The Socio-Economics of Neighbourhoods and Cities: Papers in Urban Economics

by

Jessica Lynn Burley

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy
Graduate Department of Economics
University of Toronto

Abstract

The Socio-Economics of Neighbourhoods and Cities:
Papers in Urban Economics

Jessica Lynn Burley
Doctor of Philosophy
Graduate Department of Economics
University of Toronto
2017

This thesis explores the socio-economics of neighbourhoods and cities and the role that changes to labour market opportunities and access to amenities play in determining housing demand and demographic shifts. Many cities are experiencing a spatial inversion in terms of neighbourhood demographics – city centers are becoming both wealthier and more expensive while peripheral, suburban neighbourhoods are becoming both poorer and less expensive. In a dynamic context, I consider the social and economic processes that contribute to neighbourhood change.

Chapter 1 contributes to the literature on neighbourhood change by providing a comprehensive characterization of the dynamics of New York City neighbourhoods. Using a structural breakpoint analysis I provide evidence of which neighbourhoods have changed, when they changed, and the magnitude of the change. I show a clear spatial and temporal pattern to neighbourhood change and I present several results: price increases are larger than price decreases; increases in housing demand precede increases in price growth; and downtown neighbourhoods are increasingly wealthier and a greater fraction white.

Chapter 2 explores the extent to which changing labour market characteristics contribute to the patterns of neighbourhood change. I construct a spatial-Bartik instrumental variable as an exogenous measure of neighbourhood labour demand, which is then used to generate predicted income growth rates and house price growth rates. I find that in New York City 21% of the variation in income growth rates and 41% of the variation in house price growth rates between 1990 and 2010 can be explained by exogenous labour demand shocks.

Chapter 3 focuses on the influence of the built environment on social behaviours. I find a strong and positive cross-sectional relationship between the built environment and social interactions. Once I address the endogenous decision of where to live, the significant effects found in the cross-section disappear. This implies that there is sorting of relatively social individuals into more walkable neighbourhoods. I find some evidence that this sorting is correlated with life-cycle changes that may affect both residential decisions and social relationships.

Acknowledgements

I would like to thank my advisor, William Strange, as well as my committee members, Nathaniel Baum-Snow, and Michel Serafinelli, for so much support and much needed guidance over the years. I am also grateful to Gordon Anderson, James Pesando, Michael Gilraine, Adam Lavecchia, and the graduate students and faculty in the Department of Economics for your questions, comments, and friendships over the years. Finally, thank you to my family for your unconditional support and encouragement; in particular Jacqueline for being my cheerleader, Miranda for sitting through countless practice presentations, and Sean for your calming reassurance in the presence of my less than calm tendencies.

Contents

List of Figures

List of Tables

Chapter 1

The Spatial and Temporal Patterns of Neighbourhood Change

1.1 Introduction

This chapter contributes to the literature on neighbourhood change by providing a comprehensive characterization of the evolution of New York City (NYC) neighbourhoods over the last 40 years. Since the 1990s, NYC has undergone many changes: crime rates have fallen significantly, the annual issuance of building permits for new development has quadrupled, and house prices have increased by almost 200%. These changes have been particularly concentrated in Manhattan with most neighbourhoods gaining in prices relative to the city on average. The outer boroughs on the other hand have actually seen declines in homeowner incomes and relative house prices. The primary goal of this chapter is to examine the spatial and temporal patterns of neighbourhood change; by focusing on a single city such as New York I shed light on within-city dynamics.

My initial analysis relies on trends in the sales price of properties around NYC. Using a detailed micro-level dataset for all property sales in NYC from 1974 to 2014, I construct a repeat-sales house price index and a price growth rate series. To analyze contemporaneous changes, I complement these data with homeowner incomes and race as reported on mortgage applications. I begin by presenting the patterns observed in the data: there has been a spatial inversion of neighbourhoods in NYC with central neighbourhoods becoming both richer and more expensive and peripheral neighbourhoods becoming both poorer and less expensive. While I focus specifically on NYC, this pattern has been observed in several cities around the United States. A number of recent papers address these trends in re-urbanization. In particular, Couture and Handbury (2016) show that most US cities have seen increases in young professionals living near central business districts since 2000 which has led to increasing resident incomes. Because of the relatively inelastic housing supply in downtowns, re-urbanization becomes particularly evident in increasing house prices. Between 1950 and 2000 approximately 66% of low-income census tracts across 35 cities moved up the economic ladder, while 56% of high-income census tracts moved down (Rosenthal and Ross (2015)).

To understand these patterns I begin by systematically characterizing the temporal patterns of neighbourhood change from 1974 to the present using a structural breakpoint analysis. In doing so, I provide evidence of which neighbourhoods have changed, when they changed, and the direction and magnitude of the change. The structural breakpoint procedure backs-out the point at which neighbourhoods change trajectories, as captured by movements in house prices.[1] Of the 157 NYC neighbourhoods considered in this chapter, I find statistically significant structural breaks in 94 of them. Of these, 52 are growing faster than the NYC average with house prices rising approximately 4.4 times faster in the breakpoint year relative to the year prior (I will often refer to these as the gentrifying neighbourhoods). The other 42 neighbourhoods are declining relative to NYC, with prices falling approximately 2.5 times faster at the estimated breakpoint relative to the year prior (similarly, I often refer to these as the declining neighbourhoods). After the breakpoint years, prices continue to rise or fall in the gentrifying or declining neighbourhoods, respectively.[2] The distribution of the timing and geography of the estimated breakpoints show spatial and temporal patterns that are consistent with the inversion observed in the data.

Neighbourhoods change over time for a variety of reasons. Rosenthal and Ross (2015) discuss the mechanisms that contribute to either the slow, yet relentless change of neighbourhoods, or the relatively rapid change. Several theories have been put forward to explain the dynamic process of neighbourhood change; I consider the observed changes in NYC in the context of invasion, filtering, and tipping theories. Invasion theories are initially credited to urban sociologist Burgess (1925) who developed the Concentric Zone Model. This model characterizes change as resulting from spatial pressure and competition for space. Guerrieri et al. (2013) implicitly make use of such a model. They characterize an equilibrium generated through a demand to live near richer households. Having richer neighbours generates a positive externality and as such, less expensive neighbourhoods adjacent to rich neighbourhoods are more likely to gentrify. Consistent with this, the results I find are indicative of spillovers across neighbourhoods in terms of house price appreciation rates – as increases to housing demand drive up prices in a particular neighbourhood, it is more likely that adjacent neighbourhoods will later experience the same. I find that a neighbourhood is two times more likely to have a positive estimated structural breakpoint following one in an adjacent neighbourhood and four times more likely to have a negative estimated structural breakpoint.

Closely related to invasion theories, tipping theories characterize neighbourhoods as often being observed in a state of change. Schelling (1971) proposed tipping models to explain racial segregation based on a preference function that exhibits a discontinuous 'jump', such that racial integration may be desirable until a point. Given this preference function, perfect segregation is the only stable equilibrium outcome. Sethi and Somanathan (2004) provide evidence in support of the tipping hypothesis with racial segregation remaining a persistent equilibrium despite the evolution of more tolerant preferences. Card et al. (2008) were the first to empirically estimate tipping models using a structural breakpoint analysis based on cross-sectional differences in census tract racial composition. While tipping models have

[1]The structural break estimation follows the methodology of Ferreira and Gyourko (2012). This is similar to the work by Card et al. (2008) on racial tipping. I discuss this in more detail in Section 1.3.2.

[2]Not all neighbourhoods experiencing price growth would actually be considered as having gentrified. For example, in Central Harlem, a historically low income neighbourhood in upper Manhattan, house prices have risen by about 324% between 1990 and 2014. Over the same period, in the Upper East Side, one of the most affluent neighbourhoods in New York City, house prices rose by a much smaller 48%. Both have an estimated statistically significant breakpoint but the Upper East Side would not likely be classified as a gentrifying neighbourhood.

primarily been discussed in the context of racial segregation, the idea that small changes build up and gain momentum over time can be applied to other aspects of neighbourhood change. In this chapter, I characterize this critical mass as being the changes in a neighbourhood's underlying fundamentals as they are reflected by changes in house prices.

Observing trend breaks, or tipping-points, in house price growth rates provides the first step towards characterizing neighbourhood change. After estimating the trend breaks in house price appreciation rates, I consider the change in neighbourhood demographics around these breaks; specifically, I look at homeowner incomes and racial composition. If the demand for living near downtowns is increasing and house prices are rising, the pool of residents who are able to purchase homes in these neighbourhoods will decrease. Furthermore, the correlation between incomes and race suggest that higher home prices will disproportionately affect low and middle income black families (Bouston and Margo (2013)). Therefore, a priori I expect to observe heterogeneity in the demographic characteristics of new homeowners around the estimated tipping points. I find evidence that among gentrifying neighbourhoods in New York City new homeowners are increasingly wealthier and are a greater fraction white. Among declining neighbourhoods, new homeowners are increasingly poorer but are not disproportionately black.

A different set of theoretical models – filtering models – see neighbourhood change as a process in which higher-income households prefer to live in new homes. As homes age, they are 'filtered' down to lower-income households (Hoyt (1933)). Rosenthal (2008) and Brueckner and Rosenthal (2009) provide empirical evidence in support of the filtering hypothesis. They observe how the aging and then redevelopment of the housing stock can drive income patterns seen across several MSAs; as central city housing stock is redeveloped, there is an inflow of higher-income households to previously low-income communities. While I am not able to speak to this directly, I do provide evidence that in NYC increases in the volume of sales (i.e. home demand) precedes increases in house prices within gentrifying neighbourhoods. I also provide some descriptive evidence that in NYC the pattern of building permits issued for substantial alterations or new development follow a similar distribution to that of the estimated structural breakpoints.

In summary, I provide several key results: first, neighbourhood change is spatially and temporally correlated and this relationship is stronger among declining neighbourhoods; second, price increases are larger than price decreases in absolute terms; third, gentrification is happening relatively earlier than decline; fourth, increases to housing demand precede increases in price growth among gentrifying neighbourhoods; and fifth, demographic characteristics of new homebuyers are changing in a way that is consistent with the typical gentrification anecdote. The remainder of this chapter proceeds as follows: Section 1.2 presents the data used to characterize neighbourhood change. Section 1.3 presents the key facts observed in the data, the methodology of the structural breakpoint procedure, and the resulting temporal patterns. Section 1.4 discusses the economic significance of these temporal patterns, while Section 1.5 explores contemporaneous changes in incomes and demographics. Section 1.6 discusses these results in the context of neighbourhood change theories and concludes.

1.2 Data and Descriptive Statistics

The data used in this chapter come primarily from New York City property sales transactions. From these I construct my annual measure of house price movements (*repeat sales house price index*) as well as my estimate for neighbourhood demand (*transactions volume*). In considering the contemporaneous changes to neighbourhood demographics, I supplement this with the FFIEC Home Mortgage Disclosure Act (HMDA) data. This provides a measure of the annual average incomes as well race for new homeowners within neighbourhoods. These datasets are both aggregated at the neighbourhood level according to the NYC Department of City Planning's neighbourhood definitions and are based on 2010 census tract geographies. There are 196 defined neighbourhoods in New York City and after removing airports, parks and cemeteries, and neighbourhoods comprised entirely of private or subsidized housing developments,[3] I am left with 157 neighbourhoods. These neighbourhoods are roughly the area of a zipcode and on average cover 13 census tracts. Table 1.A1 of the Appendix lists the full set of NYC neighbourhoods used throughout this chapter.

New York City Sales Data: The Furman Center at New York University maintains data for all property sales in NYC from 1974 to 2014.[4] These files include information on address (tax lot identifier), sales price, and date of sale for condominiums, single family, and multifamily homes. Using the property address this sales data is merged with publicly available data from the NYC Department of Planning; this provides building classification codes. I remove buildings coded as co-operative housing units and apartment buildings. For construction of the repeat-sales index I also remove buildings that changed either classification codes or the number of residential units over the sample period.

Using the NYC Department of Planning's neighbourhood definitions (*Neighbourhood Tabulation Areas, or NTAs*), I allocate properties to their respective neighbourhood. In the final data set, there are a total of 1,261,067 sales across all five boroughs in New York City. On average there are 288 repeat sales per year, per neighbourhood. Summary statistics are presented in Panel A of Appendix Table 1.A2. The average sales price (in 2014 dollars) increased from $148,321 in 1974 to $709,941 in 2014. As expected, this was largely driven by growth in Manhattan and Brooklyn, with prices increasing from $167,334 to $1,944,913, and $120,745 to $510,421, respectively. Median house prices increased more modestly, albeit still by about 200%.

Constructing a House Price Index

Using the NYC sales data, I construct a repeat-sales house price index for each neighbourhood in New York City following the methodology provided by the Furman Center at NYU. The repeat sales index controls for housing characteristics by using sales of the same property but comes at the cost of excluding properties that sold only once. I differentially weight properties that have sold more or less frequently and more or less recently to address potential changes to the either the physical structure or to the external environment. Details of the regression techniques used to construct the house price index (HPI) are included in Appendix 1.B. Figure 1.B1 presents the HPI for NYC and for each borough. Between

[3]These neighbourhoods include: Stuyvesant Town/Cooper Village, a private residential development; Starrett City, a housing development; and Rikers Island, a jail complex.

[4]The Furman Center has obtained this data under exclusivity from the New York City Department of Finance.

1974 and 2005 house prices were growing at roughly the same rate around NYC. Post-2005, Manhattan experienced the biggest appreciation in prices.

Because of the cyclical movement of house prices at the macroeconomic level, to calculate structural breakpoints in house price trends I use detrended (relative to NYC as a whole) house price movements. Figure 1.B2 shows the detrended house price indices for each neighbourhood in each borough. There is a large amount of heterogeneity both across boroughs and within boroughs. For example, Manhattan's house prices have been getting more expensive relative to NYC, whereas neighbourhoods in the Bronx and Queens tend to be getting less expensive. Brooklyn on the other hand has a split of neighbourhoods that are either becoming more expensive or less. Taking these indices I construct price growth rate series; this, and the structural breakpoint estimation methodology are discussed in Section 1.3.2.

Home Mortgage Disclosure Act (HMDA) Data: For my measure of neighbourhood income, I make use of the HMDA files.[5] The HMDA data include mortgage loan amounts by year, census tract, loan status, and applicant characteristics such as race and income, starting in 1990 through to 2014.[6] Summary statistics are presented in Panels B and C of Appendix Table 1.A2. Average incomes of homebuyers in New York City increased from $168,940 (median, $117,730) in 1990 to $271,330 in 2014 (median, $145,000). Note that while I do not specifically match the mortgage data to the NYC sales data, I only use originated owner-occupied mortgages and exclude mortgages that are recorded as being greater than four times the applicant's income. Using race as it is recorded in the HMDA files, the total number of approved and originated applications decreased for both black and white households between 1990 and 2014. Mortgages for black households decreased in NYC by 79% with the largest decrease in Brooklyn. Mortgages for white households on the other hand decreased by only 36%, with the largest decrease seen in Queens and Staten Island.

1.3 The Patterns of Neighbourhood Change

1.3.1 Patterns Observed in the Data

Forty years ago downtowns in the United States were declining in population, household incomes, and home prices; today this is no longer the case. Since 1974 the spatial distribution of house prices and homeowner incomes in New York City has inverted – perviously affluent suburban neighbourhoods in the outer boroughs are now declining in prices and incomes while some of the historically poorest neighbourhoods in upper Manhattan and downtown Brooklyn are rising in prices and incomes.

Using the New York City property sales data, Figure 1.1 shows the spatial inversion of house prices that has occurred in NYC between the downtown and peripheral neighbourhoods. Snapshots are shown for 1974, 1995, and 2014, each depicting the quantiles of the median house price distribution in 2014$s -

[5] The Home Mortgage and Disclosure Act was implemented in 1975 by the Federal Reserve Board and requires all lending institutions to report all mortgage applications. This was done to identify discriminatory lending practices.

[6] Using HUD x-walk files I converted 1990 and 2000 census tracts to 2010 census tract boundaries. Census tracts that were not able to be reliably applied to 2010 boundaries were removed from the sample. This represents less than half a percent of records.

darker shades are more expensive. In 1974 the peripheral neighbourhoods in Queens and Brooklyn contained the most expensive properties in NYC. By 2014 this pattern had almost completely reversed with Manhattan and downtown Brooklyn now holding the most expensive neighbourhoods in the City.

Consistent with the pattern in median house prices, initially lower income downtown neighbourhoods have also become richer while the suburban peripheral neighbourhoods have become relatively poorer. Figure 1.2 shows the corresponding income patterns using the HMDA data. In particular, upper Manhattan (i.e. Central Harlem) and downtown Brooklyn and Queens are some of the wealthiest neighbourhoods by 2010. These observed changes to neighbourhood house prices and incomes are consistent with the trending movement towards the downtown core. The remainder of this chapter estimates the spatial relationship of these changes and the temporal evolution of these patterns. In the next section I present the structural breakpoint estimation procedure.

1.3.2 The Timing of Neighbourhood Change

Structural Breakpoint Estimation

To explain the observed patterns of neighbourhood change, I begin by identifying changes as they are reflected by movements in house prices using a structural breakpoint analysis. To find breaks in the house price appreciation rate series, I use a search procedure developed by Card et al. (2008) and used by Ferreira and Gyourko (2011).[7] Under the assumption that neighbourhood change is partially characterized by rising prices, I am interested in pinning down when there is a change in house price trends.

I allow for prices to be increasing or decreasing however, I restrict my analysis to allow for at most one change-point. As such, neighbourhoods that cycle up and down even after removing the macro trend are assumed to not have a breakpoint.[8] In order to ensure that I am not picking up noise in the house price data, I estimate the price growth rates over a smoothed local polynomial of the HPI.

Price growth rates are defined as:

$$\mathrm{PG}_{n,t} = \frac{\mathrm{HPI}_{n,t}}{\mathrm{HPI}_{n,t-1}} - 1 \tag{1.1}$$

Where, $\mathrm{PG}_{n,t}$ represents the price growth in neighbourhood n at time t.

To identify the structural breakpoints I take the price growth rate series for each neighbourhood and estimate the following equation:

$$\mathrm{PG}_{n,t} = a_n + d_n \mathbb{1}[y_{n,t} \geq y^*_{n,t}] + \epsilon_{n,t}, \quad \text{for } T_{n,0} = 1974 < y^*_{n,t} < T_{n,T} = 2014 \tag{1.2}$$

[7] Outside of financial studies and macroeconomic time-series literature, the estimation of structural breaks was first used to describe racial segregation by Card et al. (2008). Ferreira and Gyourko (2011) document the heterogeneity in house price booms (pre 2008 bust) across the United States.

[8] In the context of neighbourhood change it is uncommon that within 40 years a neighbourhood both declines and gentrifies, so their exclusion from the final analysis seems appropriate. Indeed, this was confirmed by Rosenthal and Ross (2015).

The price growth rate series, $PG_{n,t}$, is regressed on a series of {1,0} dummy variables for each potential breakpoint year; the year that maximizes the R^2 is identified as the structural breakpoint. For example, to test if 1990 is a structural breakpoint, $\mathrm{PG}_{n,1990}$ is regressed on a dummy variable equalling {0} if $y < 1990$ and equalling {1} if $y \geq 1990$. This is done for all years from 1974 to 2014 and for all neighbourhoods, n. The intuition in estimating the above equation is to identify the year in which the change in the price growth rate series best predicts the growth rate series itself.

1.3.3 Statistically Significant Breakpoints

Of the 157 neighbourhoods in NYC, 94 ever have a statistically significant breakpoint. Of these, 52 are growing faster than the New York City average (i.e. gentrifying or "tipping up") and 42 are growing slower (i.e declining or "tipping down"). Figure 1.3 depicts examples of the estimated breakpoints for two neighbourhoods, Central Harlem in upper Manhattan and Canarsie, a peripheral neighbourhood in Brooklyn. Central Harlem is growing faster than New York City whereas Canarsie is declining. The top figures depict the NYC mean-detrended house price index along with the smoothed local polynomial upon which price growth rates are calculated. The bottom figures depict the corresponding price growth rate series. The vertical lines are drawn at the estimated structural break point – in both cases, 1997 was estimated as the year of the trend break.[9]

The timing of the estimated breakpoints ranges from 1978 to 2012. For both growing and declining neighbourhoods, the mass of changes are observed between 1995 and 2005. Figure 1.4 (a) shows the distribution of estimated breakpoint years for neighbourhoods growing faster than NYC. Approximately 20% of these breakpoints occurred in 2001, 13% in 2000 and 10% in 1997. Figure 1.4 (b) shows the distribution of estimated breakpoint years for neighbourhoods declining relative to NYC. Approximately 45% of these breakpoints occurred in either 2003 or 2004 and are predominantly comprised of a cluster of 10 contiguous neighbourhoods in Queens.

1.3.4 The Geographic Concentration of Structural Breakpoints

The complete spatial pattern of breakpoints is depicted in Figure 1.5: (a) shows the gentrifying neighbourhoods while (b) shows the declining neighbourhoods. There is visual evidence of the clustering of neighbourhoods around the timing of observed changes. Furthermore, the spatial patterns described earlier are reproduced here. There is a concentration of neighbourhoods getting more expensive in Manhattan and the inner parts of Brooklyn and Queens and a concentration of neighbourhoods getting less expensive in the outer parts of Brooklyn, Queens, and some of the the Bronx.

To quantify the strength of this geographic relationship, I estimate how having a bordering neighbourhood with a previously estimated positive or negative structural break increases the probability of the same. Table 1.1 presents the results. The dependent variable is an indicator equal to one in the year of the estimated structural break and zero in all other years. This is regressed on a series of additional indicator variables for (1) having a neighbour with a positive breakpoint (denoted, $d_j > 0$), (2) having

[9] Figure 1.A1 of the Appendix shows two examples of neighbourhoods where no statistically significant breakpoint is found. In both cases there is no clear trend in the house price index.

a neighbour with a negative breakpoint ($d_j < 0$), (3) having this breakpoint occur in the previous year ($t-1$), (4) having this breakpoint occur two years prior ($t-2$), or (5) having this breakpoint occur three years prior ($t-3$). In each specification neighbours are defined as contiguous neighbourhoods.

Having a neighbour with a previously estimated positive structural breakpoint ($d_j > 0$) increases the probability of tipping up by 5.5%; if this previously estimated breakpoint occurred in the year prior, this probability increases to 7.4%. Column (4) is suggestive of a strong temporal pattern among gentrifying neighbourhoods. For neighbourhoods that are declining, the spatial pattern is even stronger. Having at least one neighbour decline perviously ($d_j < 0$) increases the probability of tipping down by 15.2%; the timing of this previously estimated breakpoint does not appear to have an effect. Given the frequencies of tipping up or down in the data, having a neighbour tip up in the previous year makes it almost 2 times more likely that a neighbourhood gentrifies in the current year, while having a neighbour ever tip down makes it 3.3 times more likely that a neighbourhood declines. The rate at which prices are changing has implications for the rate at which neighbourhoods are changing. I next discuss how quickly prices are growing or falling around the estimated structural breakpoints.

1.4 The Magnitude of the Changes in Price Growth Rates

So far I've shown that neighbourhoods tip up earlier than they tip down and both types of changes are geographically correlated. Before discussing the magnitude of the changes around the breakpoints, I briefly discuss the speed of the estimated price growth changes. Among the growing neighbourhoods, between 1974 and 2014 prices rose by about 66.81%; this was driven largely by house price increases in Manhattan and Brooklyn. In the year of the breakpoint prices were rising on average by 1.08%.[10] Declining neighbourhoods on the other hand saw their prices decrease by 67.41% between 1974 and 2014, with an average price growth rate in the year of the breakpoints of -1.20%.[11] There is a large amount of heterogeneity across neighbourhoods.[12] For example, at the estimated structural breakpoint prices in Central Harlem were growing by 2.87% whereas in the Upper East Side prices were growing by 1.01%. Between 1990 and 2014, prices in Central Harlem increased by over 324% whereas in the Upper East Side they increased by 48.24%. Neighbourhoods with the biggest increases in prices are those starting with some of the lowest price levels in the City – those in upper Manhattan and close to downtown Brooklyn.

The next question I ask is how much faster are prices growing, or how much faster are prices falling, in the year of the estimated breakpoints relative to the prior years. In other words, are 1.08% and -1.20% statistically different and economically meaningful, compared to prior years. I estimate the following equation (Ferreira and Gyourko (2011)):

$$\mathrm{PG}_{n,t} = \gamma^i(y^{**}_{n,t}) + \eta_n + \epsilon_{n,t} \tag{1.3}$$

[10]Keep in mind that this is relative to prices in New York City as a whole. Average raw prices were rising at a rate of 9.08% in the breakpoint years. For reference, the S&P/Case-Schiller seasonly-adjusted national home price index increased 5% from April 2015-2016. Table 1.A3 summarizes price growth rates for neighbourhoods that are growing (i.e. prices are higher to the right of the breakpoints) and declining (i.e. prices are lower to the left of the breakpoint).

[11]As above this is relative to NYC as a whole. Average raw prices decrease by 4.9% in the breakpoint years.

[12]Table 1.A3 includes the detailed changes for each neighbourhood with an estimated structural breakpoint; the largest 20 neighbourhoods are highlighted.

As before, $\text{PG}_{n,t}$ is the price growth in neighbourhood n at time t; $\gamma^i(y^{**}_{n,t})$ are a set of year dummies relative to the estimated breakpoint indexed by i, where $y^{**}_{n,t}$ is the estimated breakpoint; η_n are neighbourhood fixed effects.

Table 1.2 presents the estimates for the magnitudes of the breakpoints. The first row is restricted to neighbourhoods that are growing faster than NYC while the second row is restricted to neighbourhoods that are declining relative to NYC. First, prices are growing approximately 1.48 percentage points faster in the two years around the breakpoint (t and $t+1$) compared to the two years prior ($t-1$ and $t-2$). Six years following the estimated structural breakpoint, prices continue to grow 2.63 percentage points faster. In the base year (i.e. the two years prior to the breakpoint) the average price growth is 0.43%. This implies that prices are growing 4.4 times faster in the two years following the breakpoint and 7.2 times faster six years following the breakpoint. Figure 1.6 (a) presents the graphical analogue to the regression results. The vertical line is drawn in the year prior to the estimated breakpoint (the reference category, year $t-1$ and $t-2$).

