (1993): “Identification of Endogenous Social Effects: The Reflection Problem,” *The Review of Economic Studies*, 531.
- MANSON, S., J. SCHROEDER, D. V. ROPER, AND S. RUGGLES. (2018): “IPUMS National Historical Geographic Information System: Version 13.0 [Database],” <http://doi.org/10.18128/D050.V13.0>.
- MELZER, B. T. (2017): “Mortgage Debt Overhang: Reduced Investment by Homeowners at Risk of Default,” *Journal of Finance*, 72.
- MIAN, A., K. RAO, AND A. SUFI (2013): “Household Balance Sheets, Consumption, and the Economic Slump,” *Quarterly Journal of Economics*, 128, 1687–1726.
- MIAN, A. AND A. SUFI (2014): “What Explains the 2007-2009 Drop in Employment?” *Econometrica*, 82, 2197–2223.
- (2018): “Credit Supply and Housing Speculation,” .

- MIAN, A., A. SUFI, AND F. TREBBI (2015): “Foreclosures, House Prices, and the Real Economy,” *Journal of Finance*, 70, 2587–2634.
- MIAN, A., A. SUFI, AND E. VERNER (2017): “How Do Credit Supply Shocks Affect the Real Economy? Evidence from the United States in the 1980s,” *SSRN Electronic Journal*.
- MONDRAGON, J. (2017): “Household Credit and Employment in the Great Recession,” .
- MUNROE, D. J. AND L. WILSE-SAMSON (2013): “Foreclosure Contagion: Measurement and Mechanisms,” *Mimeo*.
- CITY OF CHICAGO (2018): “Opportunity Zones,” [https://www.chicago.gov/city/en/depts/dcd/supp\\_info/opportunity-zones.html](https://www.chicago.gov/city/en/depts/dcd/supp_info/opportunity-zones.html).
- CLEVELAND.COM (2018): “Opportunity Zones in Cleveland,” [http://realestate.cleveland.com/realestate-news/2018/04/opportunity\\_zones\\_raise\\_hopes.html](http://realestate.cleveland.com/realestate-news/2018/04/opportunity_zones_raise_hopes.html).
- US TREASURY (2018): “CDFIF: Opportunity Zones,” <https://www.cdfifund.gov/pages/opportunity-zones.aspx>.
- ROSENTHAL, S. S. (2014): “Are Private Markets and Filtering a Viable Source of Low-Income Housing? Estimates from a ‘Repeat Income’ Model,” *American Economic Review*, 104, 687–706.
- SAIZ, A. (2010): “The Geographic Determinants of Housing Supply,” *Quarterly Journal of Economics*, 125, 1253–1296.
- SIODLA, J. (2015): “Razing San Francisco: The 1906 disaster as a natural experiment in urban redevelopment,” *Journal of Urban Economics*, 89, 48–61.
- STOKEY, N., R. E. LUCAS, AND E. C. PRESCOTT (2004): *Recursive Methods in Economic Dynamics*, Harvard University Press.
- STROEBEL, J. AND J. VAVRA (2014): “House Prices, Local Demand, and Retail Prices,” .

- TURBOTAX (2018): "Tax Law for Selling Real Estate," <https://turbotax.intuit.com/tax-tips/home-ownership/tax-law-for-selling-real-estate/L9PmDNkK5>.
- TURNBULL, G. K. (1988): "Residential Development in an Open City," *Regional Science and Urban Economics*, 18, 307–320.
- WHEATON, W. C. (1982): "Urban Residential Growth under Perfect Foresight," *Journal of Urban Economics*, 12, 1.
- ZILLOW (2017): "ZTRAX: Zillow Transaction and Assessor Dataset." <http://www.zillow.com/ztrax/>, zillow Group, Inc.
- ZORN, E. (2017): "Why it's foolish not to appeal your property tax assessment," *Chicago Tribune*.

## A. DATA APPENDIX

### A.1. CHICAGO OPEN DATA PORTAL

In recent years, detailed information on cities has become available through “open data portals,” which allow cities to disseminate public data used by their internal agencies in real time. The Chicago Open Data Portal is among the most developed of these data sources, with a wide variety of city administrative data available starting as early as 2002. Specifically, point-level crime data exists starting in 2002. Many other data sources, including some that I use, start only in 2005. Some information for these data sources is obtained via the Freedom of Information Act (FOIA), though often earlier years’ data has continuity issues, such as system or record keeping changes, that govern why historical data were not published on the open data portal.

As a free data source, collected and reported “as-is” and in real time, Chicago’s open data requires some cleaning. In all datasets, some of the data is reported as addresses, without latitudes or longitudes. To match these to census designations, I use the Google Maps API to match addresses to coordinates, and then project the coordinates onto census boundaries. This appendix explains some of the data cleaning procedures in detail.

#### A.1.1. BUILDING PERMITS

Before substantial cleaning, there are a total of 476,827 permits in the data associated with easy permits (minor improvements), renovation (major improvements), new construction, porch construction, and demolition. Using natural language processing, I manually analyze the project descriptions of permits to remove those that are clearly non-residential. Specifically, I look at all words that show up over 50 times in permits, and flag those that are associated with businesses, restaurants, schools, and events. I remove permits including these words from the dataset. This leaves 459,963 permits in the data, which is still an overstatement of the number of residen-

tial permits. I perform a second step using Zillow ZTRAX housing assessment data, matching permits to residential assessments at the address level, and removing those that fail to match. This match is nearly perfect, as both sources of addresses are originally based in the same administrative record. However, I allow for some minor amount of error in the matching using Stata's `relink2` - requiring a match on 2010 Census Block Group and street number, but allowing very minor variations in street name, such as "SAINT" and "ST" as a common prefix. This leaves 270,119 residential permits, of which 157,362 are easy permits, renovation, and porch construction.

Investors who know they are required to submit a permit but want to avoid or mitigate the city permit fee have incentives to under-report the estimated cost. Others list the estimated cost as being \$1, ostensibly because they plan to do the work primarily on their own, despite the fact that this is not a reasonable estimate of the economic cost of the work. To accurately assess the level and spatial distribution of housing investment, I correct these figures by running a series of lasso regressions, by permit type, of the estimated cost on the total fees and the words in the permit description, using the interior of the data and selecting the model that gives the lowest cross-validated error.<sup>59</sup> I then use these predictions to estimate the costs of all permits whose estimated cost is listed as being below the 5th percentile for that permit type, or a 100 dollar cutoff, whichever is larger.<sup>60</sup>

#### A.1.2. OTHER CHICAGO DATA

The building code violations data contains over 1.63 million violations from 2006-2016 at the point level. Each observation denotes the date of the violation, the violation type, the department in the city government that is associated with documenting the violation, the inspector's

---

<sup>59</sup>The interior is defined as the max of the 5th percentile and 100 dollars to the 80th percentile. I bias the interior to smaller projects because the lasso regression will be used to estimate the cost of generally lower cost projects - those that are listed as being in the 0-5th percentile of estimated cost.

<sup>60</sup>For robustness, I maintain versions of the cost variable with (i) no predicted costs, and (ii) fully predicting the cost of all permits, in case there is noise not just in the lower tail. These alternatives have no qualitative effect on the final results.

information, any comments the inspector may have had, and the precise location in either address or latitude and longitude of the violation.

Building code violations can be quite diverse, and range from things that clearly reflect negligence on the part of the building owner to those that are merely mundane problems, such as lacking the required documentation for renovations, or not providing access to the code enforcement officer on a specific day. An example of the latter is “No response/entry at the time of inspection.” Of the 1359 violation codes, I remove 147 for this reason. After this process, there are 1.29 million violations remaining.

Inspections leading to violations can also occur for multiple reasons, including complaints, periodic inspections, permit inspections, and registration. I do not actively drop any of these types of inspections, as all four types can be directly related to maintenance issues. Complaints can come from building tenants or passersby, and are generally about physically noticeable maintenance issues, while permit issues often have to do with more opaque maintenance concerns, such as the maintenance of elevators, ventilations, and air conditioning systems. Complaints are over 70% of the dataset, and are the primary means by which building inspections result in violations. Periodic inspections are substantively similar to complaints, but do not require a call from a citizen. Importantly, there are many periodic inspections that are not in this dataset because they did not result in a violation. In this sense, periodic inspections help stabilize the dataset relative to a concern that complaints may be higher in certain locations where citizens are more prone to complain to the local government. Finally, registration complaints are rare, but are almost always about the need for the landowner to register the property as a vacant or abandoned building.

The abandoned housing dataset consists of 57,290 311 calls reporting abandoned buildings from 2010-2016. I use calls reporting abandoned buildings in lieu of abandoned building violations because this data better reflects community awareness of abandonment and provides a better sense of the duration of abandonment. With explicit violations, a particular location

could be written up as abandoned once, and then remain abandoned without receiving another visit from building inspectors. By contrast, calls reporting a building that continues to be abandoned would not cease. Abandoned housing data is limited in that it starts in 2010, is relatively sparse, and is more prone to measurement error concerns given the nature of its collection.

The crime dataset consists of over 6.1 million observations from 2001-2016, which I subset to the period from 2006-2016, leaving 3.8 million observations. I specifically focus on information on primary offense types, such as narcotics, robbery, and assault as a measure of neighborhood disamenities.

Finally, I use 2010 Census data on the City of Chicago through NHGIS (Manson et al. (2018)). Specifically, I obtain population, racial composition, numbers of housing units, and the distribution of renters and property owners down to the block level.

## A.2. HMDA DETAILS

The Home Mortgage Disclosure Act (HMDA) data is a dataset comprising the universe of home mortgage loans, as reported by loan originators. The dataset is the ongoing product of a law passed by Congress in 1975 as a result of public concerns about credit shortages in some urban neighborhoods that may be due to unfair lending practices. Annual data are provided online, both for recent years at the FFIEC website and historically at the National Archives. HMDA provides loan-level information about loan applications, both accepted and rejected about loan size, loan type, lien status, census tract, MSA, applicant and co-applicant ethnicity, race, and sex, whether the loan was sold to another institution within the year, and other, optional, reporting information.

HMDA data is reported by all loan originators, such as commercial banks, credit unions, savings and loan associations, and bank holding companies (BHCs). Alongside the data, HMDA provide a reporter panel, which includes information on reporting entities' Respondent IDs, any related Replication Server System Database (RSSD) IDs, and information on the entity's parent within the year. This data is incomplete, and only lists a single parent. In the case of large banks, such as JP Morgan Chase, this can result in a subsidiary of JPMC being listed as a parent, despite the fact that the lender is ultimately owned by a larger entity. For the purposes of my analysis, I require that entities that are subject to the same, company-wide lending policies to be denoted as the same entity, so a single parent, that may not be indicative of ultimate ownership and control, is insufficient. The reporter panel is, however, useful in linking entities that are not federally regulated with a parent who is, which provides an RSSD ID that can be used to find the ultimate parent elsewhere.

To find entities' ultimate parent, I supplement the HMDA reporting panel with the National Information Center's relationship table. The relationship table reports all entities' parents, and the structure of ownership.<sup>61</sup> Using the HMDA reporter panel and the NIC tables, I run an it-

---

<sup>61</sup>This includes information such as whether the parent controls the offspring, the percentage of equity voting control, the form of ownership, etc.

erative algorithm that matches entities to their ultimate parent by working upwards through the tree to the root, defining solid links as those that involve the parent being listed as controlling the subsidiary and one of either (i) 50% equity share, or (ii) a relationship level listed as “direct.” This approach is both conservative and aggressive, but closely follows from the goal of matching entities to parents when they share common business policy. Requiring control is severely limiting, but important for ensuring that parents are not simply entities that control a large equity share, but have no say in the management of the subsidiary. By contrast, allowing for control and a direct relationship to be sufficient in lieu of a controlling equity share is inclusive, but again highlights that this matching is intended to match parents to subsidiaries when they operate as a single entity in terms of business policy.

Finally, I recast all relationships at an annual level. NIC relationships are reported at a daily frequency, but HMDA data are annual. I construct ultimate parents at the monthly level, and then define ultimate parents for the year as those that own a given subsidiary entity for a majority of months.<sup>62</sup> Importantly, I do not merger-adjust the data using the NIC transformations table, but instead define ultimate owners at an annual frequency. This is critical: entities that merge in 2008 are not the same in 2007, and as a result may operate according to very different strategies and policies. It is more accurate to keep these entities separate, and note that one stops operating in 2008, otherwise fixed effects that span both of these years would imply incorrect assumptions about the entity’s strategy.

Using this definition of entities’ ultimate parents, I construct two aggregated versions of the data at the county-year-parent level and at the tract-year-parent level. Because the data span from 2004-2015, the definition of counties changes across time according to FIPS 2000 and FIPS 2010 classifications. For county-level data, which changes substantially less across time, I incorporate changes to county definitions to make consistent time series for counties starting with FIPS 1990 by merging counties that were split, and matching counties whose names or

---

<sup>62</sup>For a majority of entities that have a complicated ownership structure, their ultimate parent is often the same, and reached through multiple separate branches of the tree, so this aggregation to annual data is often less important than it seems.

borders were changed. For tract-level data, I run a similar, but more involved, process that uses the Census 2000-2010 relationship files to match tracts that are sufficiently similar across time. Specifically, I first remove all tract-tract relationships that involve little to no change in affected households, either because they involve a minor border change, or involve the changing jurisdiction of uninhabited land. I then aggregate tracts as necessary to make a “concorded” census tract definition that allows for a matching from both FIPS 2000 and FIPS 2010 codes. This can lead to some concorded tracts that are large because they comprise many FIPS 2000 and FIPS 2010 tracts in areas that see substantial changes to census boundaries. While this is unfortunate, nationally it has a minor effect on the number of tracts.<sup>63</sup> Furthermore, this approach has the benefit, over probabilistic matching, of concurring FIPS versions across time in a way that does not risk incorrectly matching households to the wrong census tract when the designation changes.

---

<sup>63</sup>There are 65894 tracts in 2000, 73315 tracts in 2010, and 63455 concorded tracts in the US.

### A.3. TOP 10 ENTITIES IN CHICAGO - MARKET SHARE BY TRACT IN 2006

The figures in this section provide the market share of the ten largest Chicago lenders in 2006, to give a sense of the spatial heterogeneity of their lending. Lenders outside of the top 10 are often even more spatially concentrated in their lending, with a couple notable examples also included.

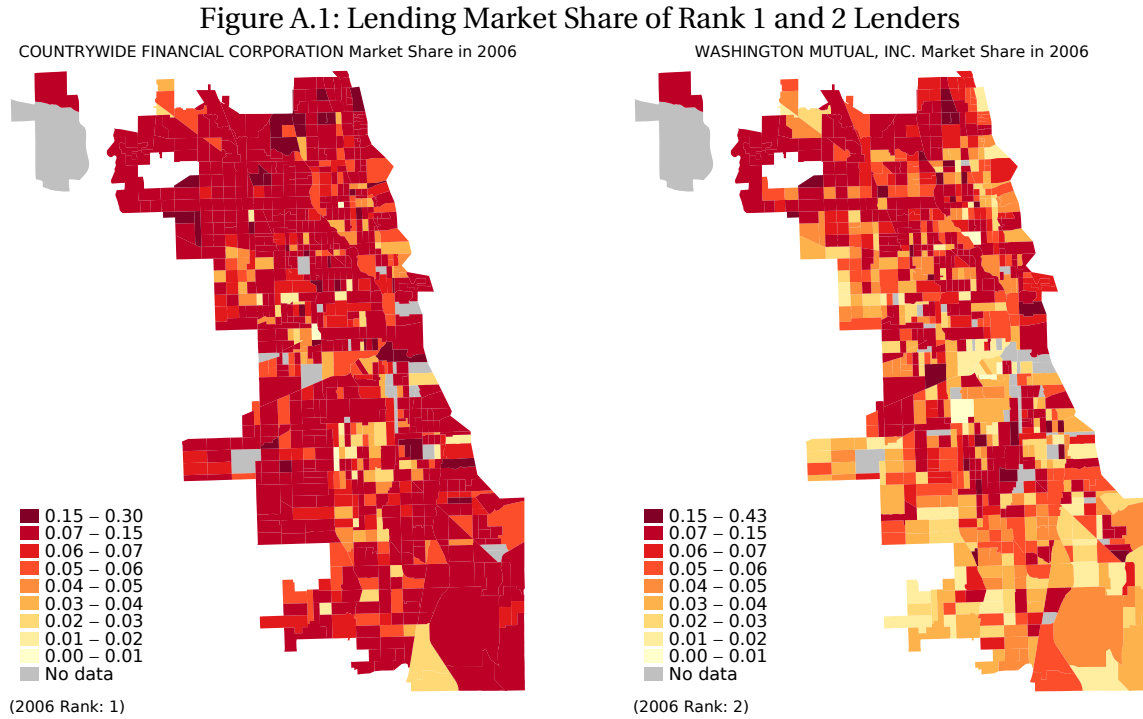


Figure A.2: Lending Market Share of Rank 3 and 4 Lenders

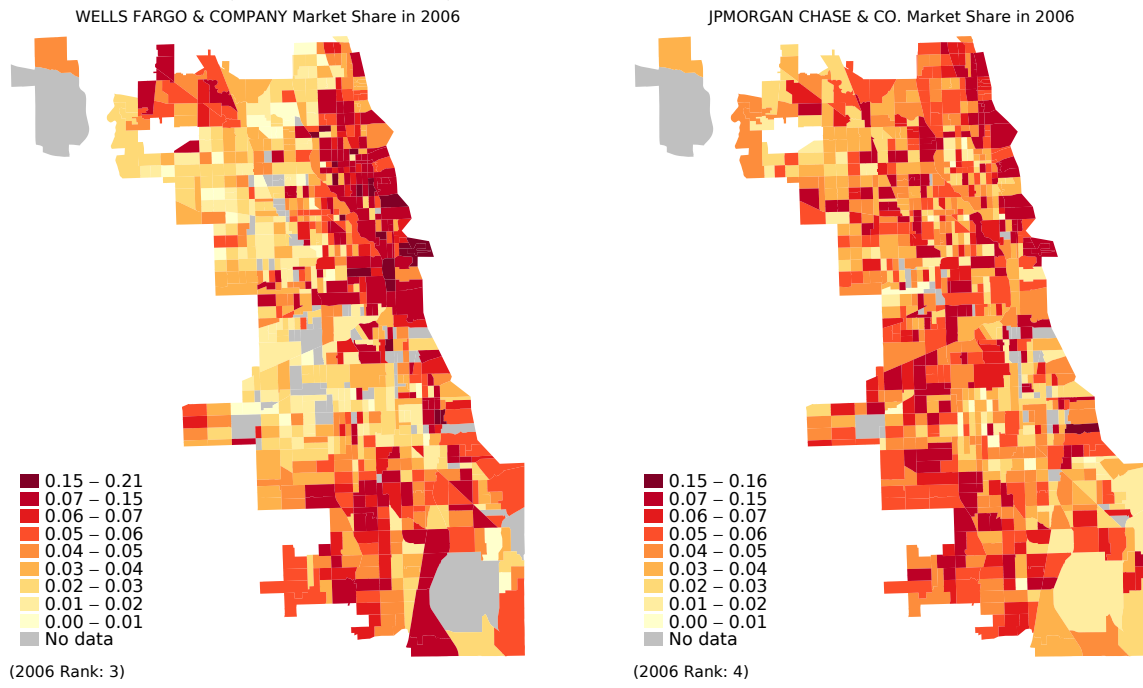


Figure A.3: Lending Market Share of Rank 5 and 6 Lenders

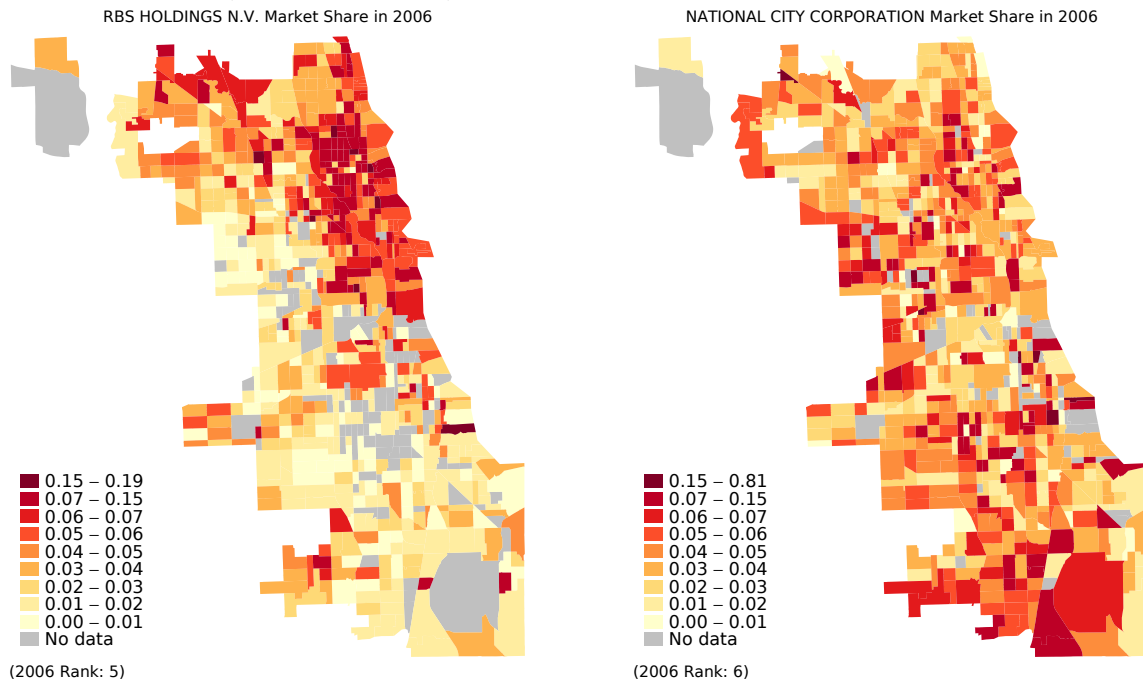


Figure A.4: Lending Market Share of Rank 7 and 8 Lenders

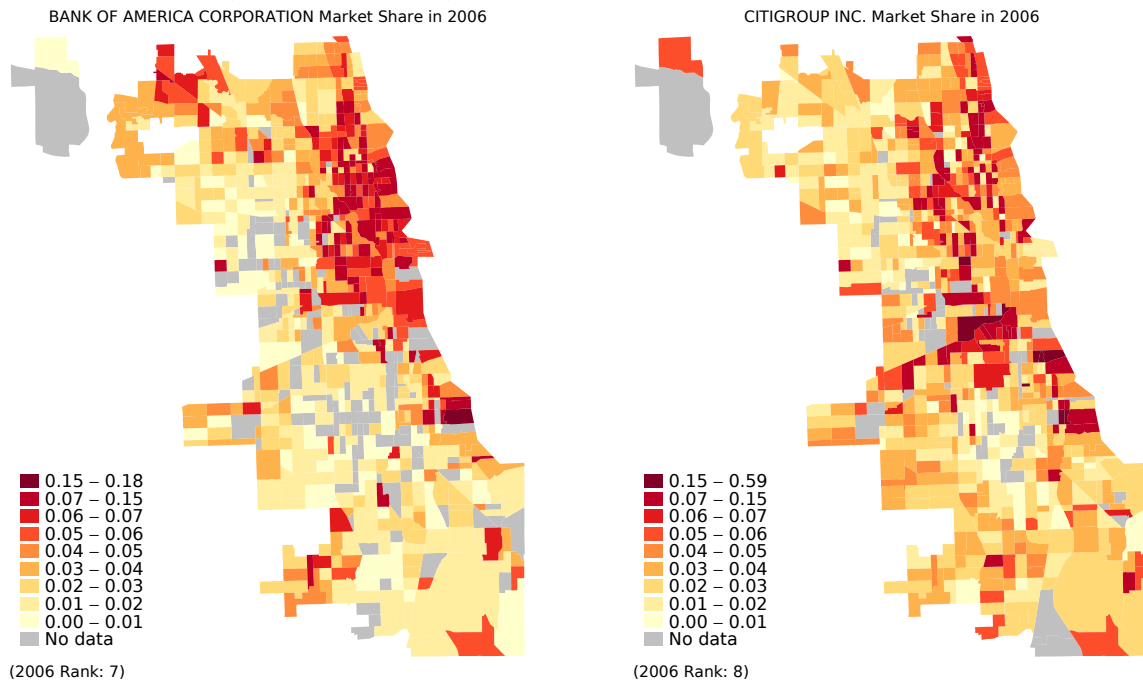


Figure A.5: Lending Market Share of Rank 9 and 10 Lenders

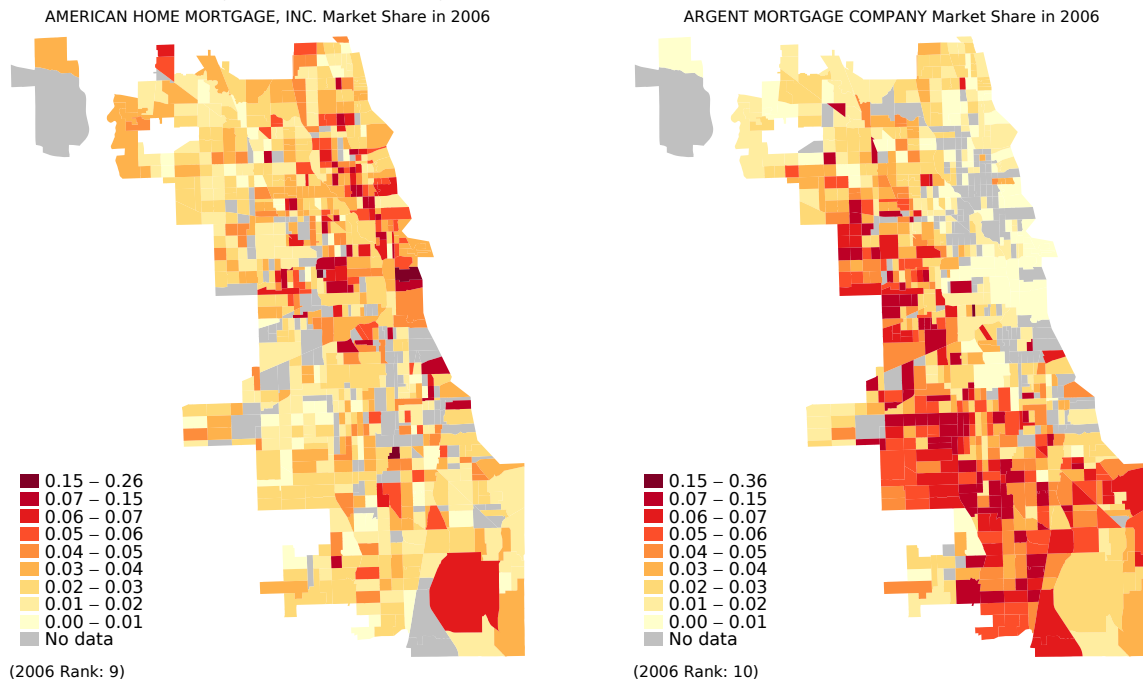
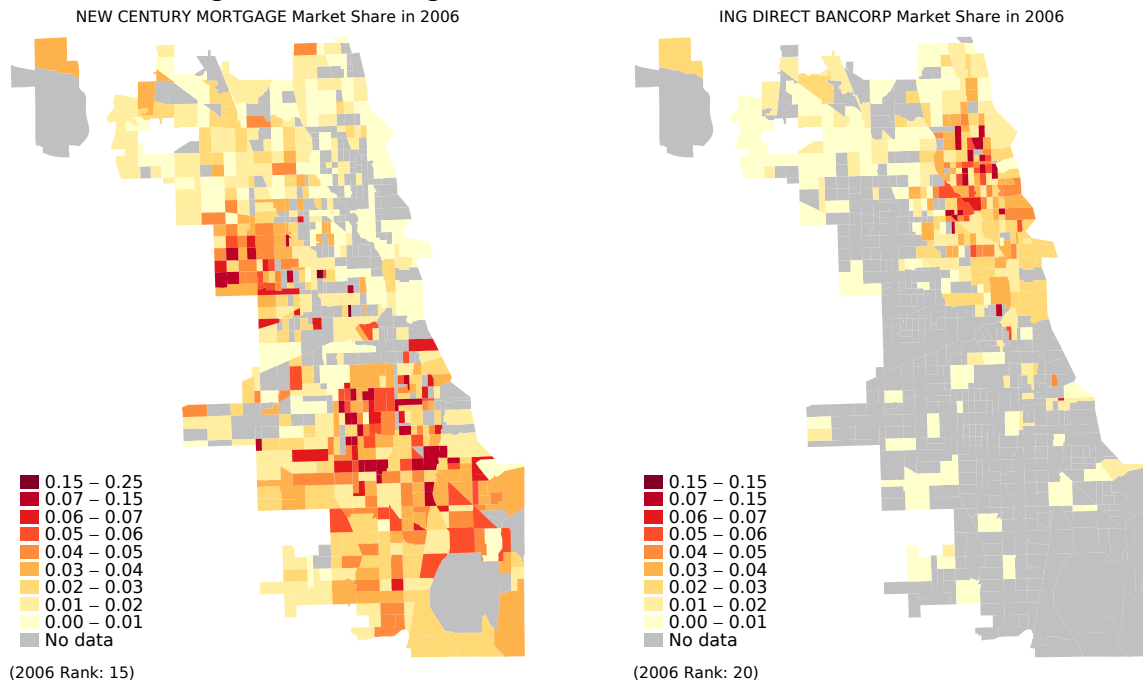


Figure A.6: Lending Market Share of Rank 15 and 20 Lenders



## A.4. OTHER DATA SOURCES

### A.4.1. BUILDFAQ

County-month level permit data from 2000-2016 are obtained from BuildFax. Permits are broken down into total, electrical, renovation, mechanical, new construction, plumbing, pool, roof, and solar, with BuildFax using internal algorithms including natural language processing of permit descriptions to homogenize classifications across permit jurisdictions as well as possible. Permits are also broken down by residential versus non-residential using similar methods. For each permit and building type, data is available on counts, counts with estimated costs, projects, and estimated costs.

Counties do not match exactly with permitting jurisdictions. As a result, for some counties, parts of the county can enter, and occasionally leave, coverage at different times. I assume that entering and exiting the dataset are roughly random.<sup>64</sup> Then, using coverage and population information at the jurisdiction-county-month level, I define the coverage percentage of a given county month as:

$$\text{Coverage}_{im} = \frac{\sum_j \text{Population}_{ijm} * I\{\text{InCoverage}\}_{ijm}}{\sum_j \text{Population}_{ijm}} \in [0, 1]$$

Given the assumption that coverage is random, I scale all data values by  $1/\text{Coverage}_{im}$  to estimate the total data value for the county, if coverage were 100%. Estimated cost data is normalized to be per household, while all other data are normalized to be per 1000 households.

### A.4.2. ZILLOW ZTRAX

The Zillow Transactions and Assessor Dataset (ZTRAX) is a national housing database including information from over 374 million public home transactions records across over 2,750 counties,

---

<sup>64</sup>In practice, data are more likely to be missing for more rural areas, so counties that comprise both urban and rural land are likely biased towards coverage of urban areas, which may have higher permitting rates. For robustness in the analysis, I subset to counties for which coverage is high (above 50% or 70%) to minimize the extrapolation effect of this random missing assumption.

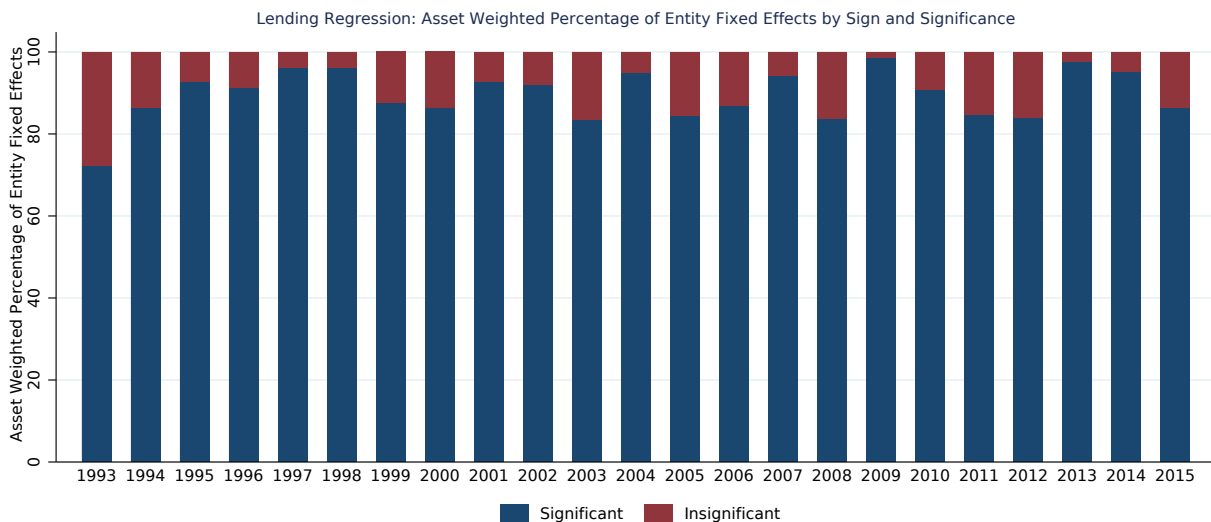
as well as matching assessment valuation panel data for approximately 200 million parcels in 3,100 counties. For this paper, I use (i) assessor data from Cook County, IL to classify addresses as residential or non-residential, and (ii) repeat home transactions matched to permits to calibrate the estimated market value of investment and the convexity of the relationship between investment and returns.