Second, growth rates are approximately 1.18 percentage points lower in the breakpoint year compared to the year prior. In the base year the average price growth is -0.79%. This implies that prices are falling approximately 2.5 times faster following the estimated breakpoint. Six years following the breakpoints, prices continue to fall by 4.33 percentage points or, 6.5 times faster than prior to the breakpoint. Figure 1.6 (b) presents the graphical analogue with the vertical line drawn at the year prior to the estimated breakpoint.

For both gentrifying and declining neighbourhoods there is no evidence of pre-trends prior to the estimated structural breakpoints. In other words, I am capturing the point in which prices truly begin to trend upwards, or downwards. Given these estimated structural breakpoints in neighbourhoods' house price appreciation rates, the remainder of this chapter is dedicated to explaining how changes to homeowner characteristics may be correlated with these spikes in house prices. I will look at both contemporaneous changes in homeowner income and race; but, prior to doing so I first look for evidence of changes in the demand for housing by looking over the volume of sales.

1.4.1 Changes in the Volume of Sales

To provide some evidence of the extent to which changes in housing demand have driven price growth, I next explore the magnitude of changes in the quantity of sales around the previously estimated breakpoints. Looking over changes in quantities has the advantage of including new residential buildings that are excluded in the repeat sales index.

Using the NYC sales data, I count the annual number of sales in each neighbourhood and I estimate the following equation:

$$\Delta ln(Quantities) = \lambda^i(y^{**}_{n,t}) + \eta_n + \epsilon_{n,t} \tag{1.4}$$

As in Equation 1.3, $\lambda^i(y^{**}_{n,t})$ are a set of year dummies relative to the breakpoints estimated in Section 1.3.2; $y^{**}_{n,t}$ are the estimated breakpoints, and η_n are neighbourhood fixed effects.

Table 1.3 presents the results and is analogous to Table 1.2 – the first row is restricted to positively growing neighbourhoods while the second row is restricted to negatively growing neighbourhoods. Quantities are increasing by 1.85 percentage points in the two years around the breakpoint (t and $t+1$) compared to the two years prior ($t-1$ and $t-2$). However, quantities are also increasing in the two years prior to the reference years (i.e. $t-3$ and $t-4$). This suggests a pre-trend with residential demand increasing prior to prices increasing.[13] In comparison, there is no evidence of demand decreasing prior to prices decreasing in neighbourhoods with a negative breakpoint. In fact, there is only evidence of demand decreasing several years following the estimated breakpoints – after six years ($t \geq 6$) quantities are 1.93 percentage points lower.

Given this evidence of demand increases preceding price increases, I search for structural breakpoints over the quantity of sales; I estimate Equation 1.2 with $\Delta ln(Quantities)$ on the left hand side. Of the 157 neighbourhoods in NYC, 132 have statistically significant breakpoints over quantities.[14] Figure 1.7 presents a plot of the estimated tipping point years for both prices – on the y-axis – and quantities – on the x-axis. In 87.5% of the cases, the year that a neighbourhood experiences a structural break in demand is earlier than the structural break in prices. Furthermore, a neighbourhood with a significant increase in the quantity of sales is on average 27% more likely to also experience a significant increase in prices and the increase in prices is more likely to follow the increase in demand.[15]

The Issuance of New Building Permits: So far I have shown evidence that for gentrifying neighbourhoods, the volume of sales is increasing prior to prices increasing. Among these neighbourhoods I am interested in seeing the corresponding pattern of newly issued building permits for evidence in support of a filtering type story – if higher income households demand new houses, new development or redevelopment could attract higher income households to the CBD. The City of New York Open Data repository contains historical records for all building permits requested since 1990. Restricting this to first-time permits that are requested and issued for major alternations or new development, I have a sample of 146,789 permits. Between 1990 and 2010 the number of permits issued increased by over 200%. Figure 1.A4 (a) shows the histogram for the total number of permits that were issued in New York City. When looking separately across the boroughs (Figure 1.A4 (b)), there is no evidence of a increasing number of permits issued in Manhattan compared to the other boroughs. However, while Manhattan has had fewer permits issued the majority of them have been for redevelopment. This suggests that the increase in prices seen across most Manhattan neighbourhoods may be due to the inelasticity of housing supply. While further work is necessary to fully understand this channel, there is some suggestive evidence that the trend breaks in price growth rates have been demand driven and are particularly prominent in inelastically supplied neighbourhoods. So far I have shown evidence that the increase in prices seen in NYC are driven by an increase in the demand to live in particular neighbourhoods. Given this, I next look at some of the observable characteristics of these demanders. In particular, I consider changes in the income and race of home buyers around the originally estimated tipping points.

[13] The graphical analogue is presented in Figure 1.A2 of the Appendix. Sales were clearly trending upwards prior to the estimated breakpoints.

[14] Note that this is larger than the 94 identified breakpoints over prices. This suggests that by using a repeat sales index I may be missing some of the demand for neighbourhoods with new housing development (i.e. new condo or apartment buildings).

[15] There is no correlation among neighbourhoods with negative structural breakpoints in house price growth rates which further supports the results presented in Table 1.3.

1.5 Contemporaneous Changes

1.5.1 Income Changes

In this section I investigate how income changes are correlated with the timing and magnitudes of the trend breaks in house price growth rates. My measure of income comes from the HMDA data files for mortgage applicants. The subset I work with is restricted to approved and originated, owner-occupied mortgages. As such the incomes I'm looking at are for new homeowners in a neighbourhood.

Panel A of Table 1.4 presents the results; Figure 1.A3 presents the graphical analogue. The outcome variable is the change in log income, $\Delta ln(Income)$. As before, to measure the magnitude of income changes around the estimated breakpoints this is regressed on a set of relative dummies (Equation 1.3). For neighbourhoods with positive breakpoints, the percentage change in incomes is 0.54 percentage points higher (i.e. growing 1.4 times faster than the base year of 1.35%) around the breakpoints than in the two years prior. Furthermore, incomes continue to rise several years following the breakpoint – four to five years following the breakpoint ($t+4$, $t+5$) the percentage change in incomes is 0.84 percentage points higher. The second row of Panel A presents the analogous results for neighbourhoods that are declining in prices. Around the breakpoints, the percentage change in neighbourhood incomes is 0.43 percentage points lower (i.e. growing 0.7 times slower than the base year of 1.41%) than in the two years prior. Following the breakpoints, incomes continued to decrease at increasing rates; after six years, the percentage change in income is 1.89 percentage points lower than it was in the two years prior to the breakpoint.

There is no observable pre-trend in income changes prior to the estimated breakpoints for either gentrifying or declining neighbourhoods. This suggests that the correlation between income growth and house price growth around the estimated tipping points is positive and rising house prices are contributing to rising neighbourhood incomes. Next I look at changes in the number of originated mortgages by black and white households around the trend breaks.

1.5.2 Racial Changes

In this section I investigate how changes in the reported race of new homebuyers is changing around the estimated breakpoints. As has been shown in the tipping points literature, race and racial preferences have strong implications for the degree of segregation in a city. If the new home buyers in the gentrifying neighbourhoods of NYC are predominately white, this may have implications for the rate at which neighbourhoods continue to change. As with income, my measure of race comes from the HMDA data files for mortgage applicants. I focus only on the number of black and white homebuyers due to inconsistencies in the coding among other racial categories, in particular hispanic.

Panel B of Table 1.4 presents the results for the percentage change in purchases by black households and Panel C presents the results for white households. For both black and white households the fraction of purchases was increasing prior to the breakpoints in both gentrifying and declining neighbourhoods. What is interesting, but perhaps not surprising, is the heterogeneity between the two groups around the estimated breakpoints. Among neighbourhoods that are gaining in prices, purchases by both black and white households were increasing several years prior to the breakpoints, however, at the breakpoint, white households are continuing to increase their purchases of new homes while there is no statistically significant change among black households. In the base year prior to the breakpoint (i.e. $t-1$, $t-2$) the average percentage change among black households is 0.98%; for white households it is 5.45%. Six years following the estimated breakpoints, purchases by both white and black households are decreasing at rates of -4.11% and -9.32%, respectively. In these gentrifying neighbourhoods, purchases by both black and white households decrease but at a faster rate for black households – the percentage decrease among black households is 2.27x larger than it is for white households.

Among neighbourhoods that are declining in prices, the percentage of black buyers decreases around the estimated breakpoints and continues to fall at increasingly greater rates in every year after. For white households there is no significant change until six years following the estimating breakpoints. In the base year prior to the breakpoints, the average percentage change among black households is 1.99%; for white households it is 5.33%. Six years following the estimated breakpoints, both groups are demanding fewer houses at rates of -17.5% and -10.2% for black and white households respectively. Among this subset of declining neighbourhoods, the percentage decrease among black households is 1.7x larger than it is for white households. Overall, the fraction of black households purchasing homes is decreasing but this fraction is decreasing faster in gentrifying neighbourhoods than in declining neighbourhoods.

1.6 Discussion and Conclusion

This chapter is an investigation of neighbourhood change in New York City. Using micro-level data I explore the within-city dynamics of both gentrification and decline. Over the last 20 years many cities in the United States have experienced a resurgence in their downtowns. Similarly, New York City exhibits a spatial inversion in neighbourhood incomes and house prices, with the downtown becoming wealthier and more expensive and the peripheral neighbourhoods becoming poorer and less expensive. There are several theories to describe the dynamic process of neighbourhood change: tipping, invasion, and filtering theories. While these theories rely on different mechanisms behind neighbourhood change, they are not mutually exclusive and I provide some evidence in support of each.

First, using a structural breakpoint analysis I characterize the spatial and temporal patterns of neighbourhood change. Overall, I find that neighbourhood decline has a stronger spatial relationship than gentrification: a neighbourhood in decline makes an adjacent neighbourhood four times more likely to also exhibit decline; this is compared to two times more likely for gentrifying neighbourhoods. This provides suggestive evidence of spillovers, or invasion, as generating the observed patterns of neighbourhood change.

In employing a structural breakpoint analysis I am testing the merits of neighbourhood change as a process involving a build up towards a critical mass which once achieved, causing a neighbourhood to "tip". By estimating trend breaks across neighbourhood house price appreciation rates, I assume that changes to underlying fundamental are absorbed into house prices. Of the neighbourhoods in New York City, I find that 52 are growing faster than the NYC average (i.e. they tip up). In these neighbourhoods prices are growing 4.4 times faster in the breakpoint year compared to the year prior and thus are economically meaningful. I find that 42 neighbourhoods are growing slower than the NYC average with prices falling 2.5 times faster in the breakpoint year compared to the year prior. These neighbourhoods also continue to decrease in price growth rates several years following the breakpoints.

Following this, I provide some evidence to the relationship between volume of sales and trend breaks in house prices. I find that in almost 88% of the cases a neighbourhood exhibits changes in the volume of sales prior to changes in the house price growth rates. While theres no clear evidence of houses being redeveloped disproportionately around the breakpoints, the majority of the building taking place in Manhattan is redevelopment (as opposed to new construction). This suggests that the increase in prices is driven by an increase in the demand to live in a particular neighbourhood and that redevelopment may be making these neighbourhoods more attractive. This leads me to consider the demographic characteristics of the demanders and the heterogeneity between gentrifying and declining neighbourhoods.

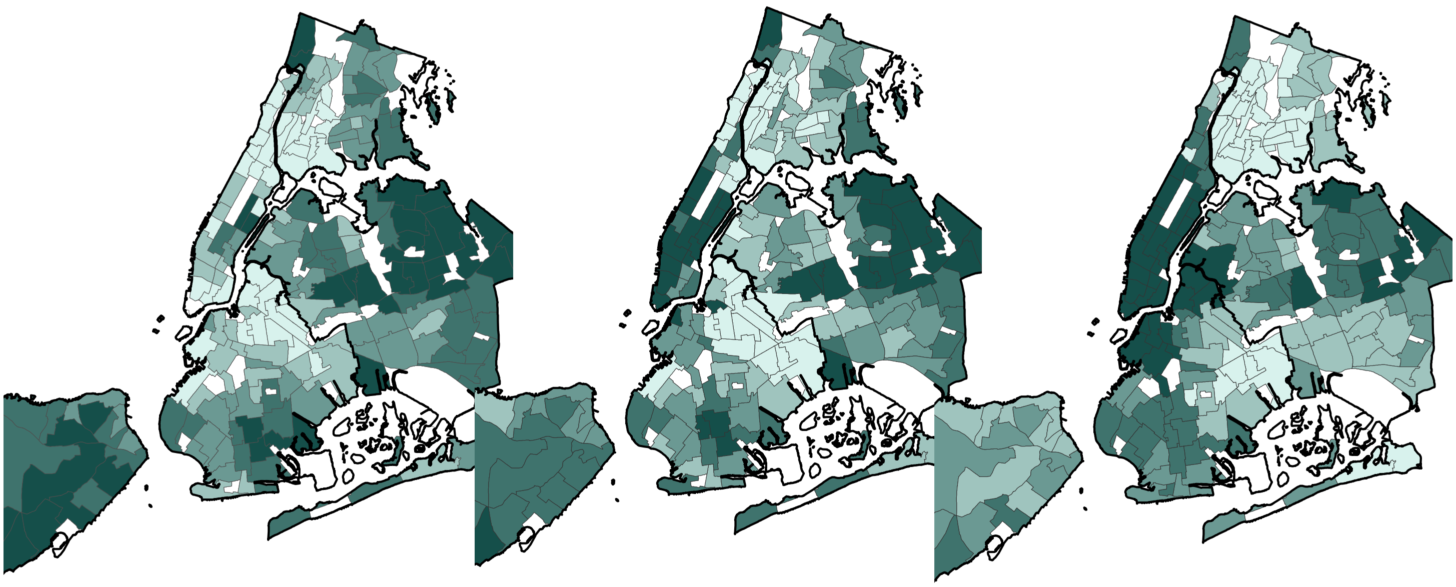
Looking first at reported incomes for new home buyers, I find that among the gentrifying neighbourhoods new homeowners are increasingly wealthier with the percentage change in homeowner income continuing to increase up to five years following the estimated breakpoint. In declining neighbourhoods, I find that the percentage change in incomes becomes increasingly lower and continues to decrease six years beyond the estimated breakpoints. As with house prices, incomes in gentrifying neighbourhoods are rising faster than they are falling in declining neighbourhoods, but the fall in incomes is more persistent.

Given the strong correlation between income and race I also look at changes in the number of black and white homebuyers around the estimated trend breaks. I find that overall homeownership is going down among both groups however, purchases by black households in the gentrifying neighbourhoods are decreasing at a faster rate than for white households. Furthermore, in the year of the estimated breakpoints purchases by white households are actually increasing. This holds true in declining neighbourhoods with purchases by black households decreasing faster than those of white households. However, in these neighbourhoods the rate of decline is slower than in gentrifying neighbourhoods. In other words, the fraction of new black homebuyers is decreasing faster in the gentrifying neighbourhoods than in the declining neighbourhoods.

This chapter sheds light on the within-city dynamics of neighbourhood change by providing a comprehensive characterization of the spatial and temporal patterns of neighbourhood change in New York City. Further work looking at why we are observing these changes across the United States is necessary. While I provide descriptive evidence in support of the anecdotal gentrification story, I have left an explanation of what is inherently driving this demographic shift in locational preferences to future work.

Figures

Figure 1.1: Median House Prices across New York City Neighbourhoods



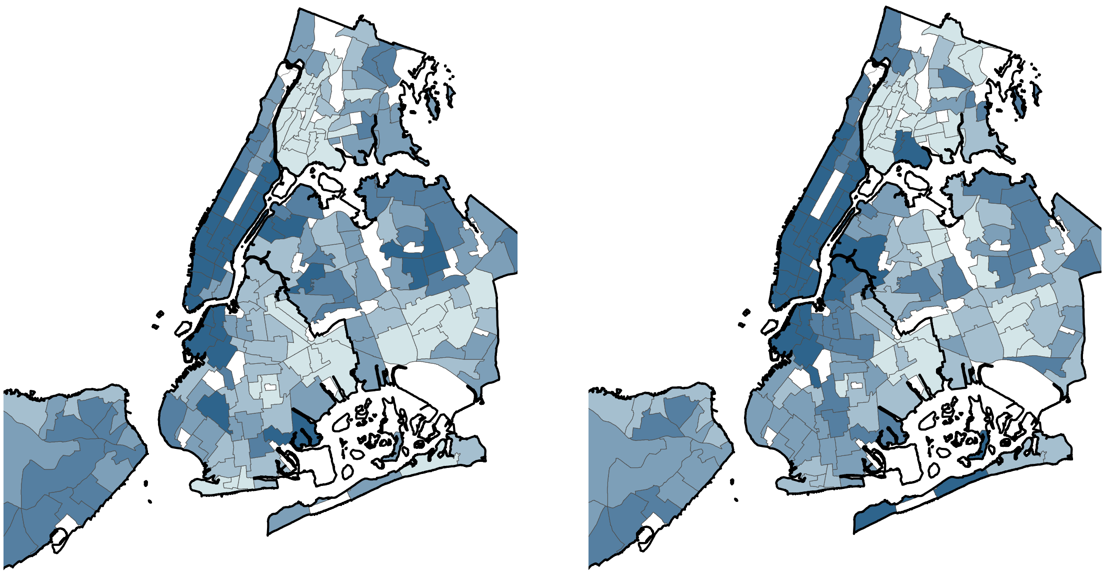
(a) 1974 **(b)** 1995 **(c)** 2014

Source: Median house prices, New York City property sales data. All values are in 2014 $s.
Notes: The above figure depicts the distribution of median house prices across New York City in 1974, 1995, and 2014. The darker-shaded neighbourhoods are the most expensive in terms of where they fall in the distribution, while the lighter-shaded neighborhoods are the least expensive. No constant-quality adjustments have been made to the house prices.

Legend for Fig. 1.1

Quantile of the distribution		
	1st	(most expensive)
	2nd	
	3rd	
	4th	
	5th	(least expensive)

Figure 1.2: Median Incomes across New York City Neighbourhoods



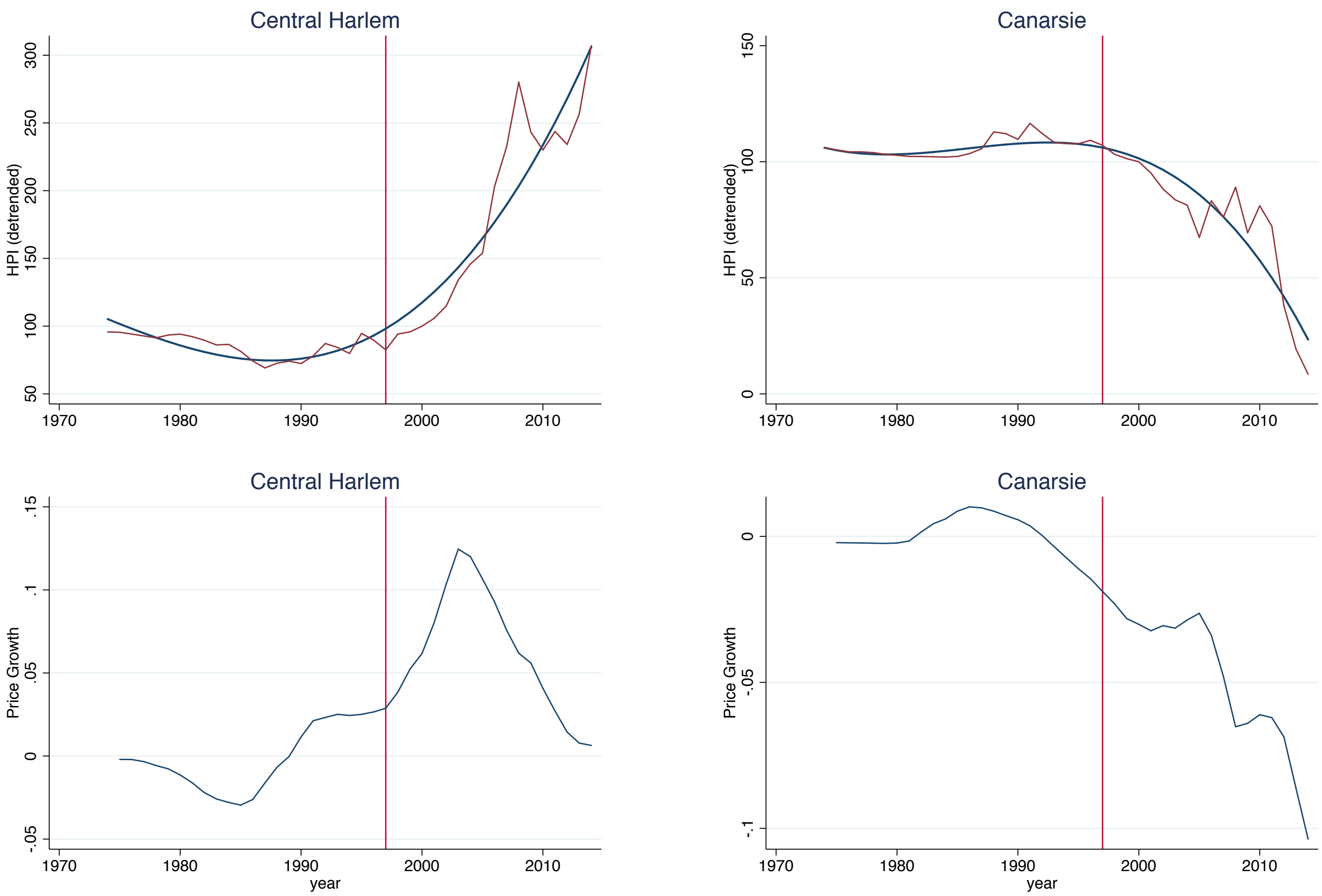
(a) 1990 **(b)** 2010

Source: Median incomes, Home Mortgage Disclosure Act data. All values are in 2014 $s.
Notes: The above figure depicts the distribution of median incomes across New York City in 1990 and 2010. The darker-shaded neighbourhoods are the wealthiest in terms of where they fall in the distribution, while the lighter-shaded neighborhoods are the poorest.

Legend for Fig. 1.2

Quantile of the distribution		
	1st	(wealthiest)
	2nd	
	3rd	
	4th	
	5th	(poorest)

Figure 1.3: Neighbourhoods with Estimated Breakpoints (examples)

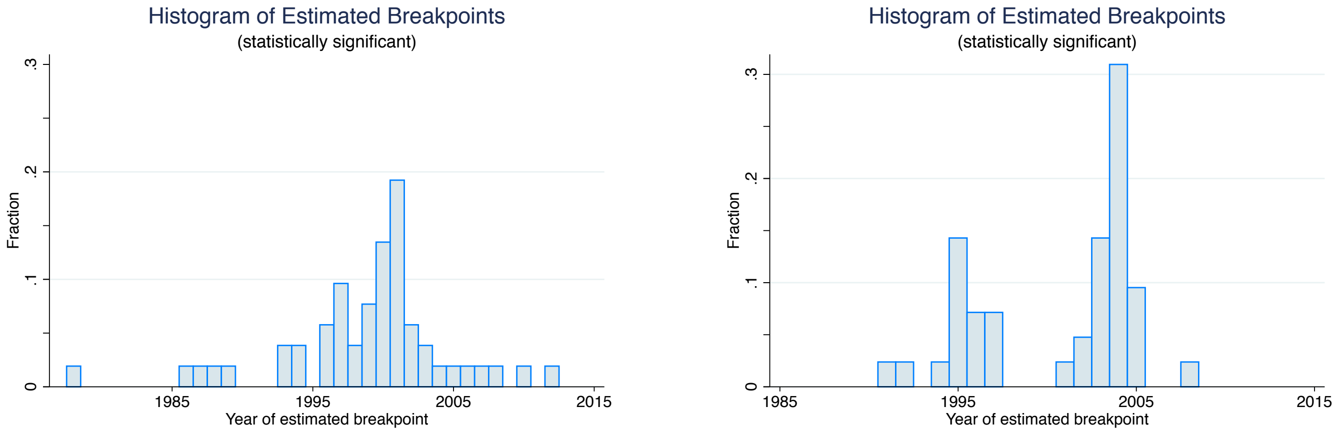


(a) Central Harlem, MN: positive breakpoint

(b) Canarsie, BK: negative breakpoint

Notes: The top figure of both panels shows the house price indexes with the smoothed local polynomials over which price growth trends are calculated; the price growth trends are in the bottom two figures. Vertical red lines are drawn at the year of the estimated breakpoints – both Central Harlem and Canarsie had structural breaks identified in 1997. Central Harlem (in Manhattan), has a positive structural breakpoint such that it is growing faster than the NYC average. Canarsie (in Brooklyn), has a negative structural breakpoint such that it is declining relative to the NYC average.

Figure 1.4: Distribution of the Timing of Statistically Significant Breakpoints



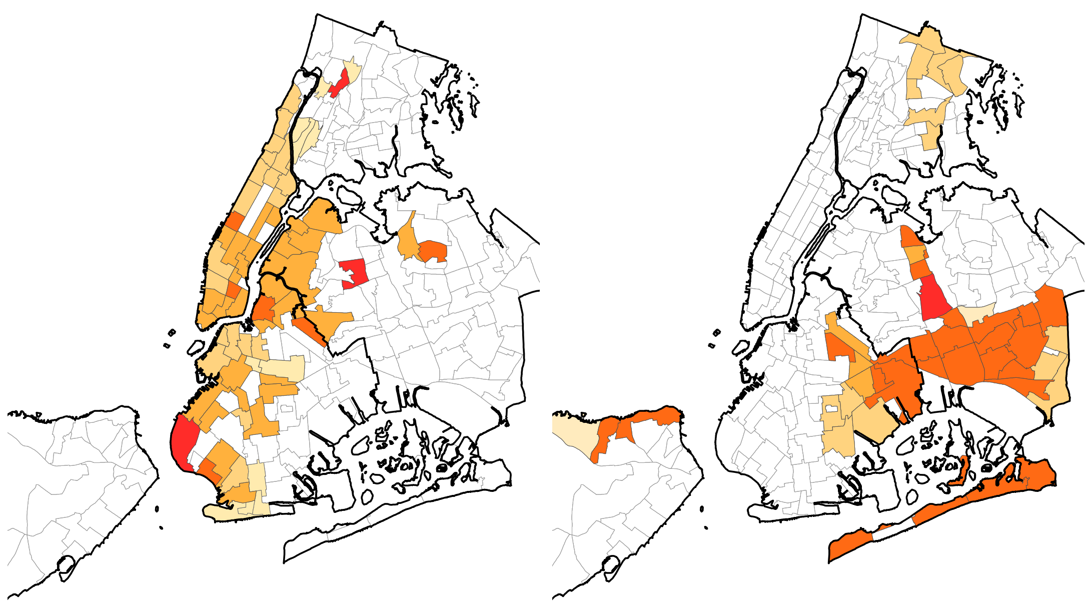
(a) Growing relative to NYC **(b)** Declining relative to NYC

Notes: Of the 157 neighbourhoods defined in this analysis, 94 had statistically significant breakpoints. The above histogram plots the distribution of the timing of the breakpoints. Breakpoints in appreciating neighbourhoods (figure a) were identified relatively earlier than in depreciating neighbourhoods (figure b).

Figure 1.5: Plot of the Timing of Statistically Significant Breakpoints

Legend for Fig. 1.5

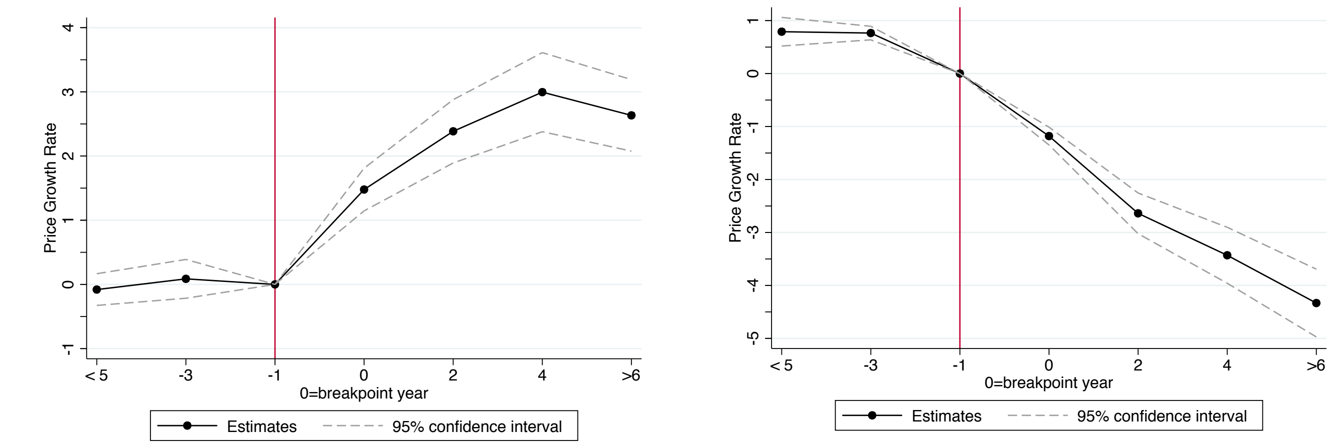
Year of the Estimated Structural Breakpoint
2007 - 2012
2002 - 2006
1997 - 2001
1992 - 1996
1978 - 1991



(a) Growing faster relative to NYC **(b)** Declining relative to NYC

Notes: The above figures depict the temporal pattern of breakpoints across NYC neighbourhoods. The occurrence of breakpoints is geographically clustered. The darker-shaded neighbourhoods have breakpoints later than the lighter-shaded neighbourhoods. Breakpoints in appreciating neighbourhoods (figure a) were identified relatively earlier and closer to the central business district (Midtown, MN) than in depreciating neighbourhoods (figure b).

Figure 1.6: Magnitude of Price Growth Rates Relative to Breakpoint

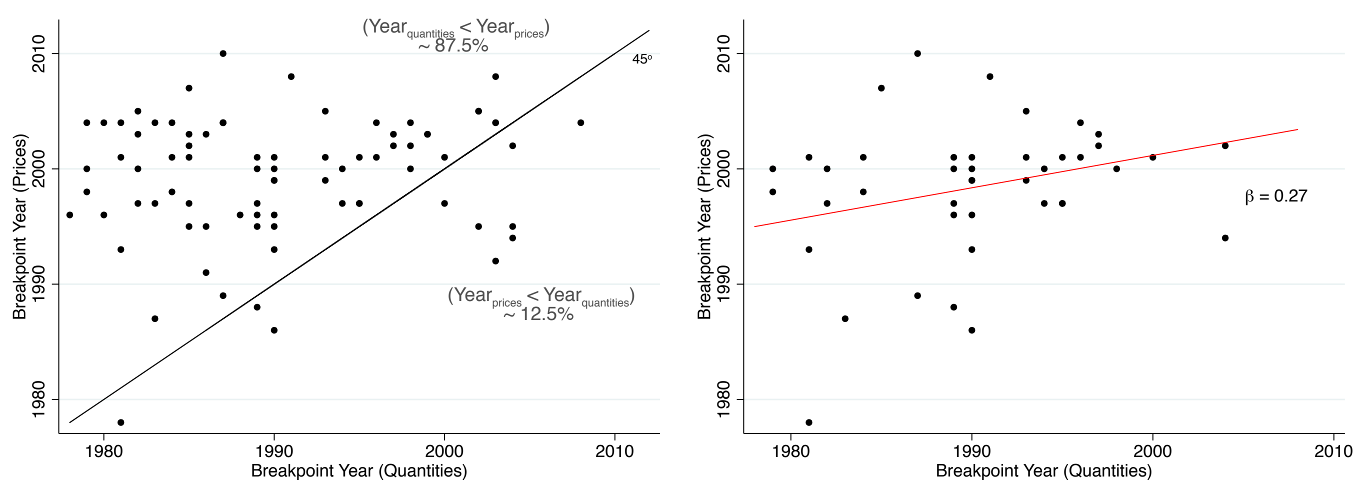


(a) Neighbourhoods growing relative to NYC average

(b) Neighbourhoods declining relative to NYC average

Notes: The above figures depict the magnitude of the changes in house prices at the breakpoint (year 0) relative to the year prior (year -1) for neighbourhoods growing faster than the NYC average (a) and neighbourhoods declining relative to the NYC average (b). The points represent the coefficients on a set of dummy regressors with the year prior to the breakpoints being the excluded category (i.e. $\beta = 0$ by construction). The dashed lines are 95% confidence intervals.

Figure 1.7: Relationship between Estimated Tipping Points in Prices and Quantities



(a) Lead/lags

(b) Regression

Notes: Figure (a) plots the years of the estimated breakpoints in house price growth rates (Y-axis) against the breakpoints in volume of sales (X-axis). The 45° line depicts the points where the two years are equal. The north-west half has the year of tipping on volume earlier than the year of tipping on price; the south-east half has the year of tipping on volume later than the year of tipping on price. Figure (b) plots the regression coefficient of the correlation between the estimated breakpoints over house price growth rates and volume of sales. The correlation is positive and is restricted to neighbourhoods that are growing relative to NYC. There is no statistical relationship among declining neighbourhoods.

Tables

Table 1.1: The Spatial Relationship between Estimated Tipping Points

			OLS		
Dependent Variable:		$\mathbb{1}$ *(when change)* $\in \{0,1\}$			
	(1)	(2)	(3)	(4)	(5)
		($d_i \geq 0$)	($d_i \leq 0$)	($d_i \geq 0$)	($d_i \leq 0$)
neighbour changed:	0.061	0.055	-0.003	0.029	
$x \in \{0,1\} \mid d_j > 0$	(0.03)**	(0.03)*	(0.01)	(0.03)	
neighbour changed:	0.137	-0.013	0.152		0.143
$x \in \{0,1\} \mid d_j < 0$	(0.02)***	(0.01)***	(0.03)***		(0.03)***
neighbour changed:				0.074	
$x \in \{0,1\} \mid d_j > 0, t-1$				(0.01)***	
neighbour changed:				0.053	
$x \in \{0,1\} \mid d_j > 0, t-2$				(0.02)***	
neighbour changed:				0.041	
$x \in \{0,1\} \mid d_j > 0, t-3$				(0.02)***	
neighbour changed:					0.033
$x \in \{0,1\} \mid d_j < 0, t-1$					(0.01)***
neighbour changed:					0.030
$x \in \{0,1\} \mid d_j < 0, t-2$					(0.02)
neighbour changed:					-0.002
$x \in \{0,1\} \mid d_j < 0, t-3$					(0.03)
R^2	0.08	0.04	0.16	0.07	0.17
N	2,335	1,886	1,852	1,886	1,852

Notes: *** statistically significant at the 1% level; standard errors in parenthesis and clustered at the neighborhood level. Regressors include a set of dummy variables equal to 1 if neighbourhood n has a neighbour ever tip up or down (columns (1)-(3)), or tip up or down in the previous few years (columns (4)-(5)). Column (1) includes the full set of neighbourhoods. Columns (2) and (4) restrict to neighbourhoods that are gaining in prices; columns (3) and (5) restrict to neighbourhoods declining in prices. All regressions include year fixed effects.

Table 1.2: Magnitude of Price Growth Relative to Estimated Breakpoints

	Dependent Variable: % Δ Prices				
	Before	At breakpoint		After	
	$(y^{**}_{n,t-3,t-4})$	$(y^{**}_{n,t,t+1})$	$(y^{**}_{n,t+2,t+3})$	$(y^{**}_{n,t+4,t+5})$	$(y^{**}_{n,t\geq 6})$
$PG_{y<y^{**}} < PG_{y>y^{**}}$	0.087	1.478	2.385	2.995	2.633
(*Gaining in prices*)	(0.15)	$(0.17)^{***}$	$(0.25)^{***}$	$(0.31)^{***}$	$(0.28)^{***}$
$PG_{y<y^{**}} > PG_{y>y^{**}}$	0.765	−1.180	−2.639	−3.431	−4.333
(*Declining in prices*)	$(0.06)^{***}$	$(0.08)^{***}$	$(0.19)^{***}$	$(0.26)^{***}$	$(0.32)^{***}$

Notes: *** statistically significant at the 1% level; standard errors in parenthesis and clustered at the neighbourhood level. The first row restricts to neighborhoods with appreciating prices such that the β coefficient on the breakpoint estimate is greater than zero. The second row restricts to neighborhoods with depreciating prices such that the β coefficient on the breakpoint estimate is less than zero. Price growth rates are regressed on a set of dummies relative to the breakpoint year with the year prior to the breakpoint being the excluded category $(t-1, t-2)$. All regressions control for borough fixed effects and coefficients are percentage point increases (or decreases) relative to the base year. The base year price growth in gentrifying neighbourhoods is 0.432%; the base year price growth in declining neighbourhoods is -0.789%.
Interpretation (row 1): at the breakpoint $PG = 1.478$ p.p. greater than at the base year
$\therefore 1.478 + 0.432 = PG$ in BP year $\implies$ prices are growing 4.4x faster than in the base year.

Table 1.3: Housing Demand Changes Relative to Estimated Breakpoints

	Dependent Variable: $\Delta ln(Quantities)$				
	Before	At breakpoint		After	
	$(y^{**}_{n,t-3,t-4})$	$(y^{**}_{n,t,t+1})$	$(y^{**}_{n,t+2,t+3})$	$(y^{**}_{n,t+4,t+5})$	$(y^{**}_{n,t\geq 6})$
$PG_{y<y^{**}} < PG_{y>y^{**}}$	0.970	1.849	1.832	1.211	−0.562
(*Gaining in prices*)	$(0.57)^{*}$	$(0.58)^{***}$	$(0.64)^{***}$	$(0.60)^{**}$	(0.43)
$PG_{y<y^{**}} > PG_{y>y^{**}}$	0.084	−0.039	−0.112	−0.436	−1.934
(*Declining in prices*)	(0.14)	(0.11)	(0.20)	$(0.26)^{*}$	$(0.45)^{***}$

Notes: *** statistically significant at the 1% level; standard errors in parenthesis and clustered at the neighbourhood level. The first row restricts to neighborhoods with appreciating prices such that the β coefficient on the breakpoint estimate is greater than zero. The second row restricts to neighborhoods with depreciating prices such that the β coefficient on the breakpoint estimate is less than zero. The changes in the quantity of sales ($\Delta ln(Quantities)$) are regressed on a set of dummies relative to the breakpoint year with the year prior to the breakpoint being the excluded category $(t-1,\ t-2)$. All regressions control for borough fixed effects and coefficients are percentage point increases (or decreases) relative to the base year.

Table 1.4: Demographic Changes Relative to Estimated Breakpoints

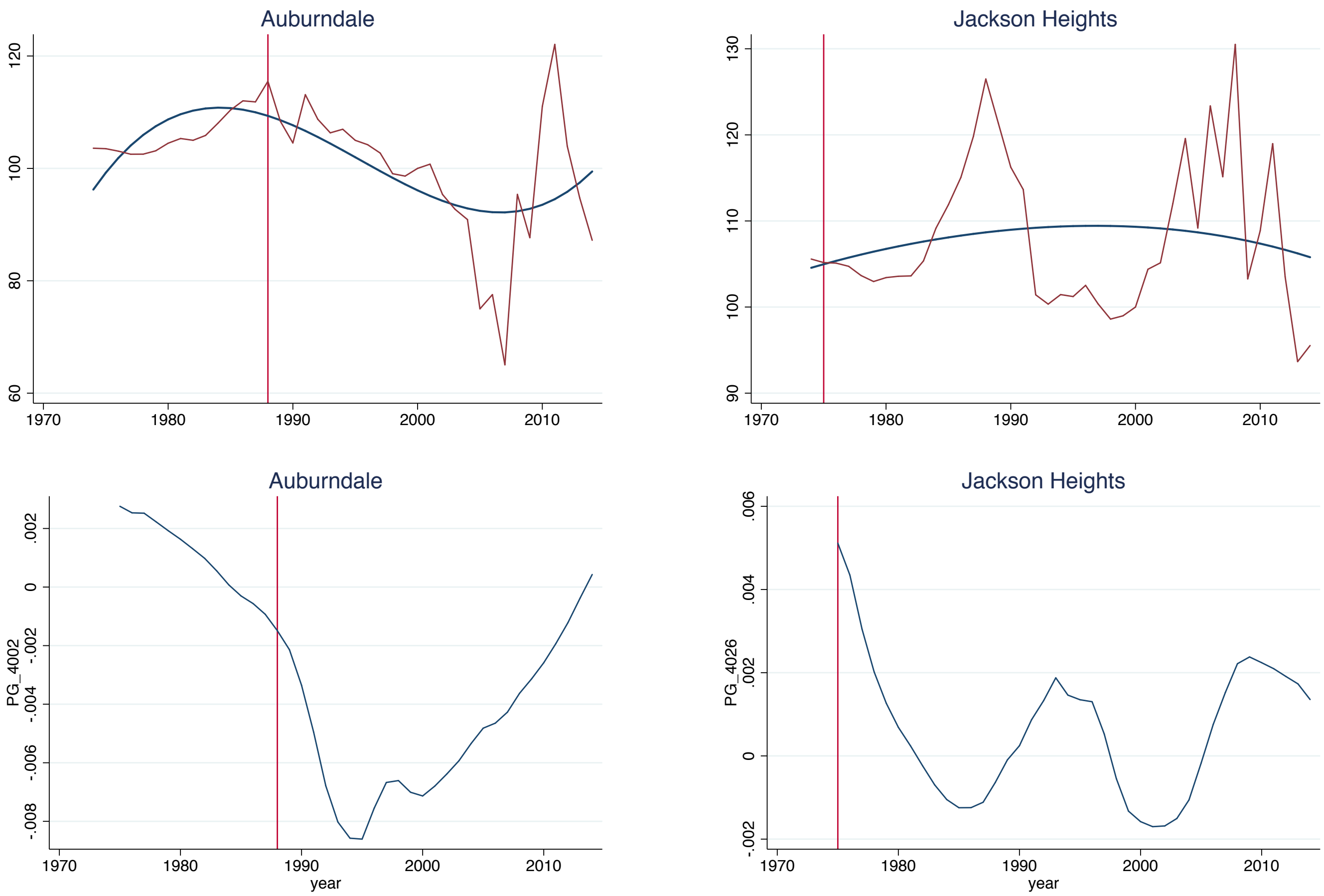
	Before ($y^{**}_{n,t-3,t-4}$)	At breakpoint ($y^{**}_{n,t,t+1}$)	($y^{**}_{n,t+2,t+3}$)	After ($y^{**}_{n,t+4,t+5}$)	($y^{**}_{n,t\geq 6}$)
Panel A:	Dependent Variable: % Δ Income				
$PG_{y<y^{**}} < PG_{y>y^{**}}$	-0.213	0.538	0.771	0.837	0.417
(*Gaining in prices*)	(0.26)	(0.29)*	(0.30)**	(0.33)**	(0.26)
$PG_{y<y^{**}} > PG_{y>y^{**}}$	-0.611	-0.431	-0.963	-1.280	-1.892
(*Declining in prices*)	(0.11)	(0.19)**	(0.35)***	(0.46)***	(0.44)***
Panel B:	Dependent Variable: % Δ Purchases by Black Households				
$PG_{y<y^{**}} < PG_{y>y^{**}}$	3.760	1.467	-0.627	-2.488	-7.332
(*Gaining in prices*)	(1.10)***	(1.24)	(1.07)	(1.03)**	(1.34)***
$PG_{y<y^{**}} > PG_{y>y^{**}}$	4.466	-3.461	-5.322	-6.563	-19.843
(*Declining in prices*)	(0.73)***	(0.52)***	(1.06)***	(1.51)***	(2.01)***
Panel C:	Dependent Variable: % Δ Purchases by White Households				
$PG_{y<y^{**}} < PG_{y>y^{**}}$	7.842	3.768	1.481	-0.464	-3.742
(*Gaining in prices*)	(1.32)***	(1.30)***	(1.22)	(1.31)	(1.42)**
$PG_{y<y^{**}} > PG_{y>y^{**}}$	6.577	-2.250	-2.979	-4.847	-15.891
(*Declining in prices*)	(1.33)***	(1.42)	(2.12)	(2.17)**	(2.37)***

Notes: *** statistically significant at the 1% level; standard errors in parenthesis and clustered at the neighbourhood level. The first row restricts to neighborhoods with appreciating prices such that the β coefficient on the breakpoint estimate is greater than zero. The second row restricts to neighborhoods with depreciating prices such that the β coefficient on the breakpoint estimate is less than zero. In Panel A income growth rates are regressed on a set of dummies relative to the breakpoint year with the year prior to the breakpoint being the excluded category ($t-1$, $t-2$). In Panel B the regressor is changes in the quantity of purchases by black households; in Panel C it is changes in the quantity of purchases by white households. All regressions control for borough fixed effects and coefficients are percentage point increases (or decreases) relative to the base year.

Appendix 1.A

Figures

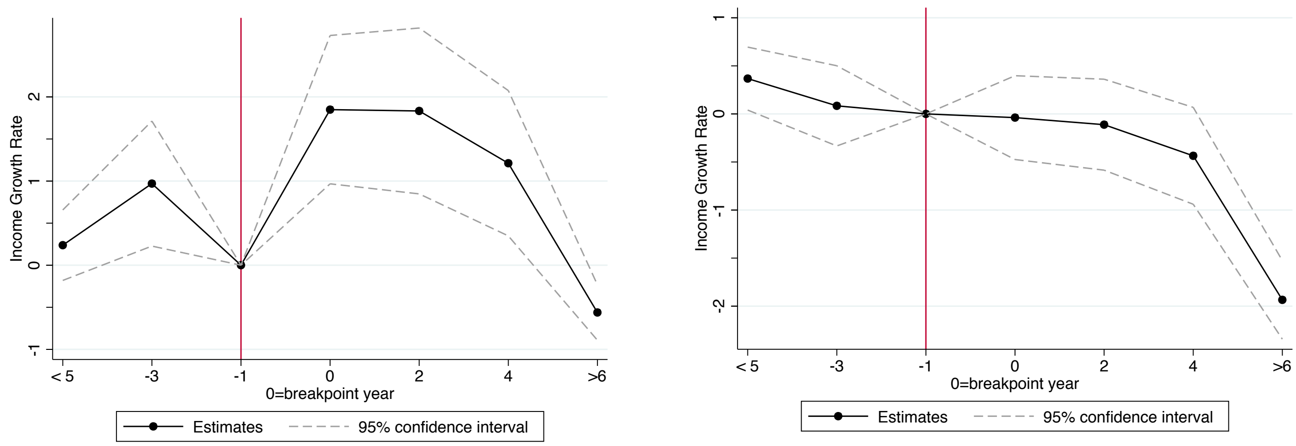
Figure 1.A1: Neighbourhoods with No Estimated Breakpoints (examples)



(a) Auburndale, QN **(b)** Jackson Heights, QN

Notes: An example of neighbourhoods with no statistically significant breakpoint. The top figure of both panels shows the house price indexes with the smoothed local polynomials over which price growth trends are calculated; the price growth trends are in the bottom two figures. Vertical red lines are drawn at the year of the insignificant breakpoint, which has no predictive power in estimating the future price growth rate series.

Figure 1.A2: Housing Demand Changes Relative to Estimated Breakpoints

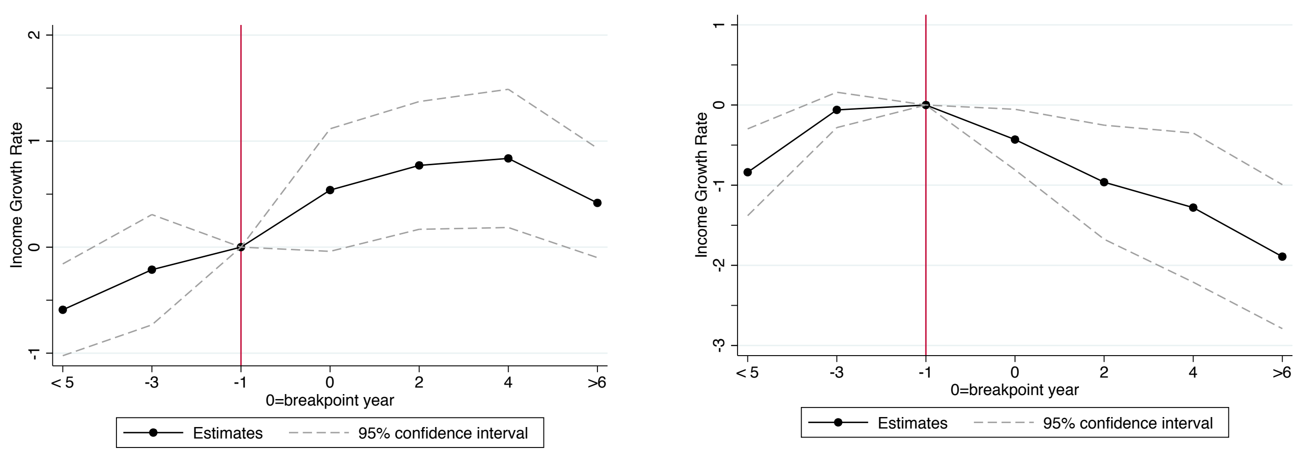


(a) Neighbourhoods growing relative to NYC average

(b) Neighbourhoods declining relative to NYC average

Notes: The above figures depict the magnitude of the changes in sales quantities at the breakpoint (year 1) relative to the year prior (year 0) for neighbourhoods growing faster than the NYC average (a) and neighbourhoods declining relative to the NYC average (b). The points represent the coefficients on a set of dummy regressors with the year prior to the breakpoints being the excluded category (i.e. $\beta = 0$ by construction). The dashed lines are 95% confidence intervals.

Figure 1.A3: Income Changes Relative to Estimated Breakpoints

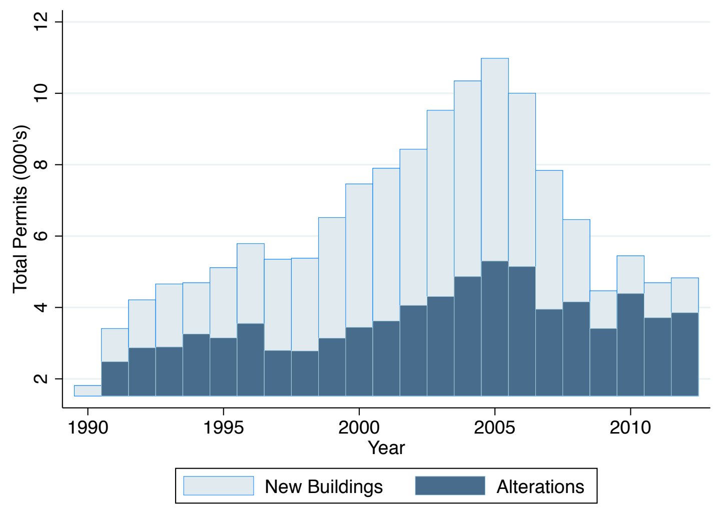


(a) Neighbourhoods growing relative to NYC average

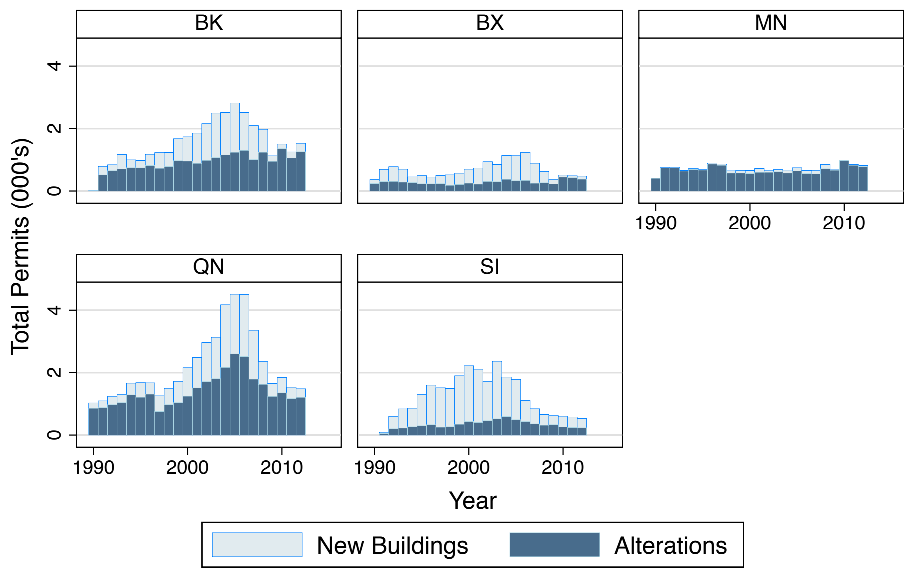
(b) Neighbourhoods declining relative to NYC average

Notes: The above figures depict the magnitude of the changes in sales quantities at the breakpoint (year 1) relative to the year prior (year 0) for neighbourhoods growing faster than the NYC average (a), and neighbourhoods declining relative to the NYC average (b). The points represent the coefficients on a set of dummy regressors with the year prior to the breakpoints being the excluded category (i.e. $\beta = 0$ by construction). The dashed lines are 95% confidence intervals.