## B. ENTITY-YEAR FE DETAILS

In this appendix, I explore the variation in the relative importance of entity FEs across years in the county-year-entity level specifications, and the significance of entity FEs.

First, figure B.1 shows that entity FEs are statistically significantly different from zero consistently in all years. To construct this figure, I run annual regressions of entity and county FEs on the log change in lending, as described in equation (3.1), clustered at the county level. I then test whether entity FEs are significantly different from zero, and calculate the fraction of total lending within each year is done by lenders with significantly nonzero effects.

Figure B.1: Within-Year Significance of entity-year FEs for Top Entities

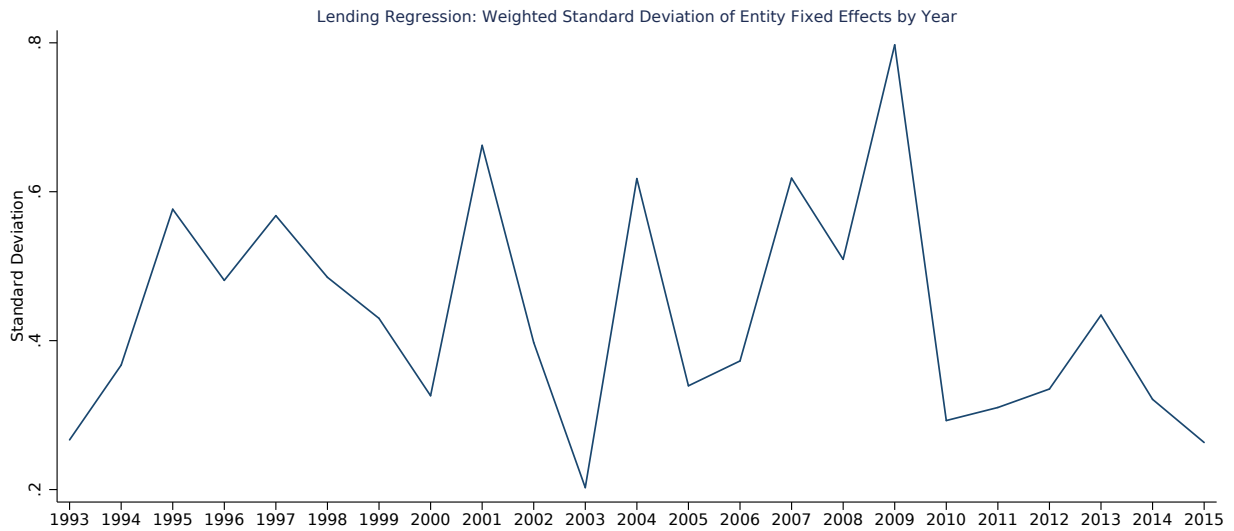


Note: Top entities are those that are in the top 50 in lending for any two years. Top entities make up over 60% of nationwide lending. Regressions to determine significance are clustered at the County level.

Figure B.2 shows that the within-year SD of supply FEs is highest during the crisis, suggesting that entities differed the most in their supply shocks during this time.

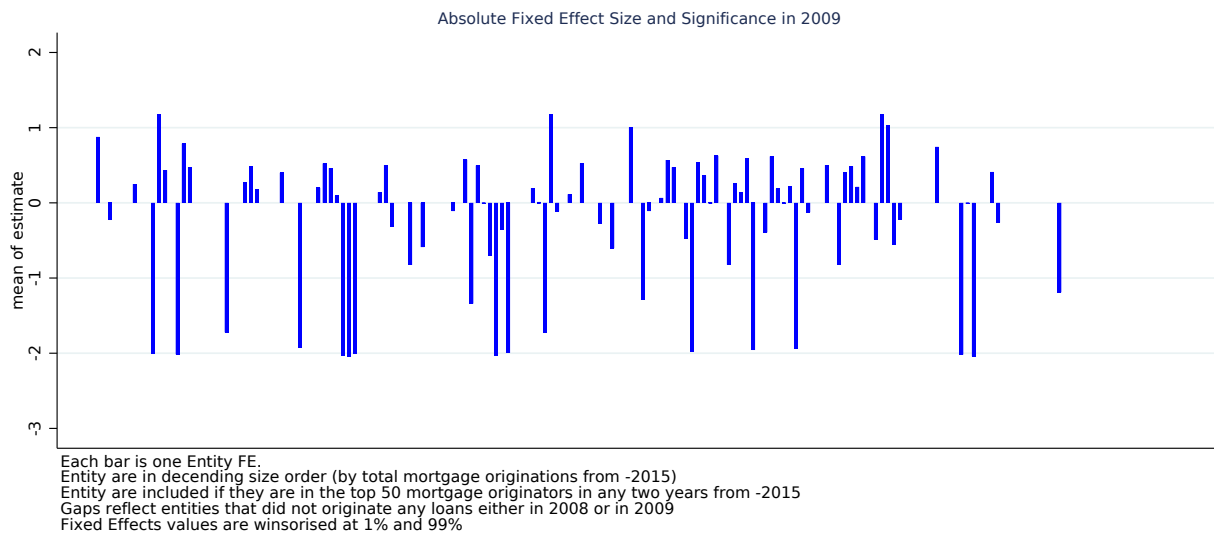
Finally, figure B.3 shows the distribution of Entity FEs for the largest entities in 2009, the year with the highest SD of entity FEs. I see that there is substantial heterogeneity, both in the sign and magnitude of credit supply shocks.

Figure B.2: Within-Year SD of entity-year FEs for Top Entities



Note: Top entities are those that are in the top 50 in lending for any two years. Top entities make up over 60% of nationwide lending.

Figure B.3: Entity FEs in 2009 for Top Entities



Note: Top entities are those that are in the top 50 in lending for any two years. Top entities make up over 60% of nationwide lending.

## C. ADDITIONAL SPECIFICATIONS

Table C.1 presents the effect of lagged credit shocks on counts of building code violations - a measure of disinvestment that is realized in the subsequent year. Though code violations are primarily the result of a longer-run process than these annual shocks capture, there still appear to be negative effects of lagged refinancing credit supply on the number of building code violations in the first column, measured as the number of violations in the tract. However, when lagged credit shortfall (standards) are also included in the second column, the significance of credit supply disappears. This is intuitive - homeowners who are most likely to fail to maintain their homes in a crisis are those who are marginally eligible for financial products in the pre-crisis period, and who are therefore likely to feel the effects of credit standards changes most acutely. In other words, unlike richer borrowers their demand for credit is low in dollar terms, but risky, so changing standards have a larger impact than changing supply. There are no significant effects on the number of houses violating code, suggesting that these effects are on the intensive margin - tightening credit standards increase the intensity of violations for households that are already struggling moreso than they draw new households to start violating code.