Figure 1.A4: Annual Issuance of Building Permits



(a) New York City



(b) By New York City Borough

Notes: Figure (a) plots the total number of new building permits issued from 1990 to 2012 in New York City. This includes permits for new buildings plus permits for significant alternations. Until 2008 the number of new building permits was greater than the number of permits for alternations. Figure (b) plots the total number of permits issued within each borough. In Manhattan, very few permits have been issued for new development whereas in the Bronx, very few permits have been issues for alterations. This is indicative of the supply elasticity within each borough. In all cases, total numbers are per thousands.

Tables

Table 1.A1: Neighbourhoods in New York City

Manhattan N = 21			
Neighbourhood name	*census tracts*	*Neighbourhood name*	*census tracts*
Battery Park City-Lower Manhattan	7	Midtown-Midtown South	16
Central Harlem	24	Morningside Heights	16
Clinton	7	Murray Hill-Kips Bay	6
East Harlem	19	SoHo-TriBeCa-Civic Center-Little Italy	10
East Village	7	Turtle Bay-East Midtown	10
Gramercy	4	Upper East Side-Carnegie Hill	14
Hudson Yards-Chelsea-Flatiron-Union Square	14	Upper West Side	16
Lenox Hill-Roosevelt Island	9	Washington Heights	23
Lincoln Square	8	West Village	14
Lower East Side (includes Chinatown)	16	Yorkville	9
Marble Hill-Inwood	7		
Excluded from analysis:			
Stuyvesant Town - Cooper Village (private residential development)			

Brooklyn N = 41			
Neighbourhood name	*census tracts*	*Neighbourhood name*	*census tracts*
Bath Beach	11	Flatbush	26
Bay Ridge	28	Flatlands	26
Bedford	17	Georgetown-Mill Basin	19
Bensonhurst East	18	Gravesend	8
Bensonhurst West	26	Greenpoint	12
Borough Park	28	Homecrest	16
Brighton Beach (includes Coney Island)	18	Kensington-Ocean Parkway	8
Bushwick North	13	Madison	15
Bushwick South	20	Midwood	18
Canarsie	32	North Side-South Side	13
Carroll Gardens-Columbia Street-Red Hook	11	Ocean Hill (includes Brownsville)	24
Clinton Hill (includes Fort Greene)	18	Ocean Parkway South	9
Crown Heights North (includes Prospect Heights)	33	Park Slope-Gowanus	20
Crown Heights South (includes Wingate)	26	Rugby-Remsen Village	17
Cypress Hills-City Line	16	Sheepshead Bay-Manhattan Beach	18
Downtown Brooklyn (includes Brooklyn Heights)	17	Stuyvesant Heights	17
Dyker Heights	17	Sunset Park East	15
East Flatbush-Farragut	19	Sunset Park West	14
East New York	22	Williamsburg (includes East)	17
East New York (Pennsylvania Ave)	9	Windsor Terrace	6
Erasmus	7		
Excluded from analysis:			
Starrett City (housing development)			

Bronx N = 23			
Neighbourhood name	*census tracts*	*Neighbourhood name*	*census tracts*
Allerton-Pelham Gardens	11	Parkchester	5
Bedford Park-Fordham North	11	Pelham Bay-Country Club-City Island	6
Belmont-East Tremont	17	Pelham Parkway	9
Bronxdale	8	Schuylerville-Edgewater Park	15
Claremont-Crotona-Melrose	21	Soundview-Bruckner	9
Concourse-Highbridge	27	Soundview-Harding Park	12
Eastchester-Baychester (includes Co-op city)	10	Van Nest-Morris Park	12
Fordham-University Heights (includes Mount Hope)	30	West Farms-Bronx River	7
Hunts Point-Longwood	13	Westchester-Unionport	7
Kingsbridge Heights-Norwood	12	Williamsbridge-Olinville	20

Continued on next page...

... table continued

Melrose-Mott Haven	19	Woodlawn-Wakefield	14
North Riverdale-Kingsbridge (includes Van Cortlandt)	27		
Excluded from analysis:			
Rikers Island (jail complex)			

Queens N = 54

Neighbourhood name	*census tracts*	*Neighbourhood name*	*census tracts*
Astoria (includes Long Island City)	30	Jamaica Estates-Holliswood	8
Auburndale	7	Kew Gardens Hills	9
Baisley Park	16	Laurelton	11
Bayside-Bayside Hills	18	Lindenwood-Howard Beach	4
Bellerose	6	Maspeth	14
Breezy Point-Broad Channel	7	Middle Village	16
Briarwood-Jamaica Hills	11	Murray Hill	13
Cambria Heights	13	North Corona	9
College Point	6	Oakland Gardens	6
Corona	10	Old Astoria	8
Douglas Manor-Douglaston-Little Neck	6	Ozone Park	6
East Elmhurst	8	Pomonok-Flushing Heights-Hillcrest	7
East Flushing	12	Queens Village	20
Elmhurst	17	Queensboro Hill	5
Elmhurst-Maspeth	6	Rego Park	8
Far Rockaway-Bayswater	8	Richmond Hill	32
Flushing	13	Ridgewood	18
Forest Hills	24	Rosedale	6
Fresh Meadows-Utopia	4	South Jamaica	14
Ft. Totten-Bay Terrace-Clearview	5	South Ozone Park	25
Glen Oaks-Floral Park-New Hyde Park	4	Springfield Gardens North	3
Glendale	13	Springfield Gardens South-Brookville	5
Hammels-Arverne-Edgemere	7	St. Albans	22
Hollis	9	Steinway	16
Hunters Point-Sunnyside-West Maspeth	18	Whitestone	8
Jackson Heights	19	Woodhaven	20
Jamaica	14	Woodside	12
Excluded from analysis:			
Airports (both JFK and LGA)			

Staten Island N = 18

Neighbourhood name	*census tracts*	*Neighbourhood name*	*census tracts*
Annadale-Huguenot-Prince's Bay-Eltingville	6	New Springville-Bloomfield-Travis	9
Arden Heights	4	Oakwood-Oakwood Beach	5
Charleston-Richmond Valley-Tottenville	4	Old Town-Dongan Hills-South Beach	5
Grasmere-Arrochar-Ft. Wadsworth	4	Port Richmond	4
Great Kills	8	Rossville-Woodrow	3
Grymes Hill-Clifton-Fox Hills	4	Stapleton-Rosebank	8
Mariner's Harbor-Arlington	8	Todt Hill-Heartland Village	7
New Brighton-Silver Lake	5	New Brighton-St. George	11
New Dorp-Midland Beach	5	Westerleigh	7

Table 1.A2: Descriptive Statistics

Panel A – New York City Sales Data: House Prices

	Mean	Sd (2014 $)	Median		Mean	Sd (2014 $)	Median
NEW YORK CITY				Manhattan			
1974	$148,321	$106,454	$136,195	1974	$167,334	$307,655	$164,854
2014	$709,941	$1,313,377	$395,000	2014	$1,944,913	$2,579,746	$1,195,000
Total Observations	*1,261,067*			*159,646*			
Brooklyn				Bronx			
1974	120,746	90,386	104,578	1974	118,842	74,607	107,011
2014	510,393	534,754	375,000	2014	222,619	184,925	183,667
Total Observations	*365,477*			*113,157*			
Queens				Staten Island			
1974	169,729	93,892	160,516	1974	187,515	94,228	187,268
2014	417,129	265,940	363,500	2014	356,802	193,142	335,000
Total Observations	*462,801*			*159,986*			

Panel B – Mortgage Application Data: Incomes

	Mean	Sd (2014 $)	Median		Mean	Sd (2014 $)	Median
NEW YORK CITY				Manhattan			
1990	$168,940	$295,380	$117,730	1990	$288,720	$508,230	$172,070
2014	$271,330	$482,830	$145,000	2014	$486,650	$733,600	$265,000
Total Observations	*557,260*			*152,982*			
Brooklyn				Bronx			
1990	128,810	120,790	106,870	1990	121,920	107,760	101,430
2014	244,060	321,220	162,000	2014	113,158	90,800	86,500
Total Observations	*121,322*			*44,041*			
Queens				Staten Island			
1990	132,890	194,780	108,680	1990	139,170	192,120	113,110
2014	123,720	103,090	99,000	2014	143,580	150,220	123,000
Total Observations	*180,791*			*58,124*			

Panel C – Mortgage Application Data: Race

	Black (total number)	White		Black (total number)	White
NEW YORK CITY			Manhattan		
1990	2,443	10,975	1990	107	3,601
2014	505	6,978	2014	42	2,478
Brooklyn			Bronx		
1990	1,128	2,319	1990	447	620
2014	168	1,558	2014	130	440
Queens			Staten Island		
1990	715	3,021	1990	46	1,414
2014	146	1,751	2014	19	751

Notes: The above table presents descriptive statistics for the New York City sales data (Panel A) and the Housing and Mortgage data (Panels B and C). All dollar values are in real 2014 $. The sales data ranges from 1974 to 2014, while the mortgage data ranges from 1990 to 2014.

Table 1.A3: Price Changes around Estimated Breakpoints (for all neighbourhoods)

Neighbourhoods gaining in prices (PG higher after estimated breakpoint):					
			$\%\Delta HPI$		breakpoint
borough	neighborhood	1974 to 2014	1990 to 2014	at breakpoint	year
MN	Central Harlem	221.68	324.31	2.87	1997
MN	Washington Heights	203.06	244.29	2.39	1997
BK	Windsor Terrace	184.91	206.10	2.30	1998
BK	Park Slope-Gowanus	180.23	200.83	3.14	1999
MN	Morningside Heights	179.79	201.27	2.15	1996
BK	Greenpoint	167.10	177.38	2.73	2001
BK	Williamsburg	158.27	142.55	1.82	1999
BK	Carroll Gardens-Red Hook	147.48	166.44	2.00	1996
BK	Clinton Hill	138.80	159.01	1.96	1997
BK	Downtown BK	127.96	152.54	1.82	1996
MN	East Harlem	122.76	155.81	2.86	1997
MN	West Village	119.05	209.21	1.51	1993
MN	Lower East Side	104.43	85.21	2.39	2001
MN	Hudson Yards-Chelsea-Flatiron	98.86	144.32	1.13	1994
BK	Sunset Park West	98.44	88.72	1.79	2000
MN	SoHo-TriBeCa-Little Italy	96.49	112.28	2.01	2002
BK	Sunset Park East	82.77	72.91	1.91	2001
QN	Hunters Point-West Maspeth	75.18	50.47	1.25	2000
MN	Upper West Side	66.21	92.99	0.6	1994
BK	North Side-South Side	64.26	114.35	1.23	2005
MN	East Village	58.24	98.89	2.08	2004
BK	Bushwick North	56.86	69.61	1.12	2003
MN	Midtown-Midtown South	55.25	53.18	0.99	2001
MN	Gramercy	51.95	79.35	1.51	2002
BK	Crown Heights North	50.62	70.31	0.42	1989
QN	Steinway	46.63	22.94	0.84	2001
QN	Ridgewood	44.41	32.57	0.62	2000
QN	Old Astoria	43.44	19.35	0.68	1999
MN	Murray Hill-Kips Bay	42.05	53.35	0.85	2001
QN	Astoria	41.26	20.38	0.61	1999
MN	Clinton	40.45	51.37	0.84	2001
MN	Upper East Side-Carnegie Hill	35.23	48.24	1.01	2002
QN	Flushing	29.73	15.29	0.73	2001
BK	Crown Heights South	29.64	31.17	0.69	2001
BK	Bensonhurst West	28.41	20.78	0.34	2001
MN	Yorkville	23.73	38.05	0.33	2000
BK	Flatbush	23.1	23.61	0.49	2000
MN	Lincoln Square	23.1	69.81	0.78	2003
MN	Turtle Bay-East Midtown	20.17	20.33	0.4	2000
BX	Kingsbridge Heights-Norwood	18.03	14.63	0	1986
BK	Gravesend	17.22	11.04	0.25	2000

Continued on next page...

... table continued

	Neighbourhoods declining in prices (PG lower after estimated breakpoint):				
			$\%\Delta HPI$		breakpoint
borough	neighborhood	1974 to 2014	1990 to 2014	at breakpoint	year
QN	South Jamaica	-105.68	-106.29	-1.79	2003
QN	Baisley Park	-100.39	-100.43	-2.12	2003
BK	East New York	-100.13	-100.14	-2.56	2004
QN	Hammels-Arverne-Edgemere	-99.59	-99.56	-2.73	2004
BK	East New York (Pennsylvania Ave)	-98.74	-98.44	-1.81	2003
BK	Ocean Hill	-92.64	-90.21	-1.48	2002
BK	Canarsie	-91.94	-92.27	-1.88	1997
BX	Williamsbridge-Olinville	-88.31	-88.35	-1.50	1997
QN	Springfield Gardens North	-86.88	-86.64	-1.79	2004
BX	Eastchester-Edenwald-Baychester	-84.42	-84.97	-1.72	1997
QN	St. Albans	-80.95	-80.25	-2.27	2004
QN	Far Rockaway-Bayswater	-79.51	-80.19	-0.56	2003
BX	Van Nest-Morris Park	-77.76	-78.41	-0.62	1995
BX	Westchester-Unionport	-77.27	-77.62	-1.63	1996
BK	East Flatbush-Farragut	-76.08	-77.71	-0.94	1995
SI	New Brighton	-75.56	-76.77	-1.87	2005
QN	Cambria Heights	-74.86	-74.72	-1.15	1996
BX	Woodlawn-Wakefield	-73.47	-74.28	-1.48	1996
BK	Rugby-Remsen Village	-71.05	-72.90	-0.64	1995
BK	Cypress Hills-City Line	-69.52	-71.87	-1.26	2004
QN	Jamaica	-69.45	-70.71	-1.74	2004
QN	Rosedale	-69.4	-70.24	-0.84	1995
QN	Springfield Gardens South-Brookville	-67.7	-67.73	-1.91	2004
SI	Mariner's Harbor	-67.5	-70.58	-0.74	1992
QN	Hollis	-66.23	-65.7	-1.57	2004
SI	Port Richmond	-63.78	-65.93	-2.55	2003
QN	South Ozone Park	-60.48	-60.35	-1.28	2004
QN	Breezy Point-Rockaway Park-	-59.59	-61.02	-0.76	2005
BK	Flatlands	-59.31	-61.09	-0.96	1995
QN	Queens Village	-56.07	-55.63	-1.01	2004
QN	Laurelton	-53.51	-55.63	-0.49	1995
BX	Bronxdale	-52.33	-58.37	-0.31	1994
QN	Woodhaven	-51.46	-54.2	-1.21	2005
BK	Bushwick South	-49.63	-37.93	-0.22	2001
QN	Richmond Hill	-49.37	-49.93	-1.11	2004
QN	Briarwood-Jamaica Hills	-49.18	-52.82	0.09	1991
QN	Ozone Park	-47.26	-50.15	-1.21	2005
QN	East Elmhurst	-41.83	-48.83	0.01	2003
QN	Corona	-40.11	-48.5	-0.64	2004
BK	Stuyvesant Heights	-28.51	-8.32	-0.18	2004
QN	North Corona	-18.3	-29.04	0.09	2002
QN	Forest Hills	-5.65	-8.54	0	2008

Appendix 1.B

Constructing the house price index

To construct the repeat sales index, I use the procedure suggested by Case and Shiller (1989) and modified by Quigley and Van Order (1995). This is also the procedure used by the Furman Center in their annual report, *State of New York City's Housing and Neighbourhoods* (NYU Furman Center). The repeat sales index is created using a weighted regression, constructed in three stages.

In the first stage, I regress the change in price between sales on a set of dummy variables for each year in the sample (Equation (1) below). The dummy variables equal +1 for the year of the second sale, -1 for the year of the first sale, and zero otherwise. In all cases, 2000 is used as the base year.

Taking the predicted errors from the first step, the second stage regresses the difference between the predicted sales price in the first stage and the actual sale price on both the time interval between sales and the squared time interval between sales (Equation (2) below). This constructs the variance matrix used to weight the regressors in Step 3.

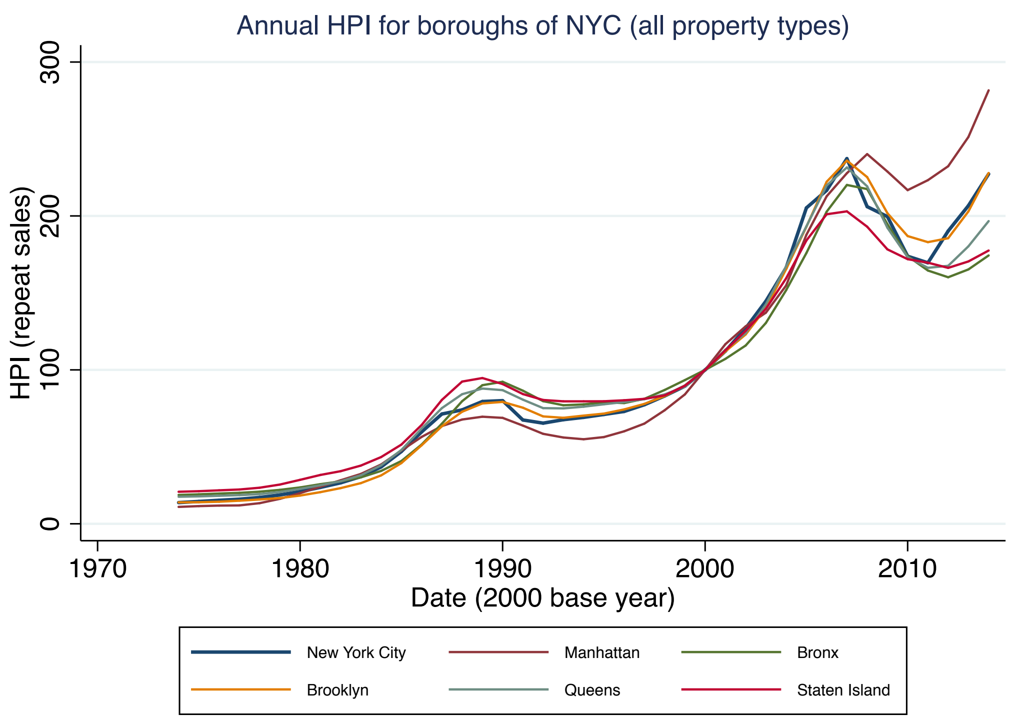
The third stage then re-estimates Equation (1), weighting each observation by the inverse of the square root of the variance predicted in the second stage (Equation (3) below). In doing so, the weighted sum of squared residuals is minimized. This mechanically puts more weight on properties that sold more frequently and gives lower weight to those properties that have longer lags between sales.

Step 1:
$$\begin{aligned} log(price_t) - log(price_{t-1}) &= \mathbb{1}[year = 1974, year = 2014] \\ &\rightarrow \text{predict}, \quad \hat{e} \end{aligned} \tag{1.1}$$

Step 2:
$$\begin{aligned} \hat{e} &= \beta_1(\text{elapsed years}) + \beta_2(\text{elapsed years})^2 \\ &\rightarrow \text{predict}, \quad \hat{\mu} \\ &\rightarrow weight = \hat{\mu}^2 \end{aligned} \tag{1.2}$$

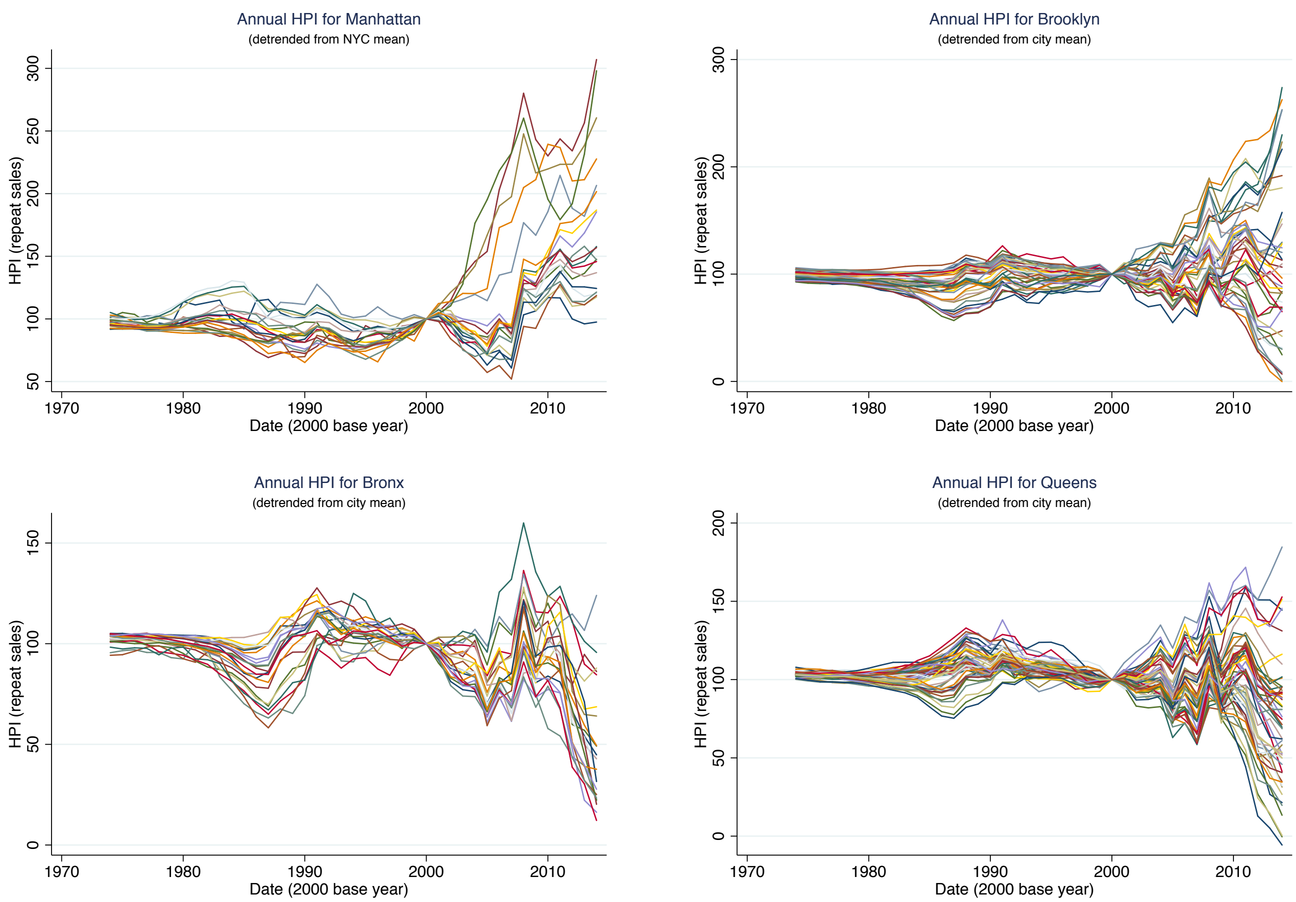
Step 3:
$$\begin{aligned} log(price_t) - log(price_{t-1}) &= \mathbb{1}[year = 1974, year = 2015], \\ aweight &= 1/weight \end{aligned} \tag{1.3}$$

Figure 1.B1: Annual House Price Index by NYC Boroughs



Notes: Annual repeat sales house price index for each borough in New York City. Each HPI is set to 2000 base year.

Figure 1.B2: Detrended Annual House Price Index by NYC Neighbourhoods



Notes: Annual house price index for each defined neighbourhood within Manhattan, Brooklyn, the Bronx, and Queens. Staten Island excluded for brevity. Each HPI is set to 2000 base year and reflects how quickly a neighbourhood is growing relative to the NYC average.

Chapter 2

Labour Demand, Housing Markets, and Neighbourhood Change

2.1 Introduction

Gentrification — the influx of higher-income residents into lower-income neighbourhoods — has received a lot of attention both politically and academically in the last decade. In the last 20 years downtowns around the United States have seen large increases in house prices and their shares of higher-income, better educated residents compared to their suburban counterparts. Focusing on changes within New York City, the goal of this chapter is to estimate the extent to which changing labour demand characteristics alter the patterns of neighbourhood change.

A number of recent papers address the trends in re-urbanization across major cities in the United States.[1] In doing so they consider the role of factors such as changes in amenities and household preferences (Baum-Snow and Hartley (2016), Couture and Handbury (2016)), tolerance for commuting (Edlund et al. (2015)), the age of the housing stock (Rosenthal (2008), Brueckner and Rosenthal (2009)), crime (Ellen et al. (2016)), and consumption externalities (Bayer et al. (2007), Guerrieri et al. (2013)). However, much of the literature looks at across-city gentrification rather than within-city neighbourhood change. While Couture and Handbury (2016) and Guerrieri et al. (2013) do consider within-city gentrification, the former specifically focuses on revival of central business districts (CBDs) as opposed to within-city changes more generally; the latter does not address the role that the CBD or access to labour markets has for neighbourhood change.

Another stream of research addresses the changing patterns in the distribution of incomes across cities. For example, Lee and Lin (2017) emphasize the role that geographic amenities play in anchoring neighbourhood incomes. Moretti (2013) shows that shifts in the demand for college-educated workers cause changes in the geographical location of skill groups. Both Baum-Snow and Hartley (2016) and Edlund et al. (2015) note the shift in jobs located in downtowns toward those employing higher-skilled

[1] For a comprehensive review of the empirical and theoretical literature on neighbourhood change, see Rosenthal and Ross (2015).

labour. Beaudry et al. (2010) argue that technological change has increased the relative productivity of skilled workers. As is observed and discussed in Arzaghi and Henderson (2008) professional services firms are drawn to high-wage locations like downtown New York because there is significant value in the information spillovers between firms. The idea of skill-biased technological change implies that as industries employing skilled workers experience positive productivity shocks, the wages and relative supply of skilled workers will increase. Due to the value generated from firm level spillovers, these wage and labour supply increases should be observed in central neighbourhoods. This chapter differs from the previous literature by focusing on the role of the changing nature of labour demand in driving neighbourhood change within a city. Characterizing these within-city dynamics is an important component of understanding the implications of gentrification.

In the context of NYC, I estimate the extent to which these changes are driven by the changing composition and spatial organization of labour demand. I construct an exogenous measure of neighbourhood labour demand which is used to generate predicted shocks to neighbourhood incomes. Following Bartik (1991), each neighbourhood's base year industry mix is interacted with national changes in industry employment.[2] This is then further interacted with a parameter for the distance between neighbourhoods to generate a spatial-Bartik instrument.[3] The intuition underlying this instrument is that the national growth in industries differentially affects neighbourhoods because of pre-existing differences in composition and as such is unrelated to potential neighbourhood labour supply shocks. The spatial component implies that labour demand shocks in one neighbourhood generate housing demand shocks in other neighbourhoods. The size of this housing demand shock depends on the distance between neighbourhoods.

I find that labour markets have a significant impact on neighbourhood incomes and house prices. Specifically, a one standard deviation increase in exogenous labour demand increases homeowner incomes by 0.28 standard deviations and house prices by 0.54 standard deviations. This suggests that exogenous variation in labour demand can explain about 21% of the observed variation in income growth rates and 41% of the observed variation in house prices in NYC between 1990 and 2010. Estimating the model with actual changes to employment suggests that endogenous changes to labour supply can explain an additional 20% and 30% of the variation in incomes and house prices, respectively.

To address concern that New York City disproportionately employs high-skilled workers in finance and real estate and therefore is not representative of the true relationship in other cities, I estimate the first stage relationship between labour demand shocks and income growth in two other cities: Detroit, Michigan and Portland, Oregon – Detroit is representative of a city that declined a lot nationally and which disproportionately employed workers in manufacturing and Portland is representative of a city that is gentrifying, but which does not have a dominant industry of employment. In both cases, I find that exogenous labour demand shocks can explain approximately one fifth of the variation in observed changes to homeowner incomes; this is consistent with the observed effect in New York City.

Finally I consider several counterfactual labour market environments. First, employment in New York

[2] Bartik instruments have been widely used in the urban literature to isolate demand for living in a CBSA. Each of the aforementioned papers, Guerrieri et al. (2013), Baum-Snow and Hartley (2016), Edlund et al. (2015), and Couture and Handbury (2016) make use of a variant of the Bartik instrument.

[3] This spatial-Bartik instrument works as follows: national growth in industry x generates predicted labour demand in neighbourhood n, which generates predicted housing demand in neighbourhoods, $j \neq n$.

City grew approximately 7% slower than the rest of the United States between 1990 and 2010. Setting employment growth across industries equal to the national rate, overall house price growth is predicted to be 18% higher and the spatial variation 12% higher. Second, if employment across all industries had grown at the same rate, overall house price growth is predicted to be almost 90% higher and the spatial variation almost 113% higher. In both counterfactual scenarios gentrification is less concentrated in upper Manhattan but rather, appears in central neighbourhoods in Brooklyn and Queens – neighbourhoods that initially had employment in industries that have largely left NYC. Finally, if labour demand had remained unchanged from its 1990 level house price growth is predicted to be 60% lower overall. This counterfactual analysis reinforces the importance of changes in labour markets on the evolution of neighbourhood change.

The remainder of this chapter proceeds as follows: Section 2.2 presents the data used to characterize neighbourhood change. Section 2.3 is dedicated to understanding the theoretical relationship between labour demand and housing demand while Section 2.4 presents the empirical strategy and the construction of the spatial-Bartik instrument. Section 2.5 discusses the main empirical results; Section 2.6 goes through the extension to Detroit and Portland; and Section 2.7 considers the counterfactual labour markets. Lastly, Section 2.8 concludes.

2.2 Data and Descriptive Statistics

The data used in this chapter come from a variety of sources: (1) New York City property sales transactions data provide my measure of house price growth rates; (2) I use the FFIEC Home Mortgage Disclosure Act (HMDA) data for a measure of annual average homeowner incomes; (3) Zip code business patterns data provide annual by-industry employment counts across zip codes; and (4) the 1990 Census Transportation Planning Package is used to calculate worker flows between census tracts. Section 1.2 of Chapter 1 discusses the New York City Sales Data and the HMDA income data; refer to Appendix B of Chapter 1 for a detailed discussion of the construction of the repeat sales house price index.

Zipcode Business Patterns Data (ZCBP): Zipcode Business Patterns data are available through the Census Bureau for the years 1994 to 2013. County Business Patterns are available starting in 1986. The US Department of Housing and Urban Development (HUD) has created crosswalk files that overlay census tract geographies with zipcodes. Using their overlay and 2010 tract boundaries, I am able to allocate the zipcode data to census tracts. The ZCBP include by-industry employment counts.[4] Appendix Table 2.A1 provides descriptive statistics of logged employment changes in New York City. Between 1990 and 2013 total employment in New York City has increased by approximately 17 percent; the rest of the United States increased by 24 percent. Compositionally, employment has moved away from industries like Manufacturing towards Business and Professional Services more rapidly than elsewhere in the United States.

[4]Prior to 1997, industries were classified based on the North American Classification System (NAICS), after which they were reclassified under the Standard Industrial Classification (SIC) system. Industries are aggregated according to their two-digit industrial classifications and include, Agriculture, Mining, Utilities, Construction, Manufacturing of non-durable and durable goods, Transportation, Wholesale Trade, Retail Trade, Finance and Real Estate, Business Services, Professional Services, Personal Services, Health Care, Education, and Arts and Entertainment.

Census Transportation Planning Packages (CTPP): The CTPP data are used primarily to calculate commuting preferences but also supplements the ZCBP data described above. The 1990 CTPP is available through the Bureau of Transportation Statistics and provides information on location of work, location of residence, and commute times between the two. It also includes industry of work based on work location.[5] In order to isolate demand shocks for living in particular neighbourhoods based on the spatial distribution of jobs and employment growth within industries, I am going to rely on commute patterns around the city. I calculate inter-neighbourhood flows – resulting in 21,620 cross tabulations of observed commutes between neighbourhoods. During peak travel times the number of commuters between neighbourhoods averages 107 with a variance of 326 and the average commute time takes approximately 40 minutes.[6]

From this I estimate the elasticity of commuters between neighbourhoods i and j, with respect to travel time as follows:

$$ln(flows_{i,j}) = \kappa(\text{commute time}_{i,j}) + \eta_i + \eta_j \tag{2.1}$$

κ is the coefficient on commute time and is estimated to be -0.026, implying that for a 10 minute increase in commute time, worker flows will decrease by approximately 26%. Appendix Figure 2.A2 depicts this relationship between worker flows and travel time.

2.3 Labour Demand and Income Growth

2.3.1 Preferences and Changes to Housing Demand

There are myriad forces driving neighbourhood change. The remainder of this chapter focuses on the impact of one of these forces – changes to labour demand. For the most part, the geographic concentration of labour is located in Manhattan around the CBD (i.e. Midtown in NYC). However, both the volume of employment and the composition of employment has changed. For example, in the CBD in 1990 the fraction of employment in manufacturing was 7.5% and professional services was 10.8%. However, by 2014, this has become 1.2% and 21.7% respectively. Across all neighbourhoods in NYC, manufacturing employment has decreased by approximately 60% while employment in finance and real estate has increased 56% and employment in professional services has increased by 198%.

Beaudry et al. (2010) argue that technological change has increased the relative productivity of skilled workers and that cities adjust to skill-intensive industries endogenously, given their supply of high-skilled workers. In NYC the financial sector is a large fraction of employment. Since the financial sector employs skilled workers, NYC should see an increase in the demand for skilled workers following skill-biased technological change. In standard residential choice models households maximize their utility by choosing among neighbourhood amenities, housing (and relocation) costs, and commute costs, given income constraints. Changes in the spatial distribution of employment and differential wage growth

[5] Industries are classified based on the 1990 Census Bureau industrial classification scheme, IND1990. As such, they differ from both the NAICS and the SIC classification schemes. I manually translate between the two and apply a correction based on the ZCBP and CTPP year 2000 relationships. Use of the 2-digit classification codes eases comparison between the two datasets.

[6] For the moment, I am restricting commute times to be no greater than 90 minutes.

across industries will affect residential demand. When the demand for an area increases, depending on the supply elasticity of housing, prices will generally rise. As incomes rise for a subset of the population, house prices are bid up and incumbent residents are crowded out.

Rosen (1979) and Roback (1982) provide the foundation for this model while Moretti (2013, 2011) provides the framework. In this model skilled and unskilled workers compete for housing in the same housing market. For ease of exposition, consider two industries: manufacturing, which employs low-skilled labour, and finance, which employs high-skilled labour; and two neighbourhoods in NYC: a neighbourhood in downtown Manhattan and a neighbourhood in outer Queens.

The indirect utility for worker i in skill group $s \in \{h, l\}$ living in neighbourhood $n \in \{m, q\}$ and working in neighbourhood $n' \in \{m, q\}$ is:

$$U_{i,s,n',n} = W_{sn'} - C_{cn'} - P_n + T_{isn} \tag{2.2}$$

where $W_{sn'}$ is the nominal wage, $C_{n'n}$ is the cost of commuting between locations n and n', P_n is the cost of housing and T_{isn} is the value of local amenities which can differ across skill groups. Each worker supplies one unit of labour and consumes one unit of housing.

As in Moretti (2013) workers have idiosyncratic preferences over location specific amenities. I also allow workers to have idiosyncratic preferences over commute tolerance.

$$T_{isn} \sim U[-t_s, t_s] \tag{2.3}$$

$$C_{in'n} \sim U[-\underline{c}, -\bar{c}] \tag{2.4}$$

A worker chooses to live and work in Manhattan if and only if,

$$U_{i,s,m,m} = max\{U_{i,s,m,m}, U_{i,s,m,q}, U_{i,s,q,m}, U_{i,s,q,q}\} \tag{2.5}$$

Appendix 2.B presents the set of equilibrium conditions. In equilibrium, inframarginal workers can earn economic rents due to idiosyncratic preferences over location specific amenities and commute tolerance however, the marginal worker of both types $s \in \{h, l\}$ must be indifferent between all possible location combinations:

$$W_{sm} - W_{sq} = (T_{sm} - T_{sq}) - (P_m - P_q) \tag{2.6}$$

Firms are assumed to be perfectly mobile price takers facing a constant returns to scale Cobb-Douglas production technology. The firms' production function in neighbourhood n, is symmetric for skilled and unskilled labour:

$$y_{sn'} = X_{sn'} N^{h}_{sn'} K^{1-h}_{sn'} \tag{2.7}$$

Following a productivity shock:
Assume that the *productivity* of high skilled workers in downtown Manhattan increases relative to the productivity of low skilled workers and therefore the *demand* for skilled workers increases in downtown Manhattan relative to the demand for low skilled workers.

Denoting X_{sn} as the skill and neighbourhood-specific productivity shifter on the production function,

$$X_{h,m,t=2} = X_{h,m,t=1} + \Delta \tag{2.8}$$

where $\Delta > 0$ is the difference in productivity between $t = 1$ and $t = 2$.

Following this, the nominal wages of high skilled workers in Manhattan increase by Δ/h, where h is the Cobb-Douglas return to labour. This may generate a change in their preference for living in downtown Manhattan relative to Queens (as per Equation 2.5). As skilled workers move to downtown Manhattan the cost of housing in downtown Manhattan increases while the cost of housing in Queens declines. The amount by which house prices change in each neighbourhood depends on the elasticity of housing supply, locational preferences, and the flows of high skilled workers.

The nominal wages of unskilled workers living in Manhattan are unchanged but the cost of housing has increased. This in turn causes a reduction in their real wage. Those low skilled workers for which Equation 2.5 no longer holds now prefer living in Queens over Manhattan. Because unskilled workers compete for scarce housing with skilled workers and the inflow of skilled workers into Manhattan residences causes some low skilled workers to leave Manhattan for Queens, the equilibrium number of skilled workers in Manhattan increases while the equilibrium number of unskilled workers decreases. Depending on individual preferences for commuting, some of these unskilled workers will continue to work in Manhattan under the assumption that their wages in Manhattan are no less than in Queens. For those with C_{mq} sufficiently high, they will search for work in Queens rather than commute.

Clearly as the number of neighbourhoods and industries increase this general equilibrium problem becomes exponentially more difficult, but the intuition remains. As long as high-skilled and low-skilled workers compete for housing, shocks to particular industries alter nominal wages, neighbourhood preferences, house prices, and thus real wages. As house prices increase, workers will relocate outside of the city center and commute to work, conditional on the distance (i.e. cost) of the commute. The remainder of this chapter evaluates the quantitative importance of industry and job growth on neighbourhood change.

2.4 Empirical Strategy

Industries that experience increases in employment demand will generate an increases in the marginal willingness to pay for housing by employees of these industries. This is due to increases in their real wages. Using the geographic distribution of jobs around NYC, I show that increasing labour demand (and employment income) can explain a large part of the observed changes in neighbourhood incomes and house prices. In doing so, I first establish the relationship between price growth and income growth across New York City neighbourhoods. Between 1990 and 2010, the relationship between the price growth and income growth in neighbourhood n can be characterized as:

$$\mathrm{PG}_{n,1990-2010} = \alpha + \beta\Delta ln(\mathrm{Income}_{n,1990-2010}) + \epsilon_n \tag{2.9}$$

Several recent papers have looked at similar relationships between urban resurgence and labour market characteristics (i.e. wages or education level) and have clearly discussed the endogeneity that is inherent in the above equation (Baum-Snow and Hartley (2016), Edlund et al. (2015), Guerrieri et al. (2013)). House prices and incomes could both change for reasons related to neighbourhood amenities or housing supply and therefore falsely attribute price growth to income growth. I instrument for $\Delta ln(\mathrm{Income})$ using a spatial-Bartik, industry shift-share measure (Bartik (1991)). In doing so, I quantify the effect of changes across labour markets on residential housing demand.

2.4.1 Spatial-Bartik Instrumental Variable

Increases to average neighbourhood incomes have both housing demand and endogenous housing supply responses. Using a spatial-Bartik instrumental variable, I am able to isolate labour demand shocks that drive housing demand. By spatially weighting the labour demand shocks, I allow for neighbourhood-specific housing demand growth that differs based on neighbourhood location and initial industry composition.

The intuition of this instrument is as follows: industries receiving positive labour demand shocks will generate increases in the incomes of workers in those industries and an increase in housing demand in nearby neighbourhoods. Based on the labour force composition across neighbourhoods in 1990, I can predict relative labour demand.[7]

The Bartik instrument is constructed as follows: for each neighbourhood i at time $t \in \{1990, 2013\}$,

$$z_{i,t} = \frac{1}{N_{i,1990}} \sum_{k=1}^{K} n_{k,i,1990}(ln(Employment_{k,-NYC,t}) - ln(Employment_{k,-NYC,1990})) \tag{2.10}$$

Where $N_{i,1990}$ is the total employment in neighbourhood i in base year 1990; $n_{k,i,1990}$ is the total employment in industry k in neighbourhood i in base year 1990. The term in brackets is the employment growth in industry k in the United States, excluding New York City. As such, based on the fraction of employment in industry k, in neighbourhood i, in 1990, national employment growth in industry k generates an exogenous predictor for labour demand growth in neighbourhood i.

The spatial component is constructed such that an increase in labour demand in neighbourhood i is predicted to increase housing demand in neighbourhood j proportional to the distance between neighbourhoods i and j. For example, a labour demand shock in Midtown (NYC's CDB) should not only affect incomes in Midtown but also in neighbourhoods around Midtown (i.e. lower Manhattan and into Brooklyn).

[7] 1990 is chosen as the base year because it is the first year for which I have industry data at the neighbourhood level (source: CTPP).

Formally, the predicted income growth in neighbourhood j is,

$$Z_{j,t} = \sum_{i \neq j} z_{i,t} e_{i,j}^{-\kappa \tau_{i,j}} \tag{2.11}$$

The instrument for income growth in neighbourhood j at time t is the summation of the expected labour demand growth in all other neighbourhoods i, weighted by their relative distance in commute times; τ is the commute time between neighbourhoods i and j; κ is the decay parameter for weighting neighbourhoods discussed and estimated in Section 2.2. I exclude the labour demand shock to own neighbourhood j in the construction of $Z_{j,t}$ to further isolate a demand shock for living in neighbourhood j that is uncorrelated with unobservables driving neighbourhood change.

Table 2.A2 of the Appendix summaries the predicted housing demand shock in 1994, 2004, and 2013, as an example. The spatial-Bartik instrument gets larger each year given that US employment has grown between 1990 and 2013. In each year, the instrument is book-ended by neighbourhoods in Brooklyn having the largest predicted increase in housing demand, and neighbourhoods in Queens or Staten Island having the smallest predicted increase in housing demand. Figure 2.A3 plots the time-series evolution of $Z_{j,t}$ for each neighbourhood in NYC detrended from growth in NYC housing demand. There is clear heterogeneity in the magnitude of the shocks across neighbourhoods, even after removing the macroeconomic time trends. Figure 2.1 (a) depicts the actual income growth across New York City, (b) shows the predicted labour demand growth from the Bartik instrument (Equation 2.10), and (c) shows the predicted housing demand from the spatial-Bartik instrument (Equation 2.11). The pattern of housing demand shocks which is an instrument for income growth, closely resembles the actual pattern of income growth. Based on this instrument, I expect the labour demand shock to operate strongest near Midtown (the CBD) resulting in both higher incomes and higher house prices.

2.4.2 Labour Demand Shocks

In this section I identify the long-run and short-run relationship between income growth and house price growth. I also discuss the implied housing supply elasticity.

Labour demand driven price growth:

$$\begin{aligned} &\mathrm{PG}_{n,1990-2010} = \alpha + \beta \Delta ln(\widehat{\mathrm{Income}_{n,1990-2010}}) + \epsilon_n \\ &\rightarrow \Delta ln(\mathrm{Income}_{n,1990-2010}) = \gamma Z_{n,2013} + \mu_n \end{aligned} \tag{2.12}$$

$$\begin{aligned} &\mathrm{PG}_{n,t} = \alpha + \beta \Delta ln(\widehat{\mathrm{Income}}_{n,t}) + a_n + a_t + \epsilon_{n,t} \\ &\rightarrow \Delta ln(\mathrm{Income}_{n,t}) = \gamma Z_{n,t} + \alpha_n + \alpha_t + \mu_{n,t} \end{aligned} \tag{2.13}$$

As before, PG is the house price growth rate in neighbourhood n at time t; $\Delta ln(\widehat{\mathrm{Income}})$ is the change in log incomes instrumented for with the spatial-Bartik instrument; a_n and a_t are neighbourhood and year fixed effects (where no t subscript is specified, regressions are the long differences over the sample period, 1990 to 2010).[8] Equations 2.12 and 2.13 estimate the price growth in a neighbourhood following an exogenous shock to labour demand.

2.5 Results

2.5.1 Income Driven Price Growth: Long Run

Table 2.1 presents the baseline results from estimating Equation 2.12. Panel A, columns (1) to (3) present the un-instrumented relationship between income growth and price growth at the neighbourhood level between 1990 and 2010; columns (4) to (6) present the corresponding instrumented results. Panel B presents the reduced form relationship between price growth and the spatial-Bartik instrument in columns (1) to (3) and the first stage for the instrumented results in columns (4) to (6). In columns (2) and (5) I also control for the distance to Midtown (NYC's CBD) and in columns (3) and (6) I control for the interaction between income growth and the distance to Midtown. Panel C offers an interpretation of the results as a fraction of the actual variation observed in the data.

Looking first at the OLS results in panel A, column (1). A one standard deviation increase in income ($1\sigma = 0.221$) is correlated with a 0.43 of a standard deviation increase in price growth rates ($1\sigma = 0.44$). Unsurprisingly, as distance to Midtown increases price growth rates decrease as does the rate of income growth. As discussed earlier, income growth is likely to affect both housing demand and housing supply and to be correlated with unobservables affecting both neighbourhood incomes and house prices. This implies that the OLS results are likely biased. To address this bias, I make use of the spatial-Bartik instrument as an exogenous predictor of labour demand. The corresponding instrumented results are thus the house price growth that is attributable to income growth generated from labour demand growth.

The coefficient on my baseline IV result is 3.790 (column (4)) and is almost 4.5 times larger than my OLS results. This implies that a one standard deviation increase in income growth ($1\sigma = 0.221$) generates a 1.94 standard deviation increase in house price growth. However, a one standard deviation increase in predicted labour demand (Z_j: $1\sigma = 1.21$) increases incomes by 0.28 of a standard deviation which in turn increases house price growth by 0.54 of a standard deviation. Imposing an assumption of normality, exogenous shocks to labour demand can explain approximately 21% of the observed income growth and 41% of the observed house price growth. Inputting actual changes to employment across labour markets into the estimated model suggests that endogenous movements across local labour markets explain an additional 30% of the variation observed in house price growth between 1990 and 2010.

A tentative explanation for the magnitude of the IV results compared to the OLS results is discussed in Eisensee and Strömberg (2007) as well as Serafinelli (2017). If there are heterogeneous effects across neighbourhoods from a labour demand shock, then it is likely that the effect of income growth on house prices is greater for those neighbourhoods that only saw their incomes rise because of a labour demand shock. In other words, OLS measures the average effect of a income growth across all neighbourhoods. However, IV estimates the average effect for the subset of neighbourhoods that would not have seen their

[8] While the house price data extends from 1974 to 2014 and all other data sets extend to 2013, for comparison with census data I restrict my analysis to changes between 1990 and 2010. Results do not qualitatively change if I vary this.

prices rise in the absence of labour demand shock. If the effect of a labour demand shock on house prices is larger for neighbourhoods that would not have appreciated otherwise the IV estimates will exceed those of consistent OLS.[9]

A potential concern with these results is the representativeness of the NYC sales data and the HMDA data. To address this concern, I estimate the same model using income and house price data from the 1990 and 2010 decadal census. While no quality adjustments have been made to the house values in the census data, the average value is close to that of the NYC sales data. While the average house price increase using the NYC sales data is about 197%, the census equivalent is 171% (Appendix Table 2.A3). On the other hand, the increase in average incomes from the HMDA data is larger than that in the census – 17% compared to 6%. This is expected given that the HMDA data exclusively includes the incomes reported on originated mortgage applications. Table 2.A4 of the appendix reproduces the results of Table 2.1 using the decadal census data. Comfortingly, the results are not quantitatively different from each other. Next, I turn to estimating the implied elasticity of housing supply that is necessary to justify the magnitude of these estimates.

2.5.2 Elasticity of Housing Supply

As housing demand increases, house prices are expected to rise. However, the magnitude with which house prices rise is not only dependent on the magnitude of the labour demand shock but also the elasticity of housing supply. For a given labour demand shock those neighbourhoods that were inelastically supplied in 1990 are expected to see larger changes in house prices.

Using neighbourhood housing quantities provided in the 1990, 2000, and 2010 Census, I construct a variable for the percentage change in housing units and estimate:

$$\Delta ln(\text{Quantity}_{1990-2010}) = \alpha + \beta \Delta ln(\widehat{\text{Income}_{1990-2010}}) + \epsilon_n \tag{2.14}$$

As above, the change in neighbourhood income is instrumented for with the spatial-Bartik. Table 2.2 presents the results as well as the corresponding implied supply elasticities. Over all housing types, I estimate a negative housing supply elasticity (column (2)). However, when I separately look at owner occupied (column (4)) and renter occupied (column (6)) housing units, the elasticity of housing supply is positive for owner occupied units and negative (and insignificant) for renter occupied units.

There are several plausible explanations for these results. First, there may be some substitution between renting and owning – when incomes rise, housing demand increases. If this also generates an increase in the supply of owner occupied housing units — either through the construction of new condominiums or the conversion of rental units into owner occupied units (both of which have happened in NYC)

[9] Making use of an example in Eisensee and Strömberg (2007), the following better illustrates this point. Neighbourhoods may be categorized into three groups: (a) those that never experience large changes to price growth, (b) those that always experience large changes to price growth, and (c) those that only experience large changes to price growth if and only if they experience income growth. OLS uses group (c) to estimate β. Neighbourhoods can also be categorized into groups according to potential income growth: (i) those that never experience income growth, irrespective of labour demand pressures, (ii) those that always experience income growth, irrespective of labour demand pressures, and (iii) those that only experience income growth if the labour demand shock is sufficiently large. IV estimates of β are therefore the share of group (c) in group (iii). This is the average effect for the group of marginal neighbourhoods.

— I would expect to find owner occupied housing increase and renter occupied housing decrease or remain unchanged. If this is indeed the case it would help to explain the aversion to gentrification by groups that tend to rent. Alternatively, if the increase in labour demand is due to favourable macroeconomic conditions, this could explain an increase in housing construction. Finally, it could be the case that neighbourhoods with high incomes also have strict regulatory practices, especially with respect to building. Therefore, when a neighbourhood gentrifies it becomes subject to supply restrictions, resulting in the estimated negative supply elasticity.

The quantity of housing increased by 15% between 1990 and 2010 however, this was largely driven by increases in owner occupied housing units which increased by 26% compared to renter occupied housing which increased by 4% (Appendix Table 2.A3). Figure 2.A5 in the Appendix includes the frequency of changes in housing units across neighbourhoods. Owner-occupied housing has a distribution that lays to the right of the renter-occupied housing, consistent with the greater increase in owner-occupied units. The estimates in Table 2.2 imply a housing supply elasticity of 0.21 for owner occupied units – for a 10% increase in prices, quantities increase by about 2%. This suggests that the IV estimates are identifying off of those neighbourhoods that are relatively inelastic and receive a large housing demand shock.

2.5.3 Income Driven Price Growth: Short Run

Table 2.3 presents the OLS and IV short run results in columns (1) to (4) and (5) to (7), respectively. The coefficient of 0.548 in column (1) implies that a one standard deviation increase in incomes ($1\sigma = 0.018$) is correlated with a 0.38 of a standard deviation increase in house price growth rates. Again, the magnitude of the IV is quite a bit larger at 3.790. A one standard deviation increase in predicted labour demand ($1\sigma = 1.721$) increases incomes by 0.33 standard deviations, which in turn increases house price growth by 0.85 standard deviations. This suggests that exogenous labour market demand shocks can explain 26% of the annual variation in income growth and 61% of the variation in house price growth. Estimating the model with actual changes to employment suggests that endogenous changes to labour supply can explain an additional 25% of the variation. Columns (6) and (7) suggest that labour market demand shocks can better explain increases in house prices than decreases.