Table C.1: Lagged Credit Supply Shocks on Building Code Violations

	Violations	Violations	HViolating	HViolating
Lagged Refinancing Credit Supply Shock	-13.027** (5.502)	-7.591 (5.304)	0.256 (0.641)	0.173 (0.639)
Lag Dollar Credit Shortfall per HH x Frac Continuing Lenders		0.336*** (0.062)		-0.005 (0.009)
Lag Perc Continuing Lenders since Pre-Crisis		0.414*** (0.146)		-0.008 (0.016)
Year FE	Yes	Yes	Yes	Yes
Tract FE	Yes	Yes	Yes	Yes
Observations	6493	6493	6493	6493

Standard errors in parentheses

Standard errors are two-way clustered at the Census Tract and Neighborhood-Year levels

Variables are winsorised at the 1 and 99th percentile to handle extreme outliers

Observations are weighted by 2010 Census housing units

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Table C.2 presents the same specification using counts of 311 calls reporting abandoned houses

as an outcome. Abandoned housing data is only available starting in 2010, but some marginally significant negative effects are still found. Together, table C.1 and table C.2 suggest that credit availability may play a role in helping struggling landlords and homeowners stave off noticeable housing quality issues.

Table C.2: Lagged Credit Supply Shocks on Abandoned Housing

Lagged Refinancing Credit Supply Shock	-1.535*	
	(0.802)	
Lagged Home Purchase Credit Supply Shock		-1.362**
		(0.677)
Year FE	Yes	Yes
Tract FE	Yes	Yes
Observations	3639	3259

Standard errors in parentheses

Standard errors are two-way clustered at the Census Tract and Neighborhood-Year levels

Variables are winsorised at the 1 and 99th percentile to handle extreme outliers

Observations are weighted by 2010 Census housing units

Standards not included due to time-frame during which abandoned housing data is available (2010 onwards)

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

## C.1. REDUCED FORM SPECIFICATIONS

Table C.3 and table C.4 present the contemporaneous and lagged reduced form specifications associated with the IV regressions presented in table 4.7 and table 4.8.

Table C.3: Reduced Form Spillovers

	HInvest	Renov	Easy
Refinancing Credit Supply Shock	22.044** (10.205)	26.365*** (9.834)	-2.644 (1.874)
Dollar Credit Shortfall per HH X Frac Continuing Lenders	-0.202 (0.195)	-0.245 (0.187)	0.038 (0.029)
Perc Continuing Lenders since Pre-Crisis	-0.053 (0.285)	-0.176 (0.275)	0.116** (0.058)
Slag Refinancing Credit Supply Shock	50.099 (40.015)	59.018 (38.951)	-12.534* (6.849)
Slag Dollar Credit Shortfall per HH X Frac Continuing Lenders	-1.294*** (0.349)	-1.497*** (0.336)	0.154*** (0.047)
Slag Perc Continuing Lenders since Pre-Crisis	0.451 (0.642)	0.253 (0.612)	0.130 (0.143)
Year FE	Yes	Yes	Yes
Tract FE	Yes	Yes	Yes
Observations	7161	7161	7161

Standard errors in parentheses

Credit shortfall measure uses linear specification, subset to loans with unspecified rejection reason

Standard errors are two-way clustered at the Census Tract and Neighborhood-Year levels

Variables are winsorised at the 1 and 99th percentile to handle extreme outliers

Observations are weighted by 2010 Census housing units

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Table C.4: Lag Reduced Form Spillovers

	HInvest	Renov	Easy
Lag Refinancing Credit Supply Shock	4.183 (11.231)	1.726 (10.543)	1.599 (2.131)
Lag Dollar Credit Shortfall per HH X Lag Frac Continuing Lenders	-0.348** (0.169)	-0.372** (0.166)	0.028 (0.027)
Lag Perc Continuing Lenders since Pre-Crisis	0.337 (0.342)	0.260 (0.329)	0.039 (0.068)
Lag Slag Refinancing Credit Supply Shock	-160.690*** (32.645)	-174.234*** (31.363)	7.322 (5.230)
Lag Slag Dollar Credit Shortfall per HH X Lag Frac Continuing Lenders	-0.797*** (0.303)	-1.009*** (0.302)	0.178*** (0.050)
Lag Slag Perc Continuing Lenders since Pre-Crisis	-0.444 (0.867)	-0.467 (0.821)	-0.077 (0.186)
Year FE	Yes	Yes	Yes
Tract FE	Yes	Yes	Yes
Observations	7138	7138	7138

Standard errors in parentheses

Credit shortfall measure uses linear specification, subset to loans with unspecified rejection reason

Standard errors are two-way clustered at the Census Tract and Neighborhood-Year levels

Variables are winsorised at the 1 and 99th percentile to handle extreme outliers

Observations are weighted by 2010 Census housing units

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

## C.2. INDIVIDUAL INSTRUMENTS

Table C.5 and table C.6 present contemporaneous and lagged versions of the IV regressions presented in table 4.7 and table 4.8 using each instrument separately.

Table C.5: IV: Instrumented Spillovers with Single Instrument

	HInvest	Renov	Easy	HInvest	Renov	Easy
Slag Housing Investment Per HH	0.982*** (0.257)			0.884*** (0.175)		
Slag Renovation Investment Per HH		0.960*** (0.208)			0.879*** (0.141)	
Slag Easy Permit Investment Per HH			0.920*** (0.202)			0.861*** (0.147)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Tract FE	Yes	Yes	Yes	Yes	Yes	Yes
Instrument	Supply	Supply	Supply	Standards	Standards	Standards
Observations	7161	7161	7161	7955	7955	7955

Standard errors in parentheses

Credit shortfall measure uses linear specification, subset to loans with unspecified rejection reason

Standard errors are two-way clustered at the Census Tract and Neighborhood-Year levels

Variables are winsorised at the 1 and 99th percentile to handle extreme outliers

Observations are weighted by 2010 Census housing units

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Table C.6: IV: Lag Instrumented Spillovers with Single Instrument

	HInvest	Renov	Easy	HInvest	Renov	Easy
Lag Slag Housing Investment Per HH	-0.021 (0.262)			0.335* (0.176)		
Lag Slag Renovation Investment Per HH		0.063 (0.215)			0.409*** (0.149)	
Lag Slag Easy Permit Investment Per HH			0.674*** (0.256)			0.535*** (0.154)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Tract FE	Yes	Yes	Yes	Yes	Yes	Yes
Instrument	Supply	Supply	Supply	Standards	Standards	Standards
Observations	6422	6422	6422	7238	7238	7238

Standard errors in parentheses

Credit shortfall measure uses linear specification, subset to loans with unspecified rejection reason

Standard errors are two-way clustered at the Census Tract and Neighborhood-Year levels

Variables are winsorised at the 1 and 99th percentile to handle extreme outliers

Observations are weighted by 2010 Census housing units

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

## D. MATH APPENDIX

### D.1. WORKERS' HOUSING DEMAND

The initial problem is:

$$\max_{c, Q, d} u(Q, c) \quad \text{s.t.} \quad A_l s - \tau d = p_c c + P(Q, d)$$

Taking first order conditions:

$$[c] \quad u_2(Q, c) = \lambda$$

$$[Q] \quad u_1(Q, c) = \lambda P_1(Q, d)$$

$$[d] \quad P_2(Q, d) = -\tau$$

From the ratio of the first two:

$$P_1(Q, d) = \frac{u_1(Q, c)}{u_2(Q, c)}$$

The third describes a linear decline in prices as distance from the city center increases, and is a simpler variant of the Alonso-Muth condition standard in the monocentric city model literature (Duranton and Puga (2015)).

The first order conditions and constraints can be summarized as

$P_1(Q, d) = \frac{u_1(Q, c)}{u_2(Q, c)}$	FOCs
$P_2(Q, d) = -\tau$	FOC/Alonso-Muth
$A_l s - \tau d = c + P(Q, d)$	Budget Constraint

## D.2. BUILDING MANAGER'S PROBLEM DETAILS

The initial Bellman equation is (compressing  $E[\bar{q}_{t+1}|\vec{q}_t]$  to  $\bar{q}_{t+1}$  for space)

$$V(q, \bar{q}, d, \zeta, \bar{\zeta}) = (1 - \omega)P(q + \zeta, \bar{q} + \bar{\zeta}, d) \\ + \max_{q_{t+1} \geq (1-\delta)q} \left\{ -c(q_{t+1} - (1 - \delta)q) + \beta V(q_{t+1}, \bar{q}_{t+1}, d, \zeta, \bar{\zeta}) \right\}$$

The FOC of the maintenance problem is:

$$c'(q_{t+1}(q, \bar{q}_{t+1}, d, \zeta, \bar{\zeta}) - (1 - \delta)q) = \beta V_q(q_{t+1}(q, \bar{q}_{t+1}, d, \zeta, \bar{\zeta}), \bar{q}_{t+1}, d, \zeta, \bar{\zeta}) \quad (\text{D.1})$$

Plugging this optimal policy into the Bellman equation:

$$V(q, \bar{q}, d, \zeta, \bar{\zeta}) = (1 - \omega)P(q + \zeta, \bar{q} + \bar{\zeta}, d) - c(q_{t+1}(q, \bar{q}_{t+1}, d, \zeta, \bar{\zeta}) - (1 - \delta)q) \\ + \beta V(q_{t+1}(q, \bar{q}_{t+1}, d, \zeta, \bar{\zeta}), \bar{q}_{t+1}, d, \zeta, \bar{\zeta})$$

The envelope conditions:

$$V_q(q, \bar{q}, d, \zeta, \bar{\zeta}) = (1 - \omega)P_q(q + \zeta, \bar{q} + \bar{\zeta}, d) + (1 - \delta)c'(q_{t+1}(q, \bar{q}_{t+1}, d, \zeta, \bar{\zeta}) - (1 - \delta)q) \quad (\text{D.2})$$

$$V_{\bar{q}}(q, \bar{q}, d, \zeta, \bar{\zeta}) = (1 - \omega)P_{\bar{q}}(q + \zeta, \bar{q} + \bar{\zeta}, d) \quad (\text{D.3})$$

$$V_d(q, \bar{q}, d, \zeta, \bar{\zeta}) = (1 - \omega)P_d(q + \zeta, \bar{q} + \bar{\zeta}, d) + \beta V_d(q_{t+1}(q, \bar{q}_{t+1}, d, \zeta, \bar{\zeta}), \bar{q}_{t+1}, d, \zeta, \bar{\zeta}) \quad (\text{D.4})$$

Plugging the envelope condition for  $V_q$  into the FOC and decompressing the notation for  $\bar{q}_{t+1}$  yields the Euler equation:

$$c'(q_{t+1} - (1 - \delta)q_t) = \beta((1 - \omega)P_q(q_{t+1} + \zeta, E[\bar{q}_{t+1}|\vec{q}_t]) \quad (\text{D.5})$$

$$+ \bar{\zeta}, d) + (1 - \delta)c'(q_{t+2} - (1 - \delta)q_{t+1})) \quad (\text{D.6})$$

### D.3. PROOF: SYMMETRIC STEADY STATE IS UNIQUE IN LONG-RUN WITHOUT CONSTRAINTS

The symmetric steady state is the unique long-run equilibrium when  $\vec{\zeta} = 0$ , in the model with no constraints. First, note that equation (5.21) holds with certainty at  $q^{SS} = \bar{q} = q^{SSS}$ .

$$(q^{SSS})^{\rho-1} = Z \left( \alpha_q + \alpha_{\bar{q}} \left( \frac{q^{SSS}}{q^{SSS}} \right)^{\frac{\sigma-1}{\sigma}} \right)^{\frac{1}{\sigma-1}} \quad (\text{D.7})$$

Consider  $\bar{q} = \gamma q^{SSS}$  where  $\gamma < 1$ . Suppose, looking for a contradiction, that  $q^{SS} \leq \bar{q}$ . Then

$$\begin{aligned} Z \left( \alpha_q + \alpha_{\bar{q}} \left( \frac{\bar{q}}{q^{SS}} \right)^{\frac{\sigma-1}{\sigma}} \right)^{\frac{1}{\sigma-1}} &\geq Z \left( \alpha_q + \alpha_{\bar{q}} \left( \frac{\bar{q}}{\bar{q}} \right)^{\frac{\sigma-1}{\sigma}} \right)^{\frac{1}{\sigma-1}} \\ &= Z \left( \alpha_q + \alpha_{\bar{q}} \left( \frac{q^{SSS}}{q^{SSS}} \right)^{\frac{\sigma-1}{\sigma}} \right)^{\frac{1}{\sigma-1}} \\ (q^{SS})^{\rho-1} &\leq (\bar{q})^{\rho-1} < (q^{SSS})^{\rho-1} \quad \forall \rho > 1 \end{aligned}$$

This contradicts equation (D.7). Therefore  $q^{SS} > \bar{q}$  whenever  $\bar{q} < q^{SSS}$ , and  $q^{SS} < \bar{q}$  whenever  $\bar{q} > q^{SSS}$  when  $\vec{\zeta} = 0$ . This implies that the symmetric steady state is the unique stable long-run equilibrium when  $\vec{\zeta} = 0$ .

### D.4. COBB-DOUGLAS CASE

Rather than assuming CES, it's sometimes easier to simply assume Cobb-Douglas utility, as this leads to more explicit functional forms.