To recap, between 1990 and 2010, I find that labour market demand shocks can account for approximately 21% of the observed variation in income growth and 41% the variation in house price growth rates. I also provide evidence that endogenous changes to labour supply across neighbourhoods can explain an additional 20% of the variation in home owner incomes and 30% of the variation in house price growth rates. Looking over the year to year variation in income growth and house price growth rates, the explained variation is similar for income growth rates but somewhat higher for house price changes. This suggests that in New York City house prices are sensitive to year to year fluctuations in the macro economy. In both the short and the long run endogenous changes to labour supply appear to reinforce the direction of neighbourhood change.[10]

[10] As robustness checks I perform the same analysis in the long run and the short run removing employment in finance and real estate given its dominance in New York City, removing Midtown, and adding a Manhattan fixed effect. Results remain unchanged.

2.6 An Extension Beyond New York City

While New York City provides an interesting case study for the labour market implications on neighbourhood change, there is some concern that NYC is itself a unique city. The specific impact that labour markets have had on the patterns of gentrification seen in NYC may not exist in other cities. The labour market in NYC disproportionately employs high-skilled labour in finance and real estate, compared to the rest of the United States. Given that in my base year (1990) NYC already had a large concentration of employment in finance and that finance grew nationally, the relationship may be different in a city such as Detroit, which in 1990 had a large concentration of employment in manufacturing – an industry that declined nationally. As such, I extend part of the above analysis to Detroit, Michigan. To address the fact that Detroit might also be a unique example of the impact of labour markets on neighbourhoods change, I also explore the relationship in Portland, Oregon. Portland is currently experiencing instances of gentrification due to the growth of technology industries along the west coast, but itself does not have a dominant industry. Using the zipcode business patterns data and the HMDA data I explore the first-stage relationship between the labour markets and neighbourhood incomes in the Detroit MSA and the Portland MSA. Lacking high-frequency sales data I am not able to properly estimate the second-stage over changes in house price indices, I do however include preliminary regression results using house prices as reported in the census data.

2.6.1 Labour Markets in Detroit, Michigan

Between 1990 and 2010 employment in the Detroit MSA decreased by about 23%; Table 2.A6 includes the by-industry breakdown of employment changes. Compared to the rest of the United States employment has decreased substantially in manufacturing industries however, similar to New York City, employment in professional and personal services have increased. In constructing my spatial-Bartik variable I once again use the CTPP data to obtain a measure of the tolerance for commuting in Detroit (κ: Equation 2.1) – I estimate that for a 10 minute increase in commute time worker flows decrease by 38% (Figure 2.A9). This implies that workers are more sensitive to commute times in Detroit than in New York City with commute times in Detroit being 20 percent lower on average than in New York City.

Figure 2.2 shows (a) the spatial pattern of actual income growth in the Detroit MSA, (b) the predicted labour demand growth, and (c) the predicted housing demand growth as estimated from the spatial-Bartik instrument. In NYC there is a visually evident positive relationship between income growth and predicted housing demand shocks; in Detroit, this relationship is inverted. The peripheral neighbourhoods further from the Detroit city center have seen the largest actual increases in homeowners incomes. Based on 1990 employment patterns and commute times, the Bartik instrument generates a negative relationship between actual income, predicted income, and predicted housing demand. This negative relationship is statistically significant and quantified in the first-stage regression results.

Table 2.4 Panel A presents the first-stage results from estimating Equations 2.12 and 2.13; column (1) considers the relationship between the long-run income growth and the instrumental variable, column (3) considers the short-run (annual) relationship. Column (2) attempts at estimating the second-stage relationship between house price growth and changes in labour demand using house prices as reported in the decadal census.[11] The coefficient for the long-run relationship is -0.073. This implies that a one standard deviation increase in predicted labour demand ($1\sigma = 0.712$) decreases incomes by 0.25 of a standard deviation. Given the 1990 base-year industry composition in Detroit, national changes in employment are expected to decrease incomes in the Detroit MSA. This suggests that exogenous shocks to labour demand can explain approximately 19% of the observed decrease in homeowner incomes in Detroit between 1990 and 2010 (in New York City this figure is 21%). Inputting actual changes to employment across labour markets suggests that endogenous factors can explain an additional 74% of the variation observed in income growth (decline) between 1990 and 2010. This is significantly larger than the 30% estimated in New York City. Column (2) shows no significant second-stage relationship between house price changes and income growth. There could be a couple of explanations for this: either there have been substantial changes to the quality of houses in Detroit which the prices as reported in the census do not control for; or, as in New York City, exogenous labour demand shocks can not explain decreases in house prices.

Looking to the short-run I find that labour market demand shocks can account for approximately 15% of the year to year changes in homeowner incomes. The coefficient of -0.003 in column (1) implies that a one standard deviation increase in predicted labour demand ($1\sigma = 1.33$) decreases incomes by 0.19 of a standard deviation.[12] Endogenous changes across labour markets explain an additional 24% of the annual variation observed in income growth rates. Compared to New York City, exogenous shocks explain somewhat less of the annual variation in incomes however, and endogenous changes explain slightly more.

As in New York City, exogenous shocks to labour demand can explain approximately one fifth of the 20-year variation observed in changes to homeowner incomes in Detroit. However, unlike New York City long-run endogenous factors explain a larger portion of the income changes in Detroit. This suggests that at least in part, the long-run decline observed in Detroit was exacerbated by endogenous decisions by businesses and households.

2.6.2 Labour Markets in Portland, Oregon

Looking now to Portland, Oregon Figure 2.3 shows (a) the spatial pattern of actual income growth in the Portland MSA, (b) the predicted labour demand growth, and (c) the predicted housing demand growth. Going through the same exercise as above, in calculating the tolerance for commuting in Portland I estimate that for a 10 minute increase in commute time, worker flows decrease by 43% (Figure 2.A10). Compared to both NYC and Detroit, workers in Portland are more sensitive to commute times with the average commute being approximately 15% lower than in Detroit.

Table 2.4 Panel C presents the main results. The coefficient from the first stage regression of income changes on predicted shocks to labour demand is 0.107 (column (1)). This implies that a one standard deviation increase in Z_{jt} ($1\sigma = 0.478$) causes incomes to increase by 0.30 of a standard deviation. As such, exogenous labour market demand shocks can explain approximately 24% of the overall variation observed in income growth rates in Portland between 1990 and 2010. However, labour demand shocks have no

[11] Using house prices as reported in the census does not control for changes in the quality of the housing stock. Furthermore it includes all homes in a neighbourhood and therefore does not necessarily represent demand driven price growth.

[12] Refer to Figure 2.A11 of the Appendix for a plot of the spatial-Bartik instrument in Detroit, Michigan.

explanatory power to the short run changes to income growth rates in Portland (column (3)).[13] When inputting actual changes to employment across labour markets I find that similar to Detroit, endogenous changes to labour supply and demand can explain an additional 70% of the variation observed in income growth between 1990 and 2010. So, while exogenous shocks to labour markets have contributed to some of income growth and gentrification seen in Portland over the last 20 years, endogenous decisions on the part of businesses and households on where to locate have had a substantial impact.

2.7 Counterfactual Labour Markets

An interesting thought experiment is to consider how the distribution of house price growth rates and patterns of neighbourhood change would look under alternative evolutions of the labour market. In this section I consider my estimated models under several counterfactual environments. The purpose of this exercise is to explore how neighbourhood change may have looked under hypothetical labour markets. In particular, I estimate how the patterns of neighbourhood change may have looked if: (i) labour demand had remained unchanged from its 1990 level; (ii) industry employment had grown at the National rate; (iii) all industries had grown at the same rate.

Recall, the distribution of prices predicted by exogenous shocks to labour markets explains 41% of the observed 20-year variation in house prices and 61% of the annual variation. Figure 2.A6 presents the distribution of predicted price growth rates between 1990 and 2010 (a) and annually (b). Labour demand shocks are not able to predict price growth at the far right-tail of the 20 year distribution or either tail of the annual distribution. These extreme values observed in the data are likely explained by other changes to neighbourhood fundamentals, such as crime rates or changing amenities.

First, if labour demand had remained unchanged from its 1990 level growth in prices is attributable entirely to common trends across New York City rather than differences in labour markets. Taking the coefficient estimates from Equation 2.12 and setting labour demand growth equal to zero, changes in prices are the same across all boroughs – predicted to be approximately 60% lower than in my baseline specification. In this counterfactual world, all neighbourhoods in New York City would be declining in incomes and prices today.

Second, I consider the scenario where industry labour demand had grown at the same rate in New York City as in the rest of the United States. In this scenario industries that saw big drops in employment in New York City between 1990 and 2010 but more modest drops elsewhere (i.e. manufacturing, wholesale trade, and education) will see increases in the predicted income growth for neighbourhoods near 1990 employment locations. The reverse is true for industries that grew more in the New York City relative to elsewhere in the United States (i.e. business, professional, and personal services). As before, I take the predicted estimates from Equations 2.12 and 2.13 and set changes in employment equal to the national growth by industry. Figure 2.A6 (a) shows a right-ward shift in the distribution of overall predicted house price growth, compared to the baseline.

Across NYC price growth rates are predicted to be 18% higher and the spatial variation 12% higher. Annually, average prices would be growing 0.5% faster but the spatial variation in price growth across

[13] Refer to Figure 2.A12 of the Appendix for a plot of the spatial-Bartik instrument in Portland, Oregon.

the city would be 38% higher. Although employment increased in New York City between 1990 and 2010, it increased more in the rest of the United States. In particular, it increased more in industries for which New York City already had a large percentage of employment in 1990 (for example, finance and real estate, health care, and education). In this counterfactual, 40% more neighbours are predicted to gentrify.

Figure 2.4 maps the concentration of price growth around the City: (a) price growth between 1990 and 2010 as predicted in my baseline specification; and (b) price growth predicted after the imposition of national employment growth rates. Comparing (a) to (b), there is a movement of price growth away from upper Manhattan and further into the peripheral neighbourhoods of Brooklyn. While outer Queens is still predicted to decline in prices, the gentrification seen in upper Manhattan (i.e. Central Harlem and Washington Heights) and the Bronx is no longer apparent.

Finally, I consider how the pattern of neighbourhood change may have looked in New York City if employment across all industries had grown at the same rate; for example, if employment in manufacturing had grown at the same rate as employment in business services. The distribution of prices in this world is quite different. The predicted price growth between 1990 and 2010 would be almost 97% higher across the City, the variation across the city would be 113% higher, but the distribution would be almost uniform. The annual patterns of growth are predicted to be 2% higher across the city, and the spatial variation is predicted to increase by 124%.

This counterfactual analysis reinforces the importance of changes in labour markets on the evolution of neighbourhood change. While other considerations — such as decreasing crime rates or improved amenities — are necessary to explain the full distribution of house price movements across New York City, labour markets are a significant determinant of a neighbourhood's trajectory.

2.8 Discussion and Conclusion

This chapter considers the relationship between the changing characteristics of labour demand and neighbourhood change. Recent literature looks at trends in the re-urbanization across major cities in the United States but is relatively absent of a discussion of neighbourhood decline as the opposite of gentrification. Characterizing these within-city dynamics is an important component towards a better understanding of gentrification. In this chapter I begin by outlining a theoretical framework within which to view the localized impact of productivity shocks on neighbourhoods: following a positive productivity shock to an industry such as finance, the wages of workers in finance increase. This wage increase in turn increases the demand to live near the CBD (where the finance industry is located). Then, this housing demand increase causes an increase in house prices. As house prices in the CBD increase, low-income workers experience a real wage decrease causing some of them to move away from the CBD. The rate at which neighbourhoods around the CBD gentrify and peripheral neighbourhoods become poorer and less expensive will depend on the tolerance for commuting, locational preferences, and the elasticity of housing supply.

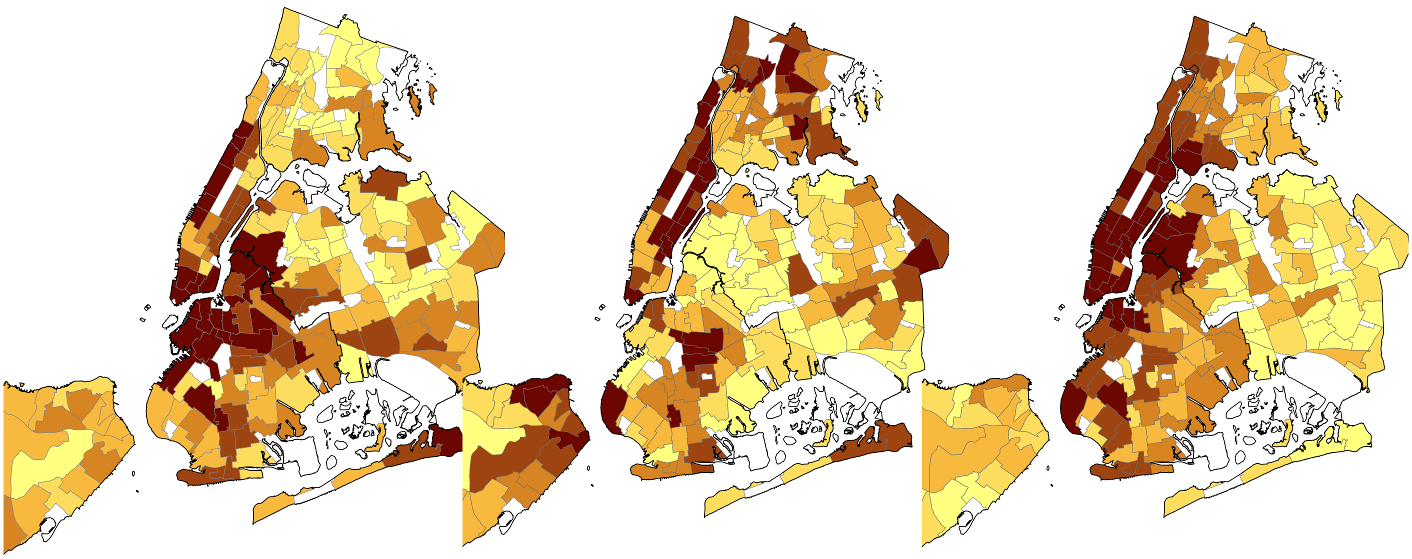
To empirically estimate the magnitude with which changes in labour demand have contributed to neighbourhood change I construct a spatially weighted Bartik instrumental variable as an exogenous measure of income growth in a neighbourhood – positive labour demand shocks generate subsequent housing demand shocks through the effect on worker incomes. I find that exogenous changes to labour demand can explain 21% of the 20-year variation in neighbourhood income growth and 41% of the 20-year variation observed in house price growth rates; a one standard deviation increase in predicted housing demand increases house price growth by 0.54 standard deviations. Estimating my model with actual changes to employment suggests that endogenous factors can explain an additional 30% of the variation observed in house price growth rates. Similarly, year to year shocks to labour demand are estimated to explain 26% of the annual variation in neighbourhood income growth and 61% of the annual variation in house price growth rates. The similarity in the magnitude of the effect of labour market shocks over the two time horizons suggests that New York City is relatively resilient to year to year macro-economic fluctuations. That said, endogenous factors do play a bigger part in explaining the 20-year variation observed in house prices than they do in explaining the year to year variation.

Following this, I extend my analysis to looking at two very different cities: Detroit, Michigan and Portland, Oregon. While I do not have detailed price level data for either of these cities, the zipcode business patterns and HMDA data allow me to estimate the first-stage relationship between labour demand shocks and income growth. I find that in Detroit, exogenous changes to labour demand can explain about 19% of the overall decline in homeowner incomes between 1990 and 2010; in Portland, the estimated effect is 24%. Both of these estimates are in line with those for New York City – exogenous labour demand shocks can explain about one fifth to one quarter of the variation in neighbourhood income growth. Endogenous factors on the other hand have a much larger impact on observed income changes in Detroit and Portland than in NYC. This suggests that at least in part, the decline in incomes in Detroit and the gentrification in Portland is strengthened by endogenous decisions on the parts of households and firms.

Finally, to really understand the importance of labour demand as a driver of price increases, I estimate the predicted changes to neighbourhood income growth and house price growth in New York City under counterfactual evolutions of the labour market. In doing so I find that had New York City labour demand remained at its 1990 level, overall house price growth would be 65% lower; had labour demand grown at the national rate, overall house price growth would be 18% higher. As such, changes across labour markets can predict very different trajectories for neighbourhood housing demand, income growth, price growth, and neighbourhood change. Overall I have shown evidence to the importance of changes in labour markets and labour demand in explaining the within-city dynamics of neighbourhood change. While a comprehensive understanding of the implications of gentrification is for future work, characterizing these dynamics is a necessary first step.

Figures

Figure 2.1: Predicted Employment Growth and Neighbourhood Demand: New York City



(a) Actual Income Changes (1990 to 2013)

(b) Predicted Employment Growth (2013, based on 1990 industry shares)

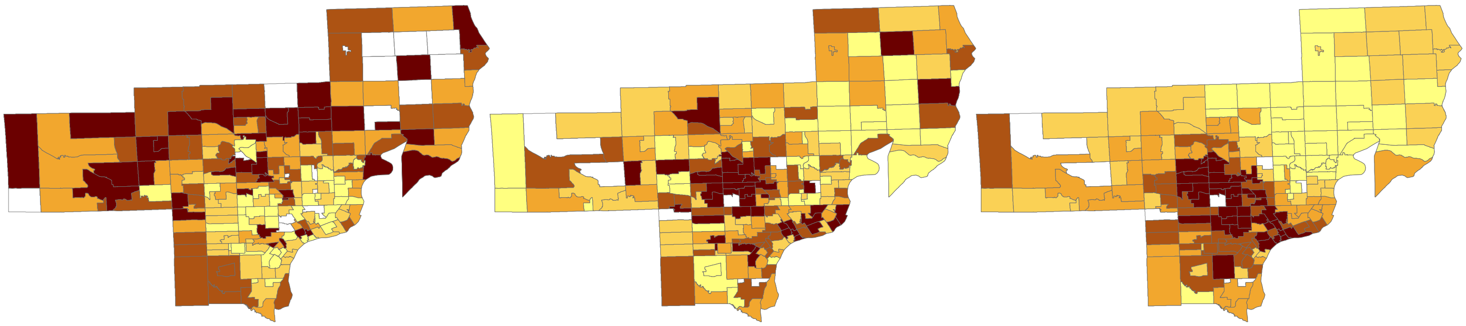
(c) Predicted Housing Demand (2013, based on 1990 industry shares)

Notes: The above figures depict the spatiality of the Bartik instrument in New York City. Figure (a) shows the actual change in neighbourhood incomes from 1990 to 2010. Figure (b) shows the predicted employment growth in 2013 based on their 1990 industry shares. Using the predicted employment growth and the distance between neighbourhoods, figure (c) shows the spatial-Bartik instrument for predicted employment growth (and thus housing demand growth). Changes are divided into six quantiles; darker shades denote the largest changes, while lighter shades denote the smallest changes.

Legend for Figs. 2.1 – 2.3

Quantile of the distribution		
	1st	(largest)
	2nd	
	3rd	
	4th	
	5th	
	6th	(smallest)

Figure 2.2: Predicted Employment Growth and Neighbourhood Demand: Detroit, MI MSA



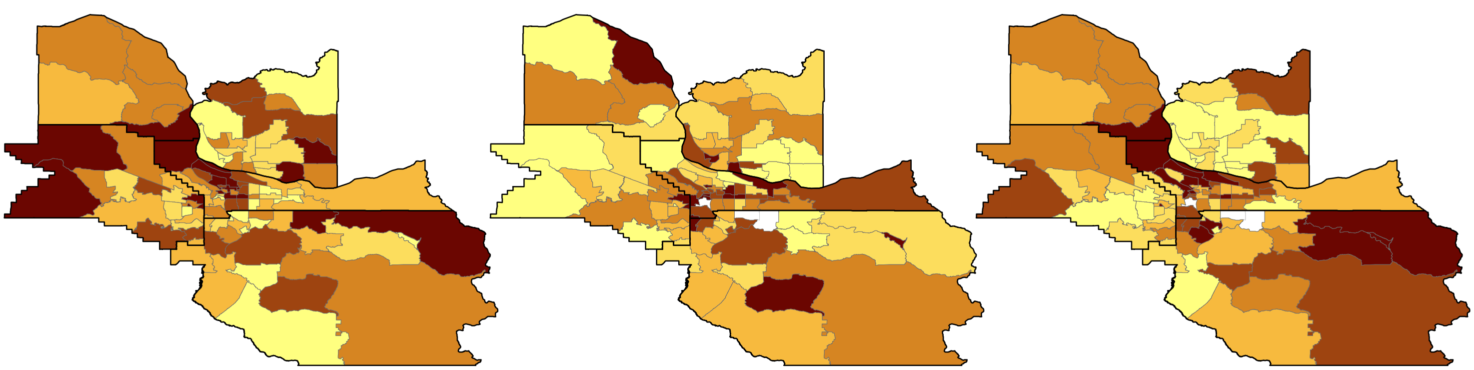
(a) Actual Income Changes (1990 to 2013)

(b) Predicted Employment Growth (2013, based on 1990 industry shares)

(c) Predicted Housing Demand (2013, based on 1990 industry shares)

Notes: The above figures depict the spatiality of the Bartik instrument in the Detroit, Michigan MSA. Figure (a) shows the actual change in zipcode incomes from 1990 to 2010. Figure (b) shows the predicted employment growth in 2013 based on their 1990 industry shares. Using the predicted employment growth and the distance between neighbourhoods, figure (c) shows the spatial-Bartik instrument for predicted employment growth (and thus housing demand growth). Changes are divided into six quantiles; darker shades denote the largest changes, while lighter shades denote the smallest changes.

Figure 2.3: Predicted Employment Growth and Neighbourhood Demand: Portland, OR MSA



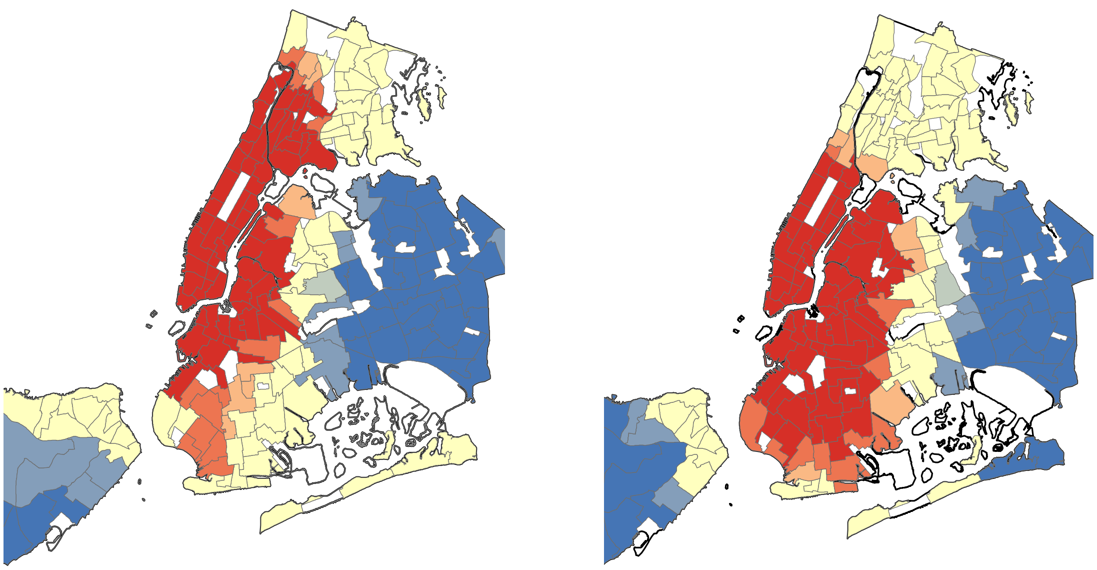
(a) Actual Income Changes (1990 to 2013)

(b) Predicted Employment Growth (2013, based on 1990 industry shares)

(c) Predicted Housing Demand (2013, based on 1990 industry shares)

Notes: The above figures depict the spatiality of the Bartik instrument in the Portland, Oregon MSA. Figure (a) shows the actual change in zipcode incomes from 1990 to 2010. Figure (b) shows the predicted employment growth in 2013 based on their 1990 industry shares. Using the predicted employment growth and the distance between neighbourhoods, figure (c) shows the spatial-Bartik instrument for predicted employment growth (and thus housing demand growth). Changes are divided into six quantiles; darker shades denote the largest changes, while lighter shades denote the smallest changes.

Figure 2.4: Counterfactual Distribution of Price Growth



(a) Baseline Price Growth (Fig. 2.1 (c)) **(b)** Counterfactual: National Growth Rates

Notes: Using ArcGIS' hotspot analysis the above figures map statistically significant spatial clusters of hot and cold spots as identified by the Getis-Ord Gi* statistic. The analysis is done over predicted price growth rates with (a) depicting the baseline results and (b) depicting the counterfactual environment in which labour demand in New York City grew at the national rate. Figure (a) is an alternative depiction of the spatial relationship shown in Figure 1(c)). Neighbourhoods in red belong to the significant cluster of high values, while neighbourhoods in blue belong to the significant cluster of low values.