Specifically, this entails redefining  $Q$  from equation (5.12) as:

$$Q_{CD}(q + \zeta, \bar{q} + \bar{\zeta}) = (q + \zeta)^{\alpha_q} (\bar{q} + \bar{\zeta})^{\alpha_{\bar{q}}}$$

Where

$$\alpha_q + \alpha_{\bar{q}} = 1$$

Recall equation (5.9), the equation for the partial steady state:

$$c'(\delta q^{SS}) = \left( \frac{(1-\omega)P_q(q^{SS} + \zeta, \bar{q}_{t+1} + \bar{\zeta}, d)}{r + \delta} \right)$$

Plugging in our functional form assumptions yields:

$$(q_{CD}^{SS})^{\rho-1} = Z \left( \frac{Q_{CD}}{q_{CD}^{SS} + \zeta} \right)$$

Where  $Z$  is defined as in equation (5.22). When  $\bar{\zeta} = 0$ , this simplifies to:

$$\begin{aligned} (q_{CD}^{SS})^{\rho-1} &= Z \left( \frac{Q_{CD}}{q_{CD}^{SS}} \right) \\ &= Z (q_{CD}^{SS})^{\alpha_q - 1} \bar{q}_{t+1}^{\alpha_{\bar{q}}} \\ q_{CD}^{SS} &= Z^{\frac{1}{(\rho-1) + \alpha_{\bar{q}}}} \bar{q}_{t+1}^{\frac{\alpha_{\bar{q}}}{(\rho-1) + \alpha_{\bar{q}}}} \end{aligned}$$

#### D.4.1. COMPARATIVE STATICS

Writing equation (5.19) in terms of maintenance:

$$\begin{aligned} (m_t)^{\rho-1} &= \beta T \left( \frac{E[Q_{t+1} | \bar{q}_t]}{m_t + (1-\delta)q_t + \zeta} \right)^{\frac{1}{\sigma}} + (1-\delta)\beta(m_{t+1})^{\rho-1} \\ T &= \left( (1-\omega) \frac{A_l}{a_u} \right)^{\frac{1}{1-\alpha_c}} \left( \frac{\alpha_q(1-\alpha_c)}{A_m} \right) \alpha_c^{\frac{\alpha_c}{1-\alpha_c}} > 0 \\ Q_{t+1} &= \left( \alpha_q(q_{t+1} + \zeta)^{\frac{\sigma-1}{\sigma}} + \alpha_{\bar{q}}(\bar{q}_{t+1} + \bar{\zeta})^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} \end{aligned}$$

In this section, I consider the effect of various parameters on optimal maintenance, taking  $\bar{q}_{t+1}$  as given. Therefore, plugging in for  $Q_{t+1}$  and simplifying yields:

$$(m_t)^{\rho-1} = \beta T \left( \alpha_q + \alpha_{\bar{q}} \left( \frac{\bar{q}_{t+1} + \bar{\zeta}}{m_t + (1-\delta)q_t + \zeta} \right)^{\frac{\sigma-1}{\sigma}} \right)^{\frac{1}{\sigma-1}} + (1-\delta)\beta(m_{t+1})^{\rho-1} \quad (\text{D.8})$$

This version of the Euler equation makes some comparative statics clear. First,  $m$  is increasing in  $\bar{q}_{t+1}$  and  $\bar{\zeta}$  entirely through the price channel. Since neighborhood quality and housing quality are complementary for consumers, higher neighborhood quality incentivizes building managers to improve their housing quality to benefit from this complementary through higher rents.<sup>65</sup> Furthermore, optimal maintenance  $m_t$  is decreasing in  $q_t$  and  $\zeta$ , holding  $\bar{q}_{t+1} + \bar{\zeta}$  fixed, since this decreases the marginal returns to investment.

An increase in  $\beta$ , signifying a higher degree of patience, or a lower interest rate, increases  $m_{t+1}$  because building managers are more willing to trade off maintenance costs today for increased revenue tomorrow. An increase in  $A_l$  increases  $q_{t+1}$  through the price channel and an inflation effect - higher city productivity translates to higher wages, which leads to higher prices and more incentives for building managers to invest in housing. Similarly, an increase in  $\alpha_u$ , increasing the value of the outside option, decreases maintenance effort, as the value of living in the city relative to elsewhere in the country falls. Increasing  $\alpha_c$  decreases  $T$ , and therefore decreases maintenance effort, as consumers are willing to spend less of their income on housing relative to other consumption goods.

A proportional decrease in  $(1-\delta)q_t + \zeta$  and  $\bar{q}_{t+1} + \bar{\zeta}$ , indicative of an tract and neighborhood of proportionally lower quality, yields a decrease in optimal maintenance  $m$ . This is because at lower quality levels, less maintenance is required to obtain the same price effect, since the price gradient is more responsive to changes in  $m$ .

Finally, one can consider changes in consumers' relative preference for housing quality ( $\alpha_q$ ) to

<sup>65</sup>In game theoretic terms, this implies that our model exhibits increasing differences, and therefore strategic complementarities.

neighborhood quality ( $\alpha_{\bar{q}}$ ), holding preference for consumption ( $\alpha_c$ ) fixed. First, there is a level effect that increasing  $\alpha_q$  increases  $T$ , which increases maintenance: more maintenance occurs when households have a higher preference for housing quality relative to neighborhood quality. This is, however, potentially offset by the ratio of neighborhood quality to location quality:

$$\frac{\bar{q}_{t+1} + \bar{\zeta}}{m_t + (1 - \delta)q_t + \zeta}$$

In particular, when the numerator (neighborhood quality) is large, increasing  $\alpha_q$  can decrease maintenance by decreasing the extent to which consumers are willing to pay for the neighborhood, and therefore decreasing building managers' incentives to match neighborhood quality. This is also governed by  $\sigma$  - if consumers' preference for neighborhood quality and housing quality are highly complementary ( $\sigma$  near 0), then incentives to match neighborhood quality are stronger.

## E. MODEL CALIBRATION DETAILS

### E.1. ACS DATA CORRECTION

Using the baseline ACS data (Manson et al. (2018)), I identify quick reversions in housing costs (in which costs change, and then revert sharply in the subsequent period) through the following metric:

$$QR_{it} = \frac{|PC_{it} - PC_{it+1}|}{\max(|PC_{it}|, |PC_{it+1}|)} * |PC_{it}| * |PC_{it+1}|$$

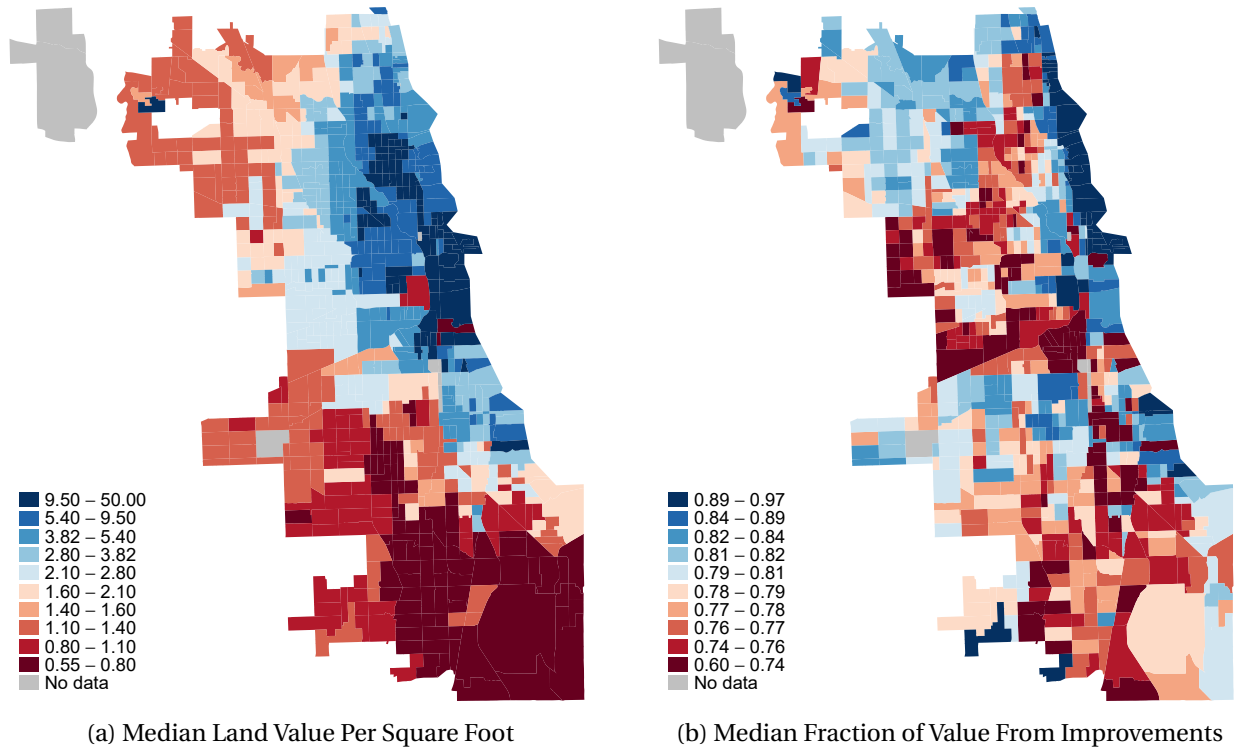
$PC_{it}$  is the percentage change in housing costs from time  $t - 1$  to  $t$  for location  $i$ :  $(HC_{it} - HC_{it-1})/HC_{it-1}$ . Using the baseline data, I identify a break in this measure at the 90th percentile -  $(i, t)$  pairs above this break have extremely large deviations that are reverted in the subsequent period, and are likely due to small sample size issues.  $PC_{it}$  is winsorised to this 95th percentile, and the underlying value of  $HC_{it}$  is updated.  $PC_{it}$  values for the whole dataset are then recalculated, and the process is repeated until no changes fall above the initial  $QR$  cutoff. This correction has a relatively minor effect on the overall series, with the two series having a correlation of over 0.999, but is extremely important when taking the data to the model. Extremely large deviations in housing costs are justified by the model through large changes in unobserved housing or neighborhood quality. When these deviations are spurious, this can lead to amplified error in the model.

## E.2. ZTRAX ON LAND AND ASSESSMENT VALUES

ZTRAX housing assessment data corroborates this confounding correlation between housing and non-housing amenities, insofar as non-housing amenities are reflected in the price of land. Note that the price of land is determined by a number of factors, including  $q$ ,  $\zeta$ ,  $\bar{q}$ , and  $\bar{\zeta}$ , as well as the cost of new construction, which is not estimated in this paper. As a result, this discussion is primarily descriptive, as land values do not map cleanly to my analysis.

Housing assessors separately estimate land value and improvement value, where land value is primarily estimated from the value of vacant land sales, and assessed value is primarily estimated as the difference between recent sales values of similar properties in the neighborhood net of the land value. The left panel of Figure E.1 shows the median land value per square foot in 2010, while the right panel shows the fraction of total assessed value attributed to improvements (built structures) rather than land. While there are differences between the two, there are also a large number of tracts in which improvements are a significant fraction of total value despite land values being high, and visa versa, indicative of the unweighted correlation between median land and improvement value of 0.673. The extent to which this correlation is attributed to local housing quality  $q$ , neighborhood quality  $\bar{q} + \bar{\zeta}$ , and local non-housing amenities  $\zeta$  is uncertain.

Figure E.1: Assessment Information: 2010 Tax Year



## F. MODEL ESTIMATION DETAILS

### F.1. ESTIMATION EQUATIONS

When obtaining implied quality from prices, given parameters, begin with the pricing equation:

$$\bar{P}(Q, d) = \left( \frac{A_l}{a_u} \right)^{\frac{1}{1-\alpha_c}} \alpha_c^{\frac{\alpha_c}{1-\alpha_c}} (1 - \alpha_c) Q - \tau d$$

Where:

$$Q = \left( \alpha_q (q + \zeta)^{\frac{\sigma-1}{\sigma}} + \alpha_{\bar{q}} (\bar{q} + \bar{\zeta})^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}$$

Rearranging, and imposing that  $P$  is data at each location  $i$ , yields the following moment at each location:

$$M_i = \frac{P_i + \tau d_i}{\left( \frac{A_l}{a_u} \right)^{\frac{1}{1-\alpha_c}} \alpha_c^{\frac{\alpha_c}{1-\alpha_c}} (1 - \alpha_c)} - Q_i = 0$$

The objective function is:

$$O_i = \sum_{i=1}^N w_i M_i^2$$

where  $w_i$  is the weight associated with location  $i$ , generally the number of households.

This is a non-trivial problem for the computer to solve, as the vector  $\vec{q}$  moves  $\vec{\bar{q}}$ , and these are conflated in  $Q$ . As such, it is helpful to provide the solver with an analytic gradient. This is constructed as:

$$\frac{\partial O_i}{\partial q_j} = 2 \sum_{i=1}^N w_i M_i \frac{\partial M_i}{\partial q_j}$$

Where

$$\begin{aligned}\frac{\partial M_i}{\partial q_j} &= -\frac{\partial Q_i}{\partial q_j} \\ &= -Q_i^{\frac{1}{\sigma}} \left( I\{i = j\} \alpha_q (q_i + \zeta_i)^{-\frac{1}{\sigma}} + I\{i \neq j\} \alpha_{\bar{q}} (\bar{q}_i + \bar{\zeta}_i)^{-\frac{1}{\sigma}} G_{ij} \right)\end{aligned}$$

Where  $G$  is the spatial weights matrix, and the fact that  $G$  has a zero diagonal is used to simplify the above expression.

In addition to providing the analytic gradient, estimation of  $q_{it} + \zeta_i$  over a grid of values of  $\alpha_q \in [0, 1]$  begins with from  $\alpha_{\bar{q}} = 0$ . At this point, neighborhood factors play no role, and  $(q_{it} + \zeta_i)$  can be solved for explicitly given  $P_{it}$  and  $d_i$ . This vector  $Q$  is then used as an input for the next grid-point  $\alpha_{\bar{q}} = 0.01$ . For subsequent grid-points, initial values are provided as a best guess based on the previous two iterations of  $Q$ . If  $Q^i$  and  $Q^{i+1}$  are the previous two quality values, the initial guess for the next step is defined as:

$$Q_{init}^{i+2} = Q^{i+1} + (Q^{i+1} - Q^i)$$

At times, it will be relevant to construct this derivative for the Cobb-Douglas case when  $\sigma = 1$  and  $Q_i$  is poorly defined. In this case, the objective is similar:

$$M_{i,CD} = \left( \frac{P_i + \tau d_i}{\left(\frac{A_i}{a_u}\right)^{\frac{1}{1-\alpha_c}} \alpha_c^{\frac{\alpha_c}{1-\alpha_c}} (1-\alpha_c)} \right)^{1-\alpha_c} - Q_i = 0$$

Where here

$$Q_i = (q_i + \zeta_i)^{\alpha_q} (\bar{q}_i + \bar{\zeta}_i)^{\alpha_{\bar{q}}}$$

The difference in the objective is because in this case  $\alpha_q + \alpha_{\bar{q}} + \alpha_c = 1$ , rather than  $\alpha_q + \alpha_{\bar{q}} = 1$  in the CES case. This is a simple rescaling.

For Cobb-Douglas, the derivative is:

$$\begin{aligned}
 \frac{\partial M_{i,CD}}{\partial q_j} &= -\frac{\partial Q_i}{\partial q_j} \\
 &= -\left( I\{i=j\} \alpha_q (q_i + \zeta_i)^{\alpha_q - 1} (\bar{q} + \bar{\zeta}_i)^{\alpha_{\bar{q}}} + I\{i \neq j\} \alpha_{\bar{q}} (q_i + \zeta_i)^{\alpha_q} (\bar{q} + \bar{\zeta}_i)^{\alpha_{\bar{q}} - 1} G_{ij} \right) \\
 &= -(q_i + \zeta_i)^{\alpha_q} (\bar{q}_i + \bar{\zeta}_i)^{\alpha_{\bar{q}}} \left( I\{i=j\} \frac{\alpha_q}{q_i + \zeta_i} + I\{i \neq j\} \frac{\alpha_{\bar{q}}}{\bar{q}_i + \bar{\zeta}_i} G_{ij} \right)
 \end{aligned}$$

## G. ADDITIONAL SIMULATIONS

In this section, I consider a number of additional simulations, adjusting  $\omega$ ,  $A_m$ ,  $\rho$ , and  $\vec{\zeta}$  relative to the baseline.

### G.1. PROPERTY TAX OR MORTGAGE RATE CHANGES

First, I consider adjusting  $\omega$  relative to the baseline. This simulates the effect of a decrease in property tax, mortgage rates, or any other flow-cost paid by landlords that does not affect the quality of the home. Specifically, I consider multipliers for  $\omega$  of 0.85, 0.9, and 0.95, similar to the multipliers used for  $A_m$  in the opportunity zone application, relative to the baseline calibrated value 0.36705. This results in values of  $\omega$  of 0.3119925, 0.330345, and 0.3486975.

Recalling the calculation used to calibrate the baseline value of  $\omega$ :

$$\omega = \frac{763.86 * (114,993/274,993) + (160000 * .02009)/12}{160000 * .01} = 0.36705$$

Of this, the following fraction of total flow costs are due to property taxes, while the rest is due to mortgage interest payments.

$$\frac{(160000 * .02009)/12}{763.86 * (114,993/274,993) + (160000 * .02009)/12} = 0.4561$$

Given this,  $\omega$  multipliers in these simulations can be interpreted either as minor changes to the tax rate or the mortgage interest rate, holding the other constant. The multiplier 0.9 specifically represents a decrease in the tax rate from 2.009% to 1.568%, per the following calculation:

$$100% * (0.36705 * 0.9 * (160000 * .01) - 763.86 * (114,993/274,993)) * (12/160000) = 1.568%$$

Calculating the mortgage interest rate change this implies is more involved. The 0.9 but requires

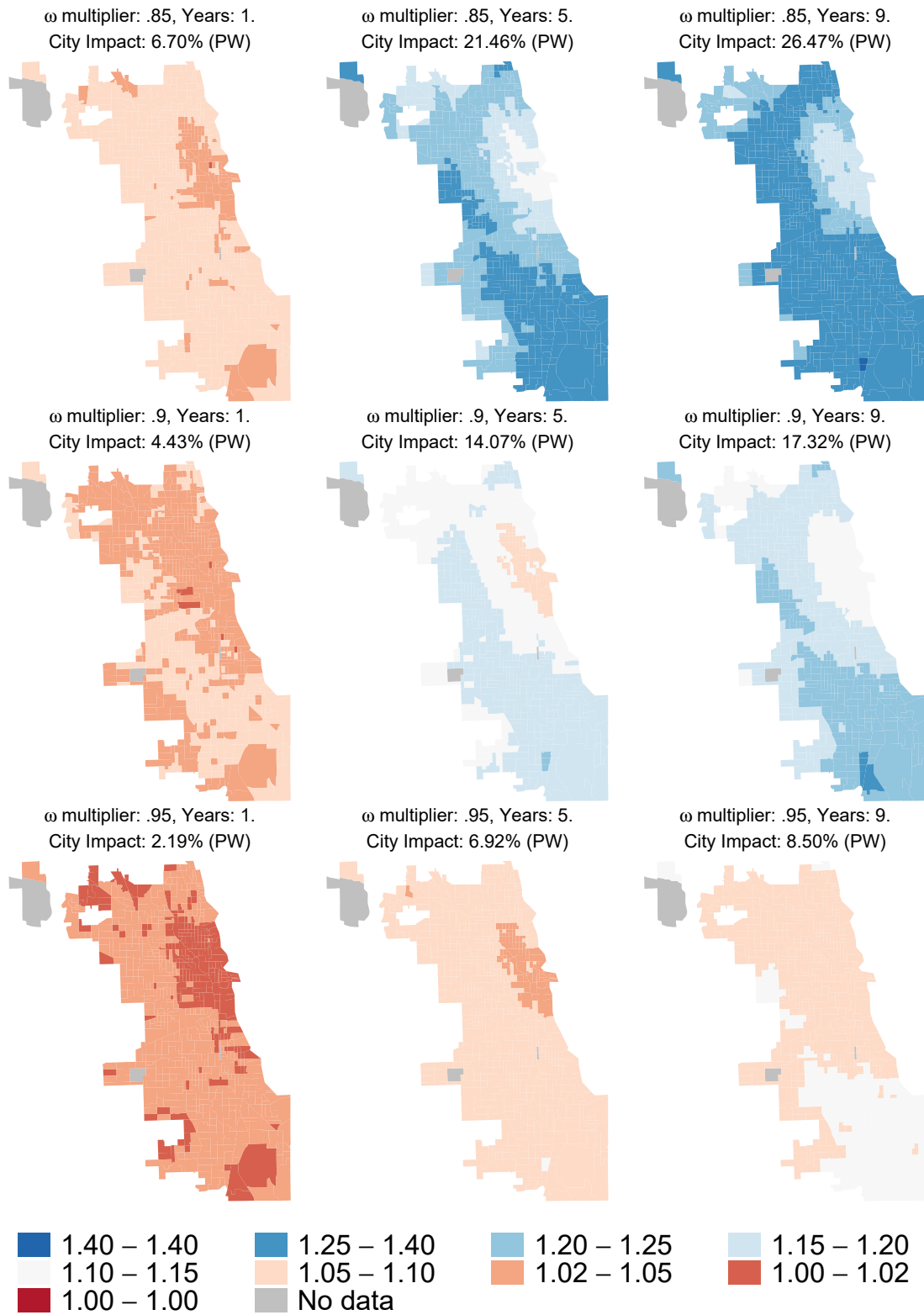
a monthly mortgage interest loss payment of:

$$(0.36705 * 0.9 * (160000 * .01) - (160000 * .02009)) / 12 = 260.69$$

Which implies a change in the mortgage rate from 4% to roughly 3.34%, leading to a monthly payment of \$704.26 dollars, of which \$260.69 is interest loss.

Figure G.1 shows an unsurprising lack of spatial heterogeneity in the long term impact of changes to  $\omega$ , due to the city-level nature of the change. The slight heterogeneity that exists is driven by driven by initial housing quality and how close tracts are to their short-run quality target  $q^{SS}$  and long-run equilibrium  $q^{SSS}$  as  $\omega$  changes. However, the city-level price increase values provide an interesting insight into the revenue-effects of these sorts of policies. Considering the second row which features an  $\omega$  multiplier of 0.9, the calculation above shows that if this were due solely to a change in the property tax rate, the change would be a decrease of  $(2.009 - 1.568) / 2.009 * 100\% = 22\%$  in the tax rate. The simulation suggests that even after 9 years, the price-increase would be only 17.32%, implying that city revenues would fall. Specifically, considering for example a \$160,000 home, tax revenues would shift from  $.02009 * \$160,000 = \$3214.4$  to  $.01568 * \$160,000 * 1.1732 = \$2943.3$ , a decline of 8.4%. In practice, tax revenues would likely decrease even more than this, due to significant frictions in adjusting assessed values to reflect increases in home values. If anything, the simulation suggests that Illinois' property tax may be too low if the city government's only goal is increasing revenue. However, tax policy depends on a number of other features external to the model, including competition across cities to attract businesses, skilled workers, and non-housing investors, as well as basic concerns such as minimizing bureaucratic overhead associated with handling high volumes of assessment-related complaints. Finally, even in the baseline model, figure 6.6 shows that a significant number of tracts are constrained. Increasing taxes further would tighten these constraints on a larger number of tracts, and could lead to steady neighborhood decline in areas where a high volume of nearby tracts are constrained simultaneously.

Figure G.1: Varying  $\omega$ : Price Impact



Note: PW is population-weighted price impact

## G.2. MENU OR INVESTMENT COST CHANGES

Next, I consider adjustments to  $A_m$  and  $\rho$ . For  $A_m$  adjustments, I follow the methodology used for the opportunity zone analysis, but with city-wide changes in this parameter rather than just in opportunity zones. This could represent many possible changes such as construction cost decreases or menu cost reductions due to increased ease of finding new renters after a value-enhancing investment. As in the opportunity zone analysis, I consider multipliers for  $A_m$  of 0.85, 0.9, and 0.95 relative to its baseline calibrated value.

By contrast to  $A_m$ , changes in  $\rho$  affect both the curvature and the slope of the investment cost curve. Recall the functional form for investment costs:

$$c(m) = \frac{A_m}{\rho} m^\rho \quad \rho > 1$$

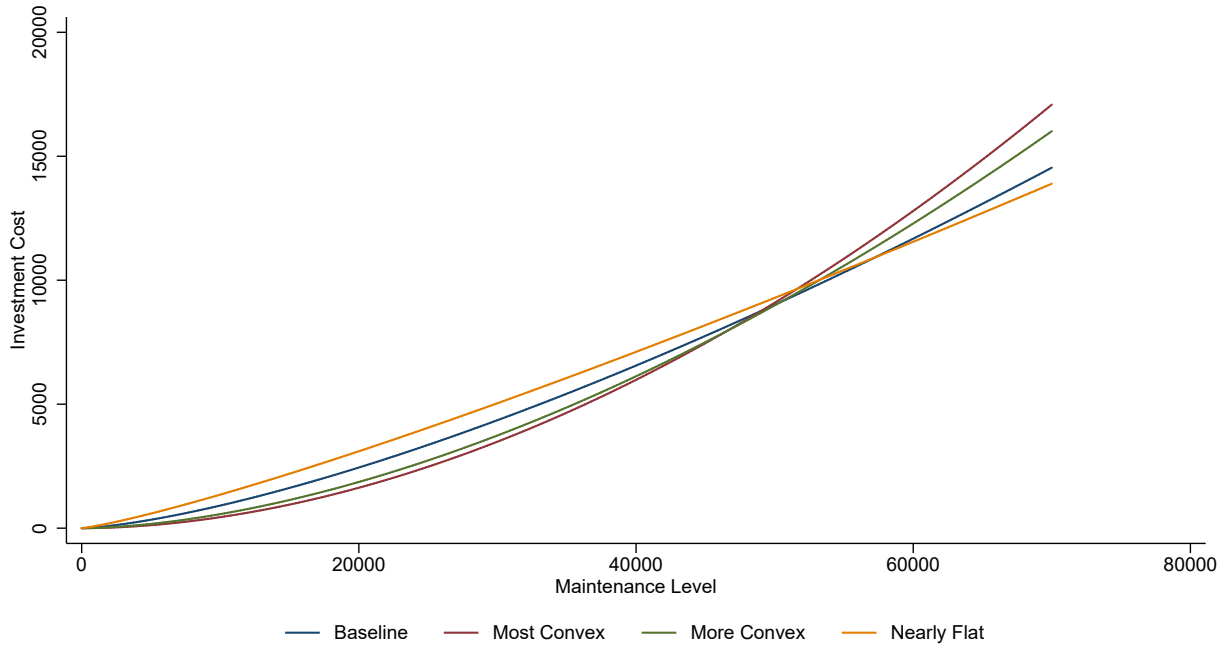
To analyze changes in  $\rho$ , I offset changes to  $\rho$  with corresponding adjustments to  $A_m$  such that  $q^{SS}$  remains unchanged. Specifically, I consider three cases:

Label	$A_m$ multiplier	$\rho$ multiplier	$\rho$ value
Baseline	1	1	1.422794
Most Convex	0.01	1.3176546	1.874751
More Convex	0.05	1.2066391	1.716799
Nearly Flat	10	0.8411727	1.196815

The first two are more convex than the baseline cost function, while the latter is nearly linear. Figure G.2 visualizes the impact that the multipliers in the table have on the investment cost curve.