Tables

Table 2.1: Income Driven House Price Growth: Long run

(% Δ Prices 1990 to 2010 in New York City)

Panel A:	OLS			IV		
Dependent Variable:		*Price Growth*			*Price Growth*	
	(1)	(2)	(3)	(4)	(5)	(6)
Δln(Income)	0.844	0.559	0.506	3.790	3.875	10.26
	(0.15)***	(0.13)***	(0.26)*	(1.10)***	(2.44)	(12.5)
Dist_Midtown (miles)		−0.052	−0.053		0.002	−1.300
		(0.01)***	(0.01)***		(0.04)	(2.42)
Dist_Midtown x Δln(Income)			−0.008			0.259
			(0.03)			(0.49)
R^2	0.18	0.40	0.40			
F-stat:				11.46	2.34	0.08
N	155	155	155	155	155	155

Panel B:	Reduced Form			First Stage		
Dependent Variable:		*Price Growth*			*Δln(Income)*	
	(1)	(2)	(3)	(4)	(5)	(6)
Z	0.194	0.105	0.100	0.051	0.027	0.029
	(0.03)***	(0.03)***	(0.03)***	(0.02)***	(0.02)	(0.02)
Dist_Midtown		−0.040	−0.044		−0.011	−0.010
		(0.01)***	(0.01)***		(0.01)*	(0.01)
Z x Dist_Midtown			0.024			−0.007
			(0.03)			(0.02)
R^2	0.29	0.37	0.37	0.09	0.11	0.11
N	155	155	155	155	155	155

Panel C:	Explained Variation					
	Income growth		Difference	House price growth		Difference
Exogenous ΔLD	0.28σ	≈ 21%		0.54σ	≈ 41%	
Actual ΔLD	0.54σ	≈ 41%	20%	1.05σ	≈ 71%	30%

Notes: *** statistically significant at the 1% level; robust standard errors. All regressions include county fixed effects. Columns (1) to (3) show the OLS relationship between price growth and income growth over the entire sample period. Columns (4) and (6) show the instrumented relationship. Panel B presents the reduced form regression results as well as the first stage relationship between the instrument ($Z_{j,t}$)and income growth. Interpretation: (OLS) a 1σ increase in $\Delta ln(Income)(= 0.22) \rightarrow 0.43\sigma$ increase in price growth ($1\sigma = 0.44$); (IV) a 1σ increase in $Z_{j,t}(= 1.21) \rightarrow 0.54\sigma$ increase in price growth.

Table 2.2: Income Driven Housing Supply Changes

(% Δ ln(Quantities) in New York City)

Panel A:						
Dependent variable:	$\Delta ln(Q\ of\ housing)$					
	All housing		owner occ.		renter occ.	
	OLS	IV	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)
Δln(Income)	0.071	-0.401	0.142	0.789	0.063	-0.381
	(0.05)	(0.28)	(0.08)*	(0.38)**	(0.05)	(0.29)
N	155	155	155	155	155	155

Panel B:	Implied Elasticity of Housing Supply *Using Table 2.1 Results*					
$\frac{\%\Delta Q}{\%\Delta P}$	0.083	-0.106	0.168	0.208	0.075	-0.084

Notes: *** statistically significant at the 1% level; ** statistically significant at the 5% level; * statistically significant at the 10% level. Using housing quantities as reported in the decadal census data and calculating the percentage change between 1990 and 2010, the implied elasticity of supply is calculated as $\frac{\%\Delta Quantity}{\%\Delta Income} \times 1/\frac{\%\Delta Price}{\%\Delta Income}$. Columns (3) and (4) suggest that for owner occupied homes, a 10% increase in the price of houses generates approximately a 2% increase in the quantity of housing supply, for owner-occupied homes. This is consistent with housing supply being quite inelastic in New York City.

Table 2.3: Income Driven House Price Growth: Short run

Panel A:	OLS				IV		
Dependent Variable:	*Price Growth*				*Price Growth*		
	(1)	(2)	(3) ($PG\uparrow$)	(4) ($PG\downarrow$)	(5)	(6) ($PG\uparrow$)	(7) ($PG\downarrow$)
Δln(Income)	0.548	0.415	0.350	0.171	3.790	3.058	-13.24
	$(0.10)^{***}$	$(0.08)^{***}$	$(0.08)^{***}$	$(0.07)^{**}$	$(0.95)^{***}$	$(0.70)^{***}$	(76.0)
Dist_Midtown (miles)		-0.003	-0.002	-0.002			
		$(0.00)^{***}$	$(0.00)^{***}$	$(0.00)^{***}$			
R^2	0.16	0.38	0.33	0.39			
F-stat:					14.24	16.07	0.03
N	3,565	3,565	2,599	2,369	3,565	2,599	2,369

Panel B:	Reduced Form				First Stage		
Dependent Variable:	*Price Growth*				*Δln(Income)*		
	(1)	(2)	(3) ($PG\uparrow$)	(4) ($PG\downarrow$)	(5)	(6) ($PG\uparrow$)	(7) ($PG\downarrow$)
Z	0.013	0.009	0.008	0.001	0.003	0.004	-0.001
	$(0.00)^{***}$	$(0.00)^{***}$	$(0.00)^{***}$	(0.01)	$(0.00)^{***}$	$(0.00)^{***}$	(0.01)
Dist_Midtown		-0.002	-0.002	-0.001			
		$(0.00)^{***}$	$(0.00)^{***}$	$(0.00)^{***}$			
R^2	0.28	0.33	0.38	0.68	0.14	0.16	0.18
N	3,565	3,565	2,599	2,369	3,565	2,599	2,369

Panel C:	Explained Variation					
	Income growth		Difference	House price growth		Difference
Exogenous ΔLD	0.33σ	$\approx 26\%$		0.85σ	$\approx 61\%$	
Actual ΔLD	0.57σ	$\approx 43\%$	17%	1.49σ	$\approx 86\%$	25%

Notes: *** statistically significant at the 1% level; standard errors in parenthesis and clustered at the neighbourhood level. All regressions include county fixed effects. Columns (1) to (3) show the OLS relationship between price growth and income growth over the entire sample period. Columns (4) and (6) show the instrumented relationship. Columns (3) and (6) restrict the sample to those neighborhoods that have increasing price growth; columns (4) and (7) are those that have decreasing price growth rates.. Panel B presents the reduced form regression results as well as the first stage results.
Interpretation: (OLS) a 1σ increase in $\Delta ln(Income)(= 0.018) \rightarrow 0.38\sigma$ increase in price growth ($1\sigma = 0.026$);
(IV) a 1σ increase in $Z_{j,t}(= 1.721) \rightarrow 0.85\sigma$ increase in price growth.

Table 2.4: Income Driven House Price Growth: A comparison to Detroit, Michigan and Portland, Oregon

Detroit, Michigan MSA			
Panel A:	Long run changes		Short run changes
Dependent Variable:	*Δln(Income)*	*Price Growth* ‡	*Δln(Income)*
	(first stage)	(second stage)	(first stage)
	(1)	(2)	(3)
Z	-0.073	0.462	-0.003
	(0.03)***	(0.42)	(0.00)***
R^2	0.22		0.30
F-stat		8.21	
N	205	205	4,632

Panel B:	Explained Variation			
	Income Growth		Difference	
Exogenous ΔLD	-0.25σ	≈ 19%		} Long run changes
Actual ΔLD	-1.79σ	≈ 93%	74%	
Exogenous ΔLD	0.19σ	≈ 15%		} Short run changes
Actual ΔLD	0.51σ	≈ 39%	24%	

Portland, Oregon MSA			
Panel C:	Long run changes		Short run changes
Dependent Variable:	*Δln(Income)*	*Price Growth* ‡	*Δln(Income)*
	(first stage)	(second stage)	(first stage)
	(1)	(2)	(3)
Z	0.107	4.820	-0.002
	(0.04)***	(1.91)**	(0.01)
R^2	0.14		0.07
F-stat		8.86	
N	113	113	2,667

Panel D:	Explained Variation			
	Income Growth		Difference	
Exogenous ΔLD	0.30σ	≈ 24%		} Long run changes
Actual ΔLD	1.91σ	≈ 94%	70%	
Exogenous ΔLD	0.09σ	≈ 6%		} Short run changes
Actual ΔLD	0.35σ	≈ 27%	21%	

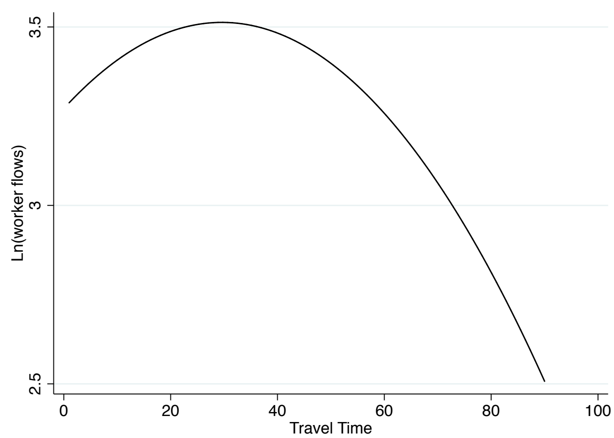
Notes: *** statistically significant at the 1% level; robust standard errors. ‡ Price growth as reported in the decadal census. All regressions include county fixed effects. Columns (1) estimate the first stage regression of long run changes in logged incomes on the spatial-Bartik instrument ($Z_{j,t}$) while columns (2) estimates the second stage regression for neighbourhood level price growth. Columns (1) uses the HMDA data for homeowner incomes while columns (2) uses average house prices as reported in the US census data. Columns (3) regresses the annual change in logged incomes on the spatial-Bartik instrument. Panels A and B present the results for Detroit, Michigan; panels C and D present the results for Portland, Oregon. Over this period, the average income growth in Detroit is -0.033; the average income growth in Portland us 0.261.
Interpretation: (Panel A(1)) a 1σ increase in $Z_{j,t}(= 0.712) \rightarrow -0.25\sigma$ increase in logged incomes ($1\sigma = 0.212$); (Panel C(1)) a 1σ increase in $Z_{j,t}(= 0.478) \rightarrow 0.30\sigma$ increase in logged incomes ($1\sigma = 0.170$).

Appendix 2.A

Figures

Figure 2.A1: Map of New York City Neighbourhoods

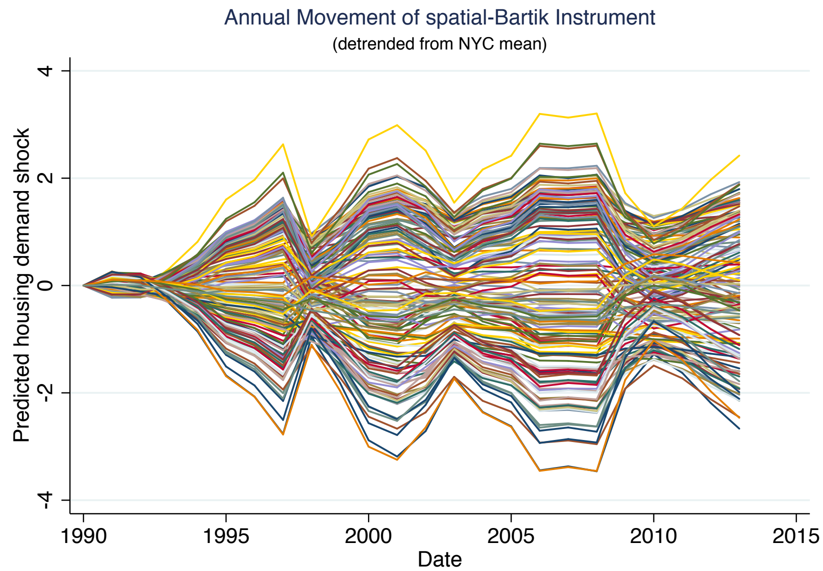
Figure 2.A2: Elasticity of worker flows with respect to commute time: New York City



Census Transportation Data (CTPP), 1990		mean	SD	min	max
flows	total workers leaving home during peak travel time	107	326	0	10,951
commute	average commute time during peak travel time	40	18	1	90
κ	decay parameter (estimated)	-0.0259			

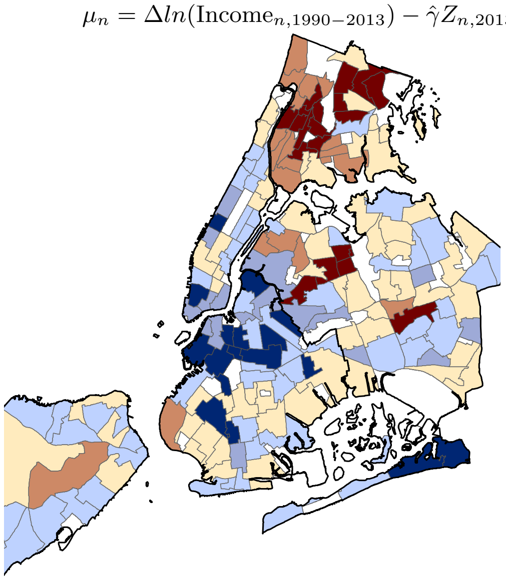
Source: 1990 Census Transportation Planning Package

Figure 2.A3: Annual Variability in Predicted Housing Demand Shocks: New York City



Notes: The above figure shows the annual variation in the spatial-Bartik instrument constructed as a measure of neighbourhood income growth in New York City. Each line depicts the instrument for a different neighbourhood.

Figure 2.A4: First Stage Residuals
$\mu_n = \Delta ln(\text{Income}_{n,1990-2013}) - \hat{\gamma} Z_{n,2013}$

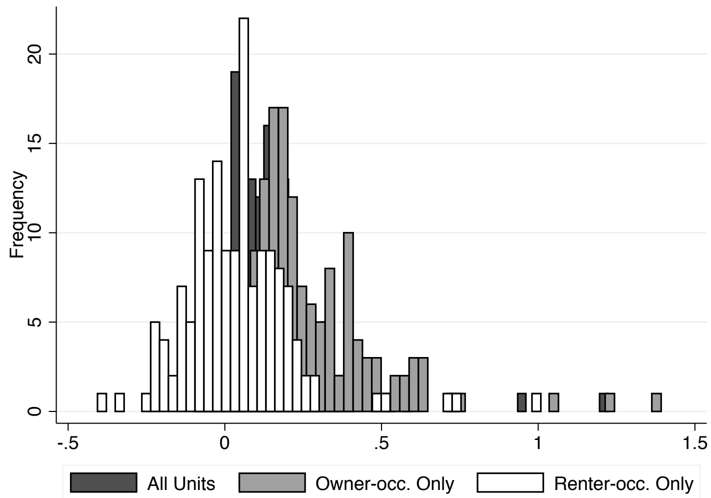


Legend for Figure 2.A4

Quantile of the distribution	
1st	(largest)
2nd	
3rd	
4th	
5th	
6th	(smallest)

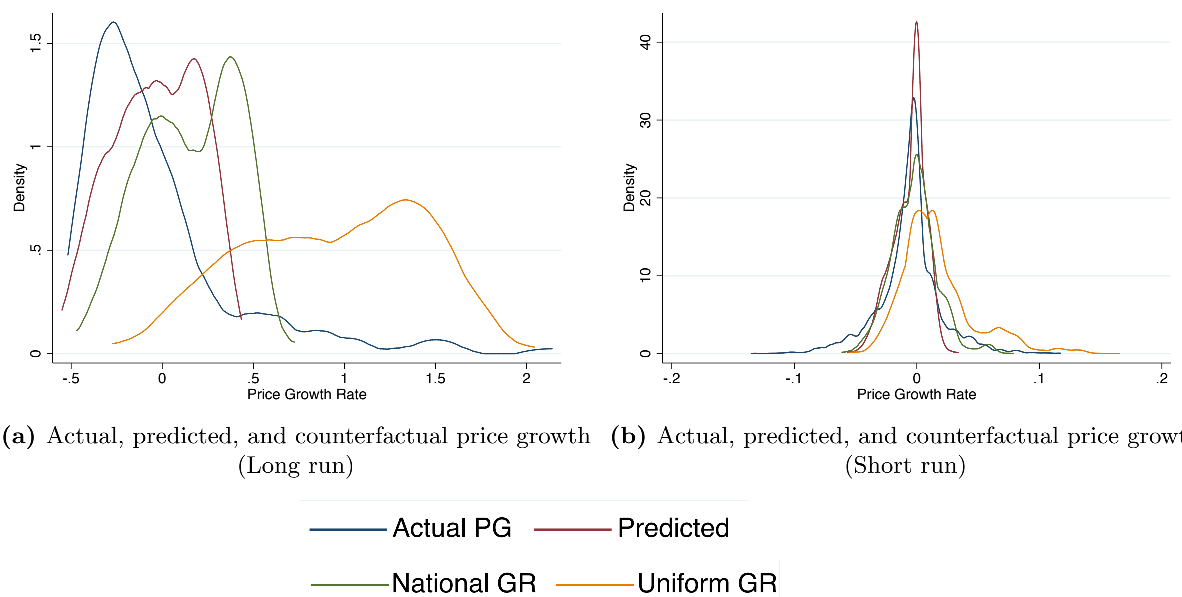
Notes: The above figure shows the first stays residuals from the regression of neighbourhood income growth on the spatial Bartik instrument. Dark blue denotes neighbourhoods for which the instrument under predicts actual income growth, whereas dark red denotes neighbourhoods whose actual income growth is over predicted.

Figure 2.A5: Changes to the Housing Supply Stock
(source: 1990 to 2010 census)



Notes: The above figure displays changes in the distribution of owner-occupied housing units and renter occupied housing units between 1990 and 2010 (as reported in the census). While the change in renter-occupied housing units is centred around zero, the change in owner-occupied housing units is on average greater than zero.

Figure 2.A6: Distribution of House Price Growth Rates
(1990 to 2010))



(a) Actual, predicted, and counterfactual price growth (Long run)

(b) Actual, predicted, and counterfactual price growth (Short run)

Notes: Actual price growth; predicted price growth from baseline specification; counterfactual price growth imposing national industry growth rates; counterfactual price growth imposing uniform growth rate across industries.

Figure 2.A7: Map of the Detroit MSA Neighbourhoods (zipcodes)

Figure 2.A8: Map of the Portland MSA Neighbourhoods (zipcodes)

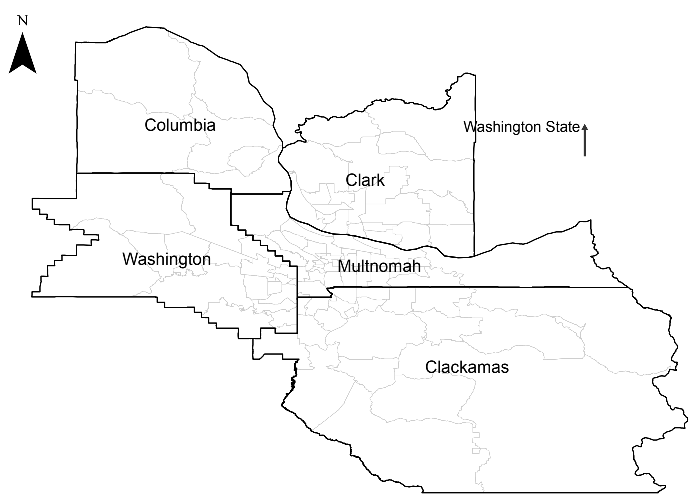
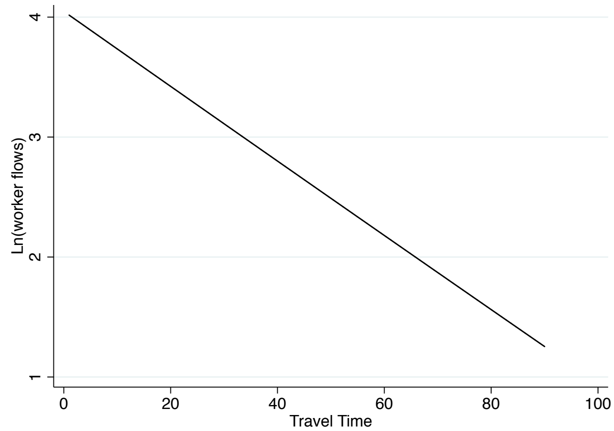


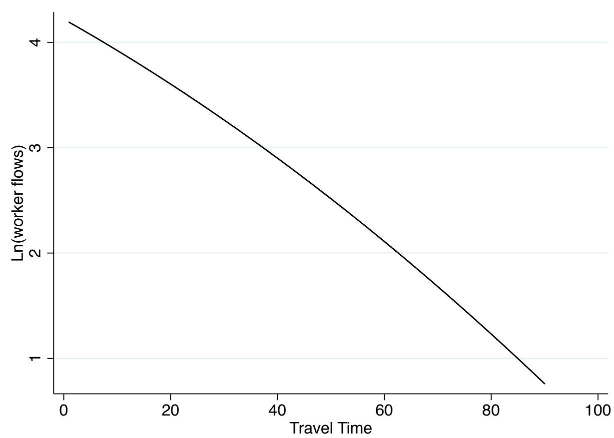
Figure 2.A9: Elasticity of worker flows with respect to commute time: Detroit, MI MSA



Census Transportation Data (CTPP), 1990		mean	SD	min	max
flows	total workers leaving home during peak travel time	39	93	0	2,853
commute	average commute time during peak travel time	32	14	1	90
κ	decay parameter (estimated)	-0.0386			

Source: 1990 Census Transportation Planning Package

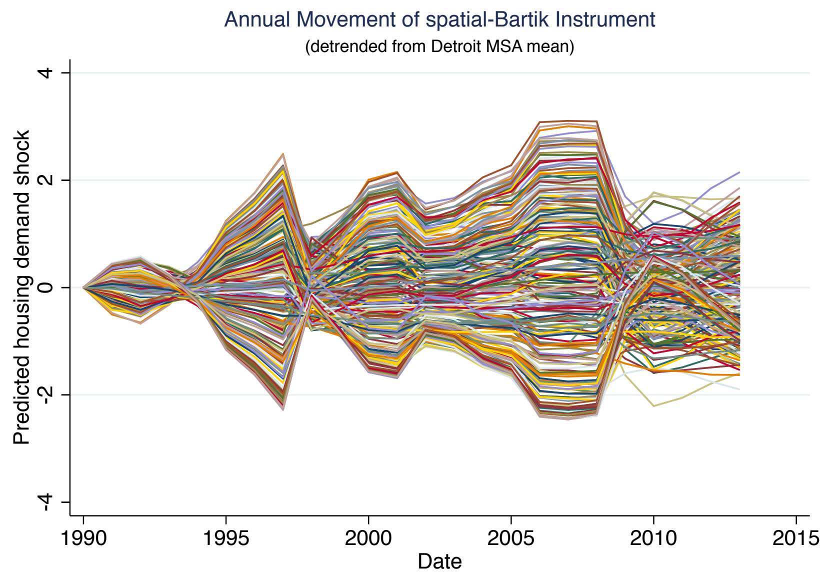
Figure 2.A10: Elasticity of worker flows with respect to commute time: Portland, OR MSA



Census Transportation Data (CTPP), 1990		mean	SD	min	max
flows	total workers leaving home during peak travel time	61	137	0	2,088
commute	average commute time during peak travel time	27	12	1	90
κ	decay parameter (estimated)	-0.0436			

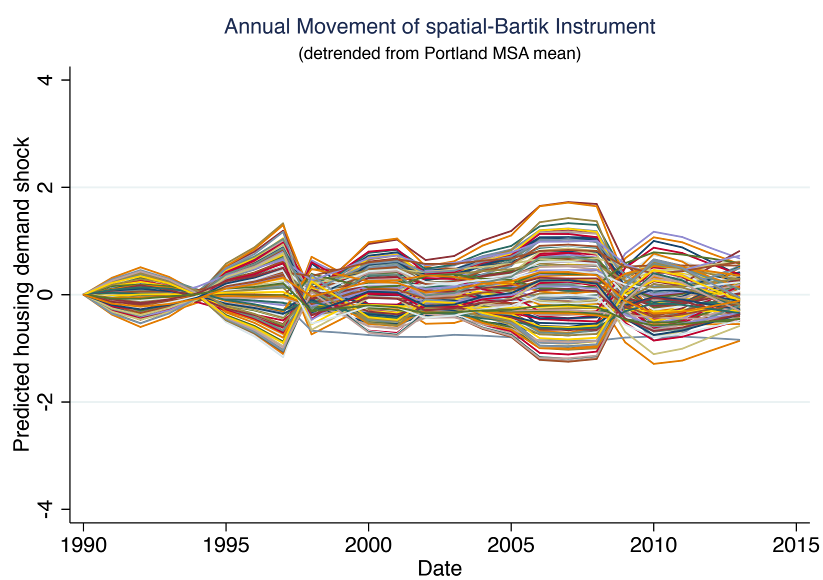
Source: 1990 Census Transportation Planning Package

Figure 2.A11: Annual Variability in Predicted Housing Demand Shocks: Detroit, MI MSA



Notes: The above figure shows the annual variation in the spatial-Bartik instrument constructed as a measure of neighbourhood income growth in the Detroit MSA. Each line depicts the instrument for a different neighbourhood.

Figure 2.A12: Annual Variability in Predicted Housing Demand Shocks: Portland, OR MSA



Notes: The above figure shows the annual variation in the spatial-Bartik instrument constructed as a measure of neighbourhood income growth in New York City. Each line depicts the instrument for a different neighbourhood.

Tables

Table 2.A1: Descriptive Statistics:
New York City Employment Data

	Mean	SD	Min	Max	Δln(NYC)	Δln(US)
Total employment:						
1990	19,411	41,674	824	333,995		
2013	21,864	47,818	1,320	402,121	0.17	0.24
Agriculture:						
1990	32	89	0	1001		
2013	0.62	4	0	42	-2.85	-1.61
Mining:						
1990	10	59	0	723		
2013	0.98	6	0	78	-2.12	-0.01
Utilities:						
1990	415	1,102	4	8303		
2013	1,035	3,560	1	31,250	0.05	1.03
Construction:						
1990	787	1,168	33	9,662		
2013	808	1,206	57	8,209	0.08	-0.06
Manufacturing (non-durable goods:)						
1990	1,291	3,013	0	23,290		
2013	268	544	0	4,424	-1.57	-0.34
Manufacturing (durable goods:)						
1990	754	1,195	0	7,552		
2013	186	347	0	2,359	-1.56	-0.29
Transportation:						
1990	806	1,557	5	11,237		
2013	494	747	10	4,599	-0.41	-0.34
Wholesale Trade:						
1990	1,363	4,028	9	43,039		
2013	1,089	3,317	2	33,293	-0.37	-0.17
Retail Trade:						
1990	2,668	4,335	111	39,825		
2013	2,216	3,390	226	26,870	-0.16	-0.38
Finance and Real Estate:						
1990	2,598	11,100	9	105,211		
2013	2,643	10,108	13	77,150	0.20	0.40
Business Services:						
1990	1,539	5,260	23	45,649		
2013	1,557	4,766	50	42,526	0.24	-0.17
Professional Services:						
1990	2,481	9,271	32	88,832		
2013	4,804	13,296	259	109,458	1.02	0.79
Personal Services:						
1990	711	1,949	0	17,166		
2013	2,602	5,662	144	39,714	1.36	1.02
Health Care:						
1990	2,012	3,289	26	19,723		
2013	2,418	2,514	56	14,647	0.41	0.82
Education:						
1990	1,509	2,023	61	11,663		
2013	980	1,651	19	9,710	-0.76	0.09
Arts and Entertainment:						
1990	431	1,391	0	13,002		
2013	760	2,274	0	18,621	0.24	-0.20

Notes: The above table presents descriptive statistics for changes in New York City and United States (excluding NYC) employment by industry. All numbers are changes in logged values.