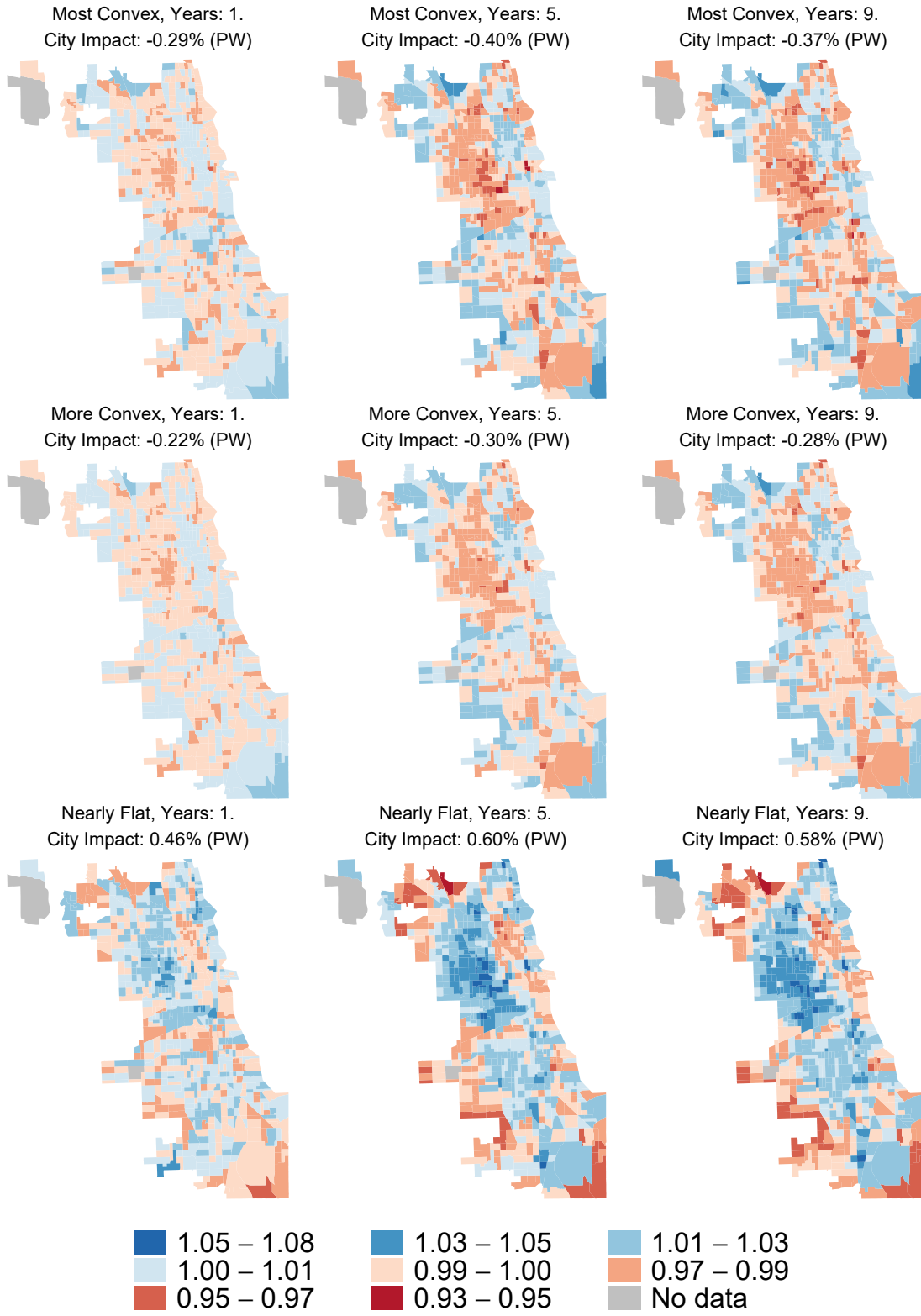
Despite the relatively similar investment cost curves in figure G.2, and the fact that all three simulations leave the long-run steady state unchanged, the short and medium-run effects of the three changes yield surprising, though minor, spatial heterogeneity as shown in figure G.3. In the first column, which depicts changes relative to the baseline simulation in the first year, the flattest cost curve has the largest initial positive effect on prices, implying that a higher level of

Figure G.2: Investment cost curve for variants of  $\rho$  and  $A_m$



investment occurred. This is somewhat counter-intuitive, since a flatter cost curve means that minor investments become relatively pricier, and therefore one might expect some households forgoing investment entirely. However, since  $q^{SSS}$  is the same for all of these simulations, the long-run target of all tracts is relatively consistent across simulations, barring credit constraints. Therefore, a flatter cost curve leads to dynamic incentives to invest in chunks if one is to invest at all, as the incentives to make minor adjustments in each period, so as to avoid the increasing marginal cost associated with large investments, are weaker. As a result, the flatter investment curve leads to faster convergence, and particularly large investments in neighborhoods that have a great deal of catching up to do, such as the south and west sides. In the first few years, these incentives can be dampened by constraints, but by year 5, as shown in the second column, many of these neighborhoods have started significant positive investment. Finally, the third column shows the slow convergence across the various scenarios towards the same long-run equilibrium due to sharing the same  $q^{SSS}$ , with all being marginally closer to the baseline after 9 years than after 5 years.

Figure G.3: Varying  $\rho$  and  $A_m$ : Price Impact



Note: PW is population-weighted price impact

By contrast to the previous exercise, changing  $A_m$  alone, without regard for the long-run symmetric steady state  $q^{SSS}$  or the exercise used to calibrate  $A_m$  initially, has dramatic effects on city-level prices. Because the model holds the outside option  $\bar{U}$  fixed, a decrease in  $A_m$  leads to a boom in low-cost quality investment. This cheap housing quality shifts consumers away from non-housing consumption, leading to a larger fraction of income spent on housing.<sup>66</sup> Though a model with new construction to increase the housing supply and more aggressive congestion forces could counteract this outcome, the general concept follows simply from supply and demand: because houses become higher quality for a lower price, competition among nationwide consumers who can only obtain utility  $\bar{U}$  necessarily leads to an increase in city-wide rents. There is some spatial heterogeneity in the impact of this change relative to the baseline that is driven by initial housing quality and how close tracts are to their short-run quality target  $q^{SS}$  and long-run equilibrium  $q^{SSS}$  as  $A_m$  changes.

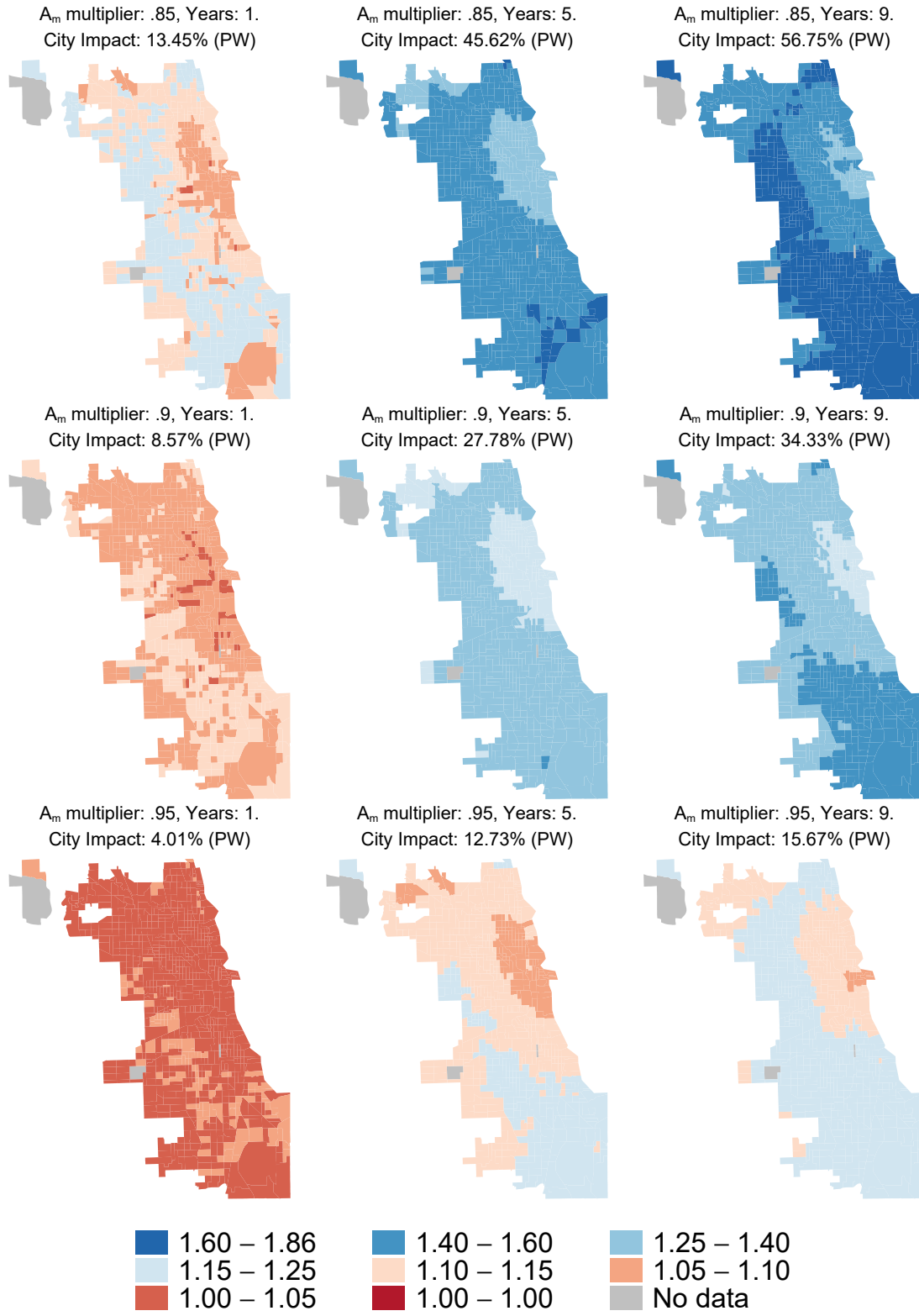
### G.3. BUILDING AMENITIES: THE OBAMA LIBRARY

In addition to tax changes, the model provides a way to consider the city-level impacts of infrastructure development in individual census tracts. Among the largest current proposals for development is the Barack Obama Presidential center, planned to be built in Jackson Park on the south-east side of Chicago. To simulate the effects of this policy, I consider a counterfactual in which  $\zeta_i$  for the tract containing the library increases by five weighted standard deviations, and the two tracts that hold the additional new golf course by two weighted standard deviations. Determining the size of changes in  $\zeta$  for counterfactuals is necessarily imprecise, as  $\vec{\zeta}$  is the combination of the common components a number of factors (crime, transit connectivity, school quality) and how they predict total quality  $q + \zeta$ , as described in equation (6.1). The large adjustment above is used primarily because it is unit-free, captures the spirit of the infrastructure project as an investment by the city that elevates a given neighborhood significantly

---

<sup>66</sup>Note that this contrasts with the general Cobb-Douglas result that constant fractions of income are spent on different goods - here the existence of market forces in bidding for houses and the fixed  $\bar{U}$  lead to changing spending distributions.

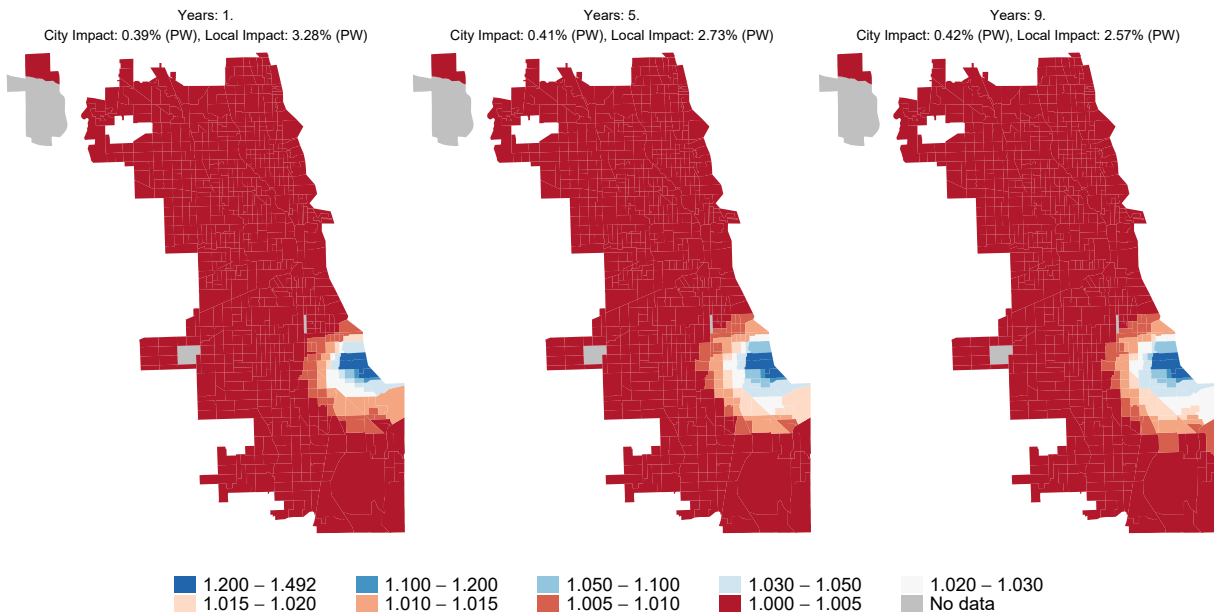
Figure G.4: Varying  $A_m$ : Price Impact



Note: PW is population-weighted price impact

relative to its peers, and reflects a reasonable upper bound on the policy’s impact in the model.

Figure G.5: Price impact of Obama Presidential Center after 5 years



Note: PW is population-weighted price impact

Figure G.5 captures the impact of this policy visually, as well as its price impact on the city as a whole and on the area nearby (tracts which experience more than a 0.1% price increase). There are a couple of main takeaways. First, prices go up in the very first year, before housing has a chance to react. This is because the model suggests that consumers are willing to pay more for houses with higher neighborhood quality, and on-impact this is capitalized into prices. Second, the effect is small - highly localized, inconsequential for the city as a whole, and minor relative to previous policy considerations such as tax changes or investment incentives. Finally, as the years pass there are two competing phenomena. Nearby tracts have incentives, and increased capacity, to invest, because their neighborhood’s improvement drives up the marginal returns of housing quality. By contrast, homes directly within the shocked tracts have less incentives to invest - their housing price is now buoyed by the presidential center - quality that requires no

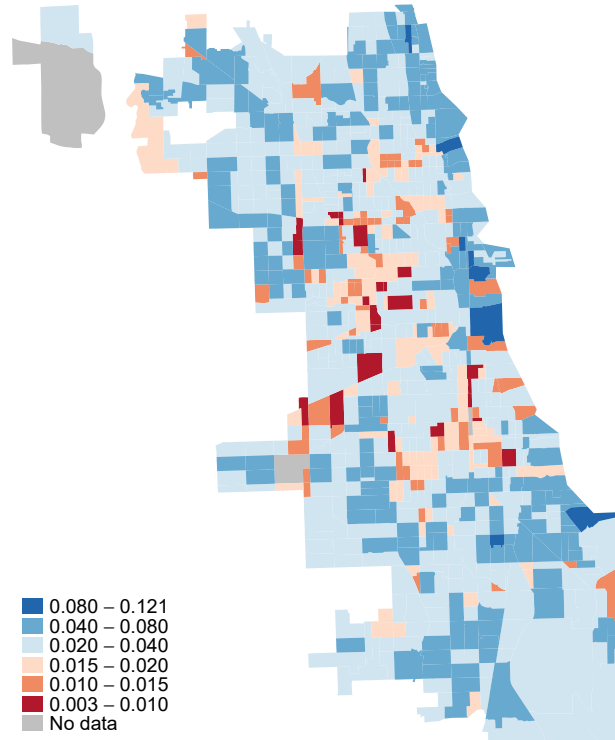
maintenance - and less driven by their own actions. In the long term, these tracts converge to having overall higher quality than before the investment, but with a larger proportion of it due to local amenities than to housing quality. Strengthening complementarities between the two types of quality in the utility function could counteract this effect.

These two effects lead to the odd pattern seen in city and local price effects - city prices continue to rise slightly over time, as minor ripple effects to further tracts have small positive effects on prices, while local prices decline relative to the baseline, as landlords shade down their maintenance spending and rely more heavily on the Obama library to sustain their desirability.

Using this same methodology, figure G.6 depicts which tracts in the city's would have the largest city-level impact if they experienced a 1 standard deviation increase in amenities. This provides a sort of counterfactual for the Obama library ("what if it were build elsewhere") as well as general guidance on what parts of the city benefit the most in the medium-run from amenity investment, which could guide policymakers. The first major takeaway is that, unsurprisingly, city-level price impacts of relatively small adjustments to amenities in individual tracts are somewhat small, capping out at around a 0.121% increase in city-level prices. While the highest impact tracts are generally concentrated in high-density areas, for whom the local population that can benefit from the amenities is large, there is a reasonable amount of spatial heterogeneity that is somewhat unpredictable, and uncorrelated with the city's demographic composition. As a result, the model could be informative for predicting what neighborhoods have the highest "multipliers" in terms of price (and therefore tax) increases engendered by amenity investment.

There are a number of caveats that are necessary for this analysis. First, improving amenities by 1SD may be heterogeneously difficult across tracts, and may be impossible for some tracts with already high levels of amenities on the one hand, or entrenched crime or school quality problems on the other. Second, the model ignores many other impacts of amenity investment, such as tourism, commercial response, and migration within the city, which can all provide

Figure G.6: City Price Impact of 1SD Shocks to Amenities after 5 years (% PW)"



additional multipliers or confounding effects. Therefore while the model can be used for this kind of exercise, it does not provide a holistic sense of the impacts of amenity investment - which are often heavily reliant on these responses “outside” of the model - too far beyond what a simple hedonic regression would also achieve.

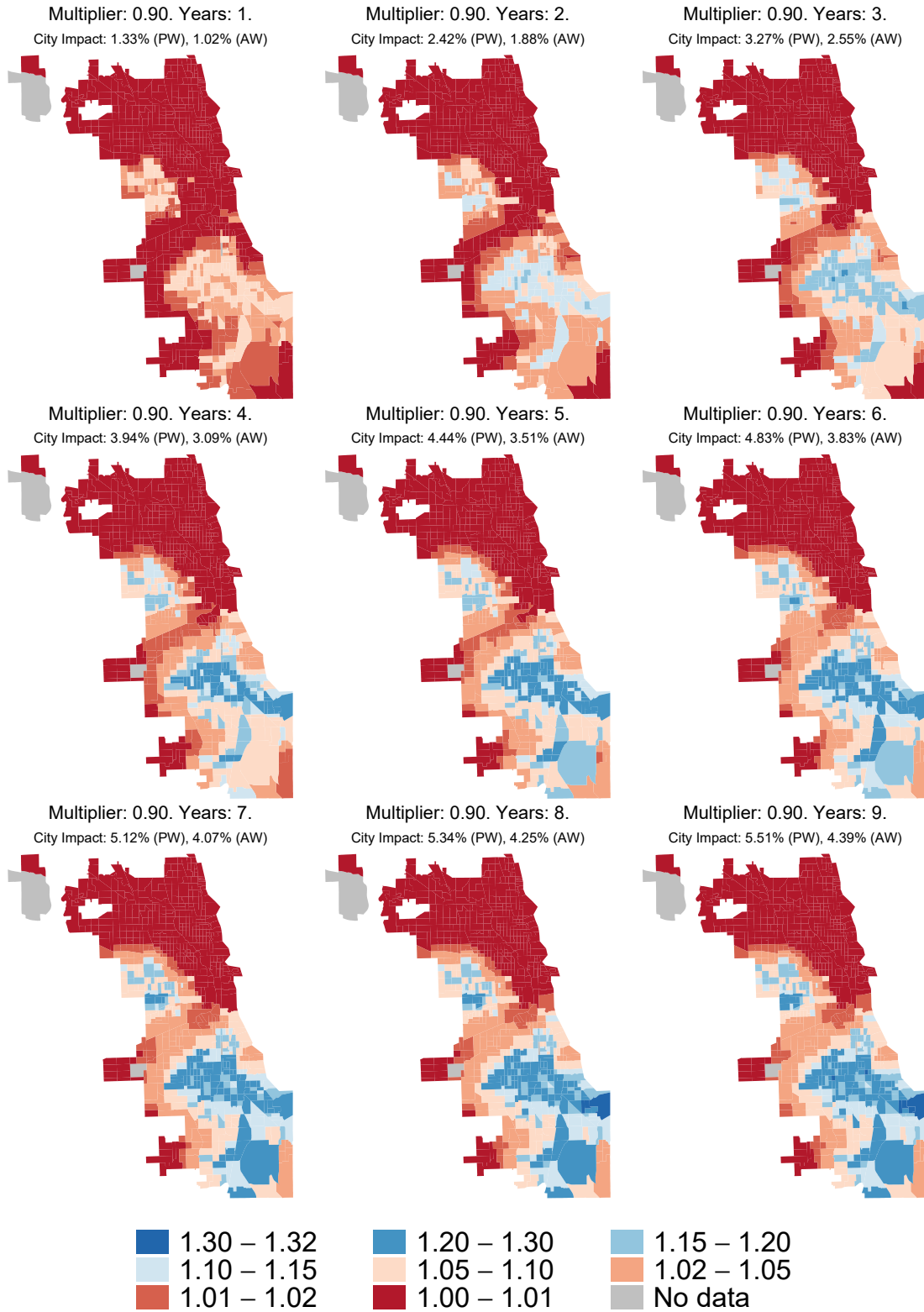
## H. ADDITIONAL SIMULATION FIGURES

In this appendix, I include a few additional simulation figures that are referenced in the main text.

Figure H.1 presents the full year-by-year results of the opportunity zones simulation for the baseline  $A_m$  multiplier 0.90. The figure shows clearly that the majority of the impact is seen in the first 5 years, after which the impact continues to grow slightly as the two simulations continue along different equilibrium paths, but generally the magnitude of the policy effect has already been realized.

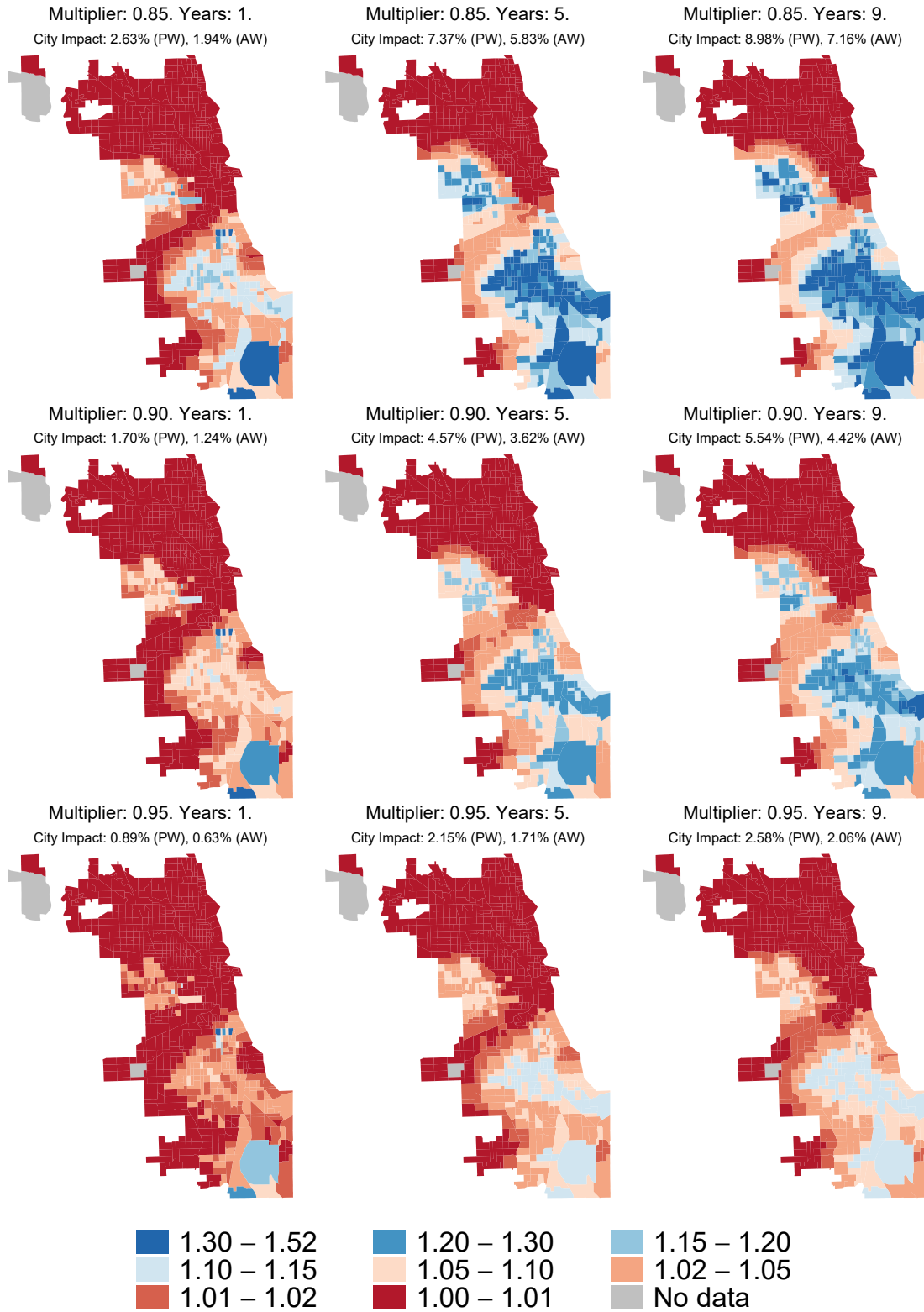
Figure H.2 presents the results of the opportunity zones simulation when opportunity zones are not subject to credit constraints. The figure shows that the long-run effects of the policy modeled in this way are nearly the same as the baseline simulation, but the effects on impact are larger, as some tracts adjust quickly to the new equilibrium path in a way that they could not with credit constraints.

Figure H.1: Opportunity Zone Simulation: Year-by-Year Price Impact



Note: PW is population-weighted price impact, AW is asset-weighted price impact

Figure H.2: Opportunity Zone Simulation: Year-by-Year Price Impact without constraints



Note: PW is population-weighted price impact, AW is asset-weighted price impact