Table 2.A2: Descriptive Statistics:
Spatial-Bartik Instrumental Variable

	1994		2004		2013	
own-Bartik (z_{it}):						
largest	Clifton-Fox Hills, SI	0.101	Battery Park, MN	0.332	Battery Park, MN	0.349
median	Cypress Hills, BK	0.032	Clinton, MN	0.082	Elmhurst, QN	0.089
smallest	Whitestone, QN	-0.032	Baisley Park, ON	-0.097	Rosedale, QN	-0.160
	mean:	**0.031**	**mean:**	**0.088**	**mean:**	**0.097**
	SD:	**0.021**	**SD:**	**0.083**	**SD:**	**0.106**
spatial-Bartik ($Z_{j,t}$)						
largest	Bay Ridge, BK	1.858	Bay Ridge, BK	5.002	Bay Ridge, BK	5.486
median	Homecrest, BK	1.034	Westchester, BX	2.840	Woodlawn, BX	3.070
smallest	Cambria Heights, QN	0.187	Arden Heights, SI	0.476	Cambria Heights, QN	0.397
	mean:	**1.041**	**mean:**	**2.840**	**mean:**	**3.066**
	SD:	**0.362**	**SD:**	**1.066**	**SD:**	**1.208**

Notes: The above table presents descriptive statistics of the spatial-Bartik instrument for predicted housing demand in a neighbourhood for 1994, 2004, and 2013.

Table 2.A3: Descriptive Statistics:
Price Growth and Income Data

Variable	mean	SD	min	max	N	*source*
$\%\Delta HPI_{1990-2010}$	1.965	1.061	0.192	6.412	155	*NYC Sales Data*
$\%\Delta$avg home value$_{1990-2010}$	1.707	1.261	0.453	7.375	155	*Census*
$\%\Delta$avg rent$_{1990-2010}$	0.192	0.124	-0.114	0.630	155	*Census*
$\%\Delta Income_{1990-2010}$	0.172	0.256	-0.320	1.280	155	*HMDA*
$\%\Delta$avg income$_{1990-2010}$	0.061	0.172	-0.230	0.945	155	*Census*
$\%\Delta HPI_{t,t-1}$	-0.006	0.026	-0.135	0.118	3,720	*NYC Sales Data*
$\%\Delta Income_{t,t-1}$	0.008	0.018	-0.070	0.106	3,565	*HMDA*
$\%\Delta Q_{totalunits,1990-2010}$	0.147	0.154	-0.084	1.222	155	*Census*
$\%\Delta$own occ. units$_{1990-2010}$	0.255	0.218	-0.097	1.392	155	*Census*
$\%\Delta$rent occ. units$_{1990-2010}$	0.041	0.178	-0.406	1.008	155	*Census*

Notes: The above table presents descriptive statistics for changes in neighbourhood house prices, average rents, and average incomes as reported in the NYC Sales Data, the HMDA data, or the Census Data. Changes are from 1990 to 2010.

Table 2.A4: Income Driven House Price Growth: Census comparison

(Long run: % Δ 1990 to 2010)

Panel A:	OLS			IV		
Dependent Variable:	*Price Growth*			*Price Growth*		
	(1) *(Table 2.1)*	$(2)^{\dagger}$	$(3)^{\ddagger}$	(4) *(Table 2.1)*	$(5)^{\dagger}$	$(6)^{\ddagger}$
Δln(Income)	0.844	4.110	2.385	3.790	7.636	10.010
	$(0.15)^{***}$	$(0.59)^{***}$	$(0.41)^{***}$	$(1.10)^{***}$	$(1.09)^{***}$	$(2.48)^{***}$
R^2	0.26	0.49	0.25			
F-stat:				11.46	29.88	14.23
N	155	155	155	155	155	155

Notes: *** statistically significant at the 1% level; robust standard errors in parenthesis.
† Decadal census income changes
‡ Decadal census average house prices; no quality controls
Columns (1) and (4) reproduce the baseline results from Table 2.1. Columns (2) and (5) use changes in average incomes from the census in place of the HMDA data, while columns (3) and (6) use changes in average house prices from the census as opposed to the NYC Sales data. While the magnitudes of the coefficients are larger in the census, the interpretations are quite similar (recall that the NYC Sales data has been detrended from the NYC mean).
Interpretation: (column 3) a 1σ increase in $\Delta ln(Income) \rightarrow 0.42\sigma$ increase in price growth;
(column 6): a 1σ increase in $Z_{j,t} \rightarrow 0.49\sigma$ increase in price growth.

Table 2.A5: Robustness Checks

Panel A:	IV			
Dependent Variable:	$PriceGrowth_{n,\Delta(1990to2010)}$			
	(1) Baseline *(Table 2.1)*	(2) w/out Finance	(3) w/out Midtown	(4) Manhatan FE
Δln(Income)	3.790	3.821	3.810	4.457
	(1.10)***	(1.16)***	(1.11)***	(2.22)**
Manhattan fixed effect				-0.256
				(0.58)
F-stat	11.46	10.66	11.10	3.11
N	155	155	155	155

Panel B:	IV			
Dependent Variable:	$PriceGrowth_{n,t}$			
	(1) Baseline *(Table 2.3)*	(2) w/out Finance	(3) w/out Midtown	(4) Manhattan FE
Δln(Income)	3.790	3.810	3.805	4.584
	(0.95)***	(1.01)***	(0.97)***	(1.85)**
Manhattan fixed effect				-0.020
				(0.03)
F-stat	14.24	12.92	13.68	5.15
N	3,565	3,565	3,565	3,565

Notes: *** statistically significant at the 1% level; standard errors in parenthesis and clustered at the neighbourhood level where appropriate. The dependent variable in Panel A is the price growth between 1990 and 2010; the dependent variable in Panel B is the annual price growth. Columns (1) reproduce the results from Tables 2.1 and 2.3; columns (2) exclude finance and real estate from the instrumental variable while columns (3) exclude Midtown (the CBD). Columns (4) include a dummy variable equal to 1 for neighbourhoods in Manhattan, and 0 otherwise. Results are robust across specifications.

Table 2.A6: Descriptive Statistics:
Detroit Employment Data

	Mean	SD	Min	Max	Δln(Detroit MSA)	Δln(US)
Total employment:						
1990	7,036	9,086	10	59,775		
2013	6,403	8,260	3	47,966	-0.23	0.24
Agriculture:						
1990	36	45	3	2,202		
2013	1	10	0	155	-1.87	-1.60
Mining:						
1990	2	6	0	307		
2013	2	6	0	41	-0.22	-0.01
Utilities:						
1990	148	433	0	5,240		
2013	183	665	0	9,070	0.05	1.04
Construction:						
1990	272	315	0	2,110		
2013	230	321	0	2,571	-0.35	-0.06
Manufacturing (non-durable goods:)						
1990	317	629	0	5,343		
2013	157	309	0	2,723	-0.64	-0.38
Manufacturing (durable goods:)						
1990	1,191	2077	0	12,683		
2013	615	1135	0	6,672	-0.75	-0.30
Transportation:						
1990	195	661	0	8,253		
2013	210	684	0	7,532	-0.09	-0.36
Wholesale Trade:						
1990	462	812	0	7,037		
2013	334	672	0	5,149	-0.45	-0.17
Retail Trade:						
1990	1495	1,890	0	11,477		
2013	819	1,109	0	7,575	-0.67	-0.37
Finance and Real Estate:						
1990	487	1,240	0	9,876		
2013	390	883	0	7,857	-0.15	0.38
Business Services:						
1990	692	1,372	0	9,012		
2013	623	1,159	9,673		-0.03	-0.19
Professional Services:						
1990	697	1,231	0	9,363		
2013	1,155	2,071	0	18,248	0.40	0.79
Personal Services:						
1990	155	245	0	1,596		
2013	727	908	0	6,762	1.83	1.03
Health Care:						
1990	672	1,237	0	10,114		
2013	737	1,295	0	11,435	0.12	0.83
Education:						
1990	74	204	0	1,518		
2013	111	243	0	2,400	0.65	0.07
Arts and Entertainment:						
1990	107	174	0	1,403		
2013	107	266	0	2,333	0.65	-0.19

Notes: The above table presents descriptive statistics for changes in the Detroit MSA and United States (excluding Detroit) employment by industry. All numbers are changes in logged values.

Table 2.A7: Descriptive Statistics:
Portland Employment Data

	Mean	SD	Min	Max	Δln(Portland MSA)	Δln(US)
Total employment:						
1990	5,216	7,218	3	39,068		
2013	6,448	8,172	3	37,348	0.13	0.23
Agriculture:						
1990	38	49	0	258		
2013	13	41	0	420	-1.17	-1.62
Mining:						
1990	5	16	0	115		
2013	3	9	0	66	-0.62	-0.01
Utilities:						
1990	115	389	0	3,817		
2013	224	529	0	2,964	0.66	1.04
Construction:						
1990	211	363	0	1,768		
2013	377	503	0	3,338	-0.04	-0.07
Manufacturing (non-durable goods:)						
1990	303	503	0	2,250		
2013	198	362	0	2,433	-0.28	-0.38
Manufacturing (durable goods:)						
1990	616	951	0	4,942		
2013	479	947	0	6,392	-0.37	-0.30
Transportation:						
1990	222	506	0	3,523		
2013	242	590	0	4,005	-0.09	-0.36
Wholesale Trade:						
1990	465	964	0	6,677		
2013	426	811	0	5,476	-0.15	-0.18
Retail Trade:						
1990	1079	1,666	0	10,847		
2013	785	1,119	0	6,179	-0.30	-0.38
Finance and Real Estate:						
1990	456	1212	0	10,819		
2013	425	876	0	6,375	0.12	0.38
Business Services:						
1990	433	779	0	4,354		
2013	524	773	0	4,478	0.39	-0.20
Professional Services:						
1990	483	857	0	5,898		
2013	1,144	1,741	0	9,186	0.64	0.78
Personal Services:						
1990	118	235	0	1,412		
2013	733	958	0	4,155	2.04	1.03
Health Care:						
1990	402	702	0	3,303		
2013	573	1,008	0	4,970	0.45	0.82
Education:						
1990	82	209	0	1,414		
2013	178	338	0	1,856	1.20	0.07
Arts and Entertainment:						
1990	78	127	0	843		
2013	121	200	0	1,342	0.52	-0.20

Notes: The above table presents descriptive statistics for changes in the Portland MSA and United States (excluding Portland) employment by industry. All numbers are changes in logged values.

Appendix 2.B

Theory

The following theory has been adapted from Moretti (2013) to examine the spatial implications within a city, following skill-biased technological change. While Moretti (2013) focused on movements between cities, I am considering movements within cities. As such, the cost of commuting between the suburbs and the downtown will be part of an individual's utility function.

For simplicity, consider two skill groups $s \in \{h, l\}$, and two locations $c, n \in \{m, q\}$ where c denotes work location, and n denotes home location. We can think of m as representing Manhattan (i.e. downtown) and q as representing Queens (i.e. the suburbs). I define the indirect utility for worker i as:

$$V_{i,s,c,n} = V(W_{sc}, C_{icn}, P_n, T_{isn}) \tag{2.15}$$

$$= W_{sc} - C_{cn} - P_n + T_{sn} \tag{2.16}$$

Assume that utility is increasing in wages (W_s) and locational preferences (T_s) and decreasing in commute time (C) and price of housing (P). Wages and preferences differ by skill group (s), whereas commute costs and housing costs do not.[14] However, commute costs and locational preferences represent idiosyncratic preferences that differ by individual.

$$C_{icn} \sim \mathcal{U}[-c, c] \tag{2.17}$$

$$T_{isn} \sim \mathcal{U}[-t_s, t_s] \tag{2.18}$$

$$\tag{2.19}$$

Workers of each skill group s choose:

$$V_{n,c} = max\{V_{mm}, V_{mq}, V_{qq}, V_{qm}\} \tag{2.20}$$

And, the optimality condition requires that:

$$V(W_{c^*}, C_{c^*n^*}, P_{n^*}, T_{n^*}) \geq V(W_c, C_{cn}, P_n, T_n) \qquad \forall c \neq c^*, n \neq n^* \tag{2.21}$$

The above implies that for an individual to choose a commuting job over a local job, or a more expensive residential location, they must be receiving a wage premium in their work location. Next, I will discuss some scenarios that will determine the equilibrium distribution of workers between Manhattan and Queens. Following this, I will consider a skill-biased technological shock that increases the wages for skilled workers working in Manhattan and the redistributional implications of this.

[14] Allowing commute tolerance to differ by skill group, or to be a function of wage, will not dramatically change the intuition of the results. However, I will discuss this further below.

Assumptions:
Manhattan is more expensive than Queens: $P_m > P_q$
Wages in Manhattan are higher than in Queens, for both skill groups: $W_{hm} > W_{hq}; \quad W_{lm} > W_{lq}$
Commute costs are symmetric: $C_{qm} = C_{mq} > C_{mm} = C_{qq}$

A worker will choose to live in Queens and work in Manhattan, over living and working in Manhattan iff:

$$V(W_{sm}, C_{imq}, P_q, T_{isq}) \geq V(W_{sm}, C_{imm}, P_m, T_{ism})$$

$$\begin{aligned} \Rightarrow W_{sm} - C_{imq} - P_q + T_{isq} &\geq W_{sm} - C_{imm} - P_m + T_{ism} \\ P_m - P_q &\geq (C_{imq} - C_{imm}) + (T_{ism} - T_{isq}) \end{aligned}$$

Assuming commute costs are fixed, $C_{mm} < C_{mq}$, and $P_m > P_q$, the house price differential between Manhattan and Queens must be high enough to compensate for the cost of commuting plus any preference for living in Manhattan over Queens. The marginal worker must be indifferent between the two.

A worker will choose to live in Queens and work in Manhattan, over living and working in Queens iff:

$$V(W_{sm}, C_{imq}, P_q, T_{isq}) \geq V(W_{sq}, C_{iqq}, P_q, T_{isq})$$

$$\begin{aligned} \Rightarrow W_{sm} - C_{imq} - P_q + T_{isq} &\geq W_{sq} - C_{iqq} - P_q + T_{isq} \\ W_{sm} - W_{sq} &\geq C_{imq} - C_{iqq} \end{aligned}$$

For this to hold, the wage premium in Manhattan must compensate for the commute costs.

A worker will choose to live in Manhattan and work in Queens, over living and working in Queens iff:[15]

$$V(W_{sq}, C_{iqm}, P_m, T_{ism}) \geq V(W_{sq}, C_{iqq}, P_q, T_{isq})$$

$$\begin{aligned} \Rightarrow W_{sq} - C_{iqm} - P_m + T_{ism} &\geq W_{sq} - C_{iqq} - P_q + T_{isq} \\ T_{ism} - T_{isq} &\geq (P_m - P_q) + (C_{iqm} - C_{iqq}) \end{aligned}$$

If $C_{qq} < C_{qm}$ and $P_m > P_q$, idiosyncratic preferences for living in Manhattan must be high enough to overcome the higher cost of housing in Manhattan and the commute costs or these workers must make $E(W_{sm}) + C$ in Queens to be enticed to commute from Manhattan to Queens.[16]

[15] Note: a worker will never choose to live in Manhattan and work in Queens, over living and working in Manhattan.

[16] As discussed in So, Orazem, and Otto (2001), if on average $W_m > W_q$, there will be relatively few jobs in Queens that would entice someone to commute from Manhattan. They find this to be true in the data.

A worker will choose to live and work in Manhattan, over living and working in Queens iff:

$$V(W_{sm}, C_{imm}, P_m, T_{ism}) \geq V(W_{sq}, C_{iqq}, P_q, T_{isq})$$

$$W_{sm} - C_{imm} - P_m + T_{ism} \geq W_{sq} - C_{iqq} - P_q + T_{isq}$$
$$W_{sm} - W_{sq} \geq (P_m - P_q) - (T_{ism} - T_{isq})$$

Assuming $C_{mm} = C_{qq}$, this group needs to be compensated for higher cost of living, with higher wages in MN. As in all of the above the marginal worker must be indifferent between the two while the inframarginal workers are able to earn economic rents given their idiosyncratic preferences.

Theoretical Predictions:

- A worker will *commute* to Manhattan if the differential cost of living between Manhattan and Queens exceeds the cost of commuting and any preference to live in Manhattan over Queens; or, the wage differential between Manhattan and Queens is greater than the cost of commuting
- A worker will *commute* to Queens only if the preference for living in Manhattan is high enough to compensate for the higher cost of living and the commute cost
- A worker will live and work in Manhattan if the wage differential compensates for the higher cost of living and any preference for living in Queens over Manhattan
- In equilibrium, the marginal worker needs to be indifferent between each of the above:

$$\text{If} \quad C_{mq} = C_{qm}, \quad \text{then in equilibrium}$$
$$W_{sm} = W_{sq} - (P_m - P_q) + (T_{sm} - T_{sq}) \tag{2.22}$$

Firms production function and wages:

As in Moretti (2013), high skilled and low skilled workers are employed by different firms. Firms employing labour type $s \in \{h, l\}$, operating in location $c \in \{m, q\}$ have a Cobb-Douglas production function and constant returns to scale:

$$y_{sc} = X_{sc} N_{sc}^{h} K_{sc}^{1-h} \tag{2.23}$$

Firms are price takers and labor is paid its marginal product:

$$MPL = \frac{\partial y_{sc}}{\partial N_{sc}} = X_{sc} h N_{sc}^{h-1} K_{sc}^{1-h} \tag{2.24}$$

$$\therefore \tilde{w}_{sc} = \tilde{X}_{sc} - (1-h)\tilde{N}_{sc} + (1-h)\tilde{K}_{sc} + ln(h) \tag{2.25}$$

where $\tilde{Var}$ denotes logged values.

Equilibrium in the labour market:

$$W_{sm} = W_{sq} - P_m + P_q + T_{sm} - T_{sq} \tag{2.26}$$
$$= X_{sm} - (1-h)N_{sm} + (1-h)K_{sm} + ln(h) \tag{2.27}$$

subbing in for W_{sq} by symmetry, and solving for $N_{sm} = N_s - N_{sq}$:

$$N_{sm} = \frac{N_s + K_{sm} - K_{sq}}{2} + \frac{1}{2(1-h)}[(X_{sm} - X_{sq}) - (P_m - P_q) + (T_{sm} - T_{sq})] \qquad \forall s \in \{h, l\} \tag{2.28}$$

and by symmetry,

$$N_{sq} = \frac{N_s + K_{sq} - K_{sm}}{2} + \frac{1}{2(1-h)}[(X_{sq} - X_{sm}) - (P_q - P_m) + (T_{sq} - T_{sm})] \qquad \forall s \in \{h, l\} \tag{2.29}$$
$$\tag{2.30}$$

Housing market:
Each worker consumes one unit of housing such that the demand for housing equals the supply of skilled and unskilled workers in a location. Without formally modelling the housing market, assume that the supply of housing is determined by:

$$P_m = z + \kappa_m N_m \tag{2.31}$$
$$P_q = z + \kappa_q N_q \tag{2.32}$$

For $n \in \{m, q\}$, all worker types pay the same price for housing (r_n); the number of housing units in the neighbourhood equals the number of workers ($N_n = N_{hn} + N_{ln}$); and κ_n is the elasticity of housing supply.

Skill-biased technological change:
Suppose that there is an increase in the productivity of skilled workers who work in Manhattan.[17]

Assume that $X_{hm2} = X_{hm1} + \Delta$, where $\Delta > 0$ is the productivity shock to industry that employs high-skilled labour in Manhattan. Given the C-D production function, nominal wages of skilled labour will increase by Δ/h, where h is the C-D return to labour. When nominal wages increase, the supply of skilled workers in Manhattan will increase as some workers will now find it optimal to live in Manhattan.

This increase in the supply of labour will increase the demand for housing by high skilled workers which will in turn cause P_m to increase by $\Delta N_m \kappa_m \geq 0$ Low-skilled workers living in Manhattan now face a real-wage decrease as their nominal wages are unchanged but their cost of housing has increased. This increase in the cost of housing will cause some low-skilled workers to demand housing outside of Manhattan (i.e. to move to Queens). The fraction of movements for high skilled and low skilled workers is determined by their tolerance for commuting and locational preferences.

[17]While firms employing skilled workers in Queens will also be affected by a productivity shock because of the relative density in Manhattan, I assume that firms located close to the city center are disproportionately affected by this positive shock.

Equilibrium adjustments following a productivity shock to high-skilled workers, employed in Manhattan:

Who will be effected?

(1) Some high income workers living and working in Queens will begin commuting to Manhattan:

$$\begin{aligned} &\text{In period 1:} \quad W_{hm1} - W_{hq1} = C_{mq} - C_{qq} \\ &\text{In period 2:} \quad W_{hm2} - W_{hq2} \geq C_{mq} - C_{qq} \end{aligned}$$

(2) Some high income workers living and working in Queens will move to Manhattan:

$$\begin{aligned} &\text{In period 1:} \quad W_{hm1} - W_{hq1} = (P_{m1} - P_{q1}) - (T_{sm} - T_{hq}) \\ &\text{In period 2:} \quad W_{hm2} - W_{hq2} \geq (P_{m2} - P_{q2}) - (T_{sm} - T_{hq}) \end{aligned}$$

When wages in Manhattan increase, there will be an inflow of high-skilled workers into Manhattan. This increases the demand for housing in Manhattan and causes prices to rise (by an amount dependent on the elasticity of housing). Workers will continue to move until the above equation once again equates for the marginal worker.[18] For inframarginal workers, if their preference for living in Queens over Manhattan is large enough, provided that the wage premium exceeds the cost of commuting, they may choose to commute.

(3) Some low incomes workers living and working in Manhattan move to Queens:

$$\begin{aligned} &\text{In period 1:} \quad W_{lm1} - P_{m1} = (W_{lq1} - P_{q1}) - (T_{lm} - T_{lq}) \\ &\text{In period 2:} \quad W_{lm2} - P_{m2} \leq (W_{lq2} - P_{q2}) - (T_{lm} - T_{lq}) \end{aligned}$$

(4) For some of the low income workers who move to Queens, they will commute to work in Manhattan if:

$$\begin{aligned} \text{Assume:} \quad & W_{lm1} = W_{lm2} \\ \therefore & W_{lm1} - C_{qm} - P_{q1} + T_{lq} \geq W_{lq1} - C_{qq} - P_{q1} - T_{lq} \\ & W_{lm} - W_{lq} \geq C_{qm} - C_{qq} \end{aligned}$$

When P_m increases, the real wage of low income workers in Manhattan will decrease. As a result low-income workers will relocate to Queens conditional on their preference for living in Manhattan. But, these workers do not necessarily choose to work in Queens. Now the low income worker can choose to live and work in Queens, or live in Queens and commute to Manhattan if the wage differential is such that it compensates for the cost of commuting. It must be that the preference for living in Manhattan is still sufficiently high to compensate for both the higher cost of housing and the commute time.

[18]Note: Moretti (2013) shows that the real wage differential increases for high skilled workers – the increase in nominal wages is going to be larger than the increase in house prices.

Chapter 3

The Built Environment and Social Interactions

3.1 Introduction

How the features of one's neighbourhood can be conducive to social interactions and foster social engagement has long been a question of theorists, activists, and policy makers alike. Some of the more prominent discussions around the subject have come from Jacobs (1961) who criticized urban planning practices as not catering to the needs of the residents, Putnam (2000) who criticized sprawl for destroying civic engagement, and Olson (1965) who saw high-density neighbourhoods as contributing to free-rider problems in the provision of public goods. Lacking conclusive evidence, the optimal design of neighbourhoods is still the subject of debate. Jacobs in particular emphasized the importance of neighbourhoods in generating interactions between residents from different socio-economic backgrounds.

The belief in a causal relationship between the built environment and positive social outcomes has been encapsulated in the New Urbanist approach to urban planning and has influenced land-use policies and real-estate development. As a reaction to Modernism, city planners all over the world are redeveloping and redesigning neighbourhoods to increase green-space and walkability with the intention of increasing community involvement and neighbourhood safety - a Jacobs-esque idea of getting people out of their houses and onto the streets. While this is undoubtedly a desired outcome, this chapter questions whether having these new public spaces actually *causes* people to interact more and subsequently become more involved in their community.

Previous literature has often concluded that because more walkable communities have more social interactions on average, these walkable communities cause more social interactions.[1] While I don't question whether walkable communities facilitate social interactions, I question whether they are changing who interacts and how as opposed to whether they interact at all or how often; in other words, whether they are changing people, or changing peoples' location decisions. It is important to recognize that the

[1] Previous cross-sectional studies include, French et al. (2013), Leyden (2003), Maas et al. (2006), and Wood et al. (2008).

location decisions of individuals are not random and may be driven by unobservable propensities to be around a certain group of people or to engage in particular behaviours. As such, previous estimates of the social interaction effects attributed to environmental characteristics likely suffer from endogeneity due to a sorting of social people into socially conducive neighbourhoods. Thus far only a handful of papers have addressed this issue of endogeneity.

Brueckner and Largey (2008) look at the relationship between social interactions and population density. Using density at the MSA level as an instrument for density at the census tract level, they estimate a negative effect of urban density on social interactions. This result contradicts Putnam's critique of suburbs as destroying social capital relationships. In similar work, Borck (2007) looks at the relationship between city size and social interactions, consumption opportunities, and group memberships. Using lagged population density as an instrument for current population density, results are largely ambiguous with some evidence of consumption externalities. Glaeser and Gottlieb (2006) also consider the impact of density on amenities, as well as civic engagement and political involvement. They conclude that although density is correlated with consumer amenities, there is no apparent relationship with civic engagement. This result is similar to Borck (2007) and is also in contradiction of Putnam's theory that urban sprawl destroys social capital relationships.

Using an extensive panel dataset to address issues of endogeneity, I look at the relationship between social interactions and access to amenities (i.e. neighbourhood walkability) as well as the relationship between own sociability and average neighbourhood sociability. This differs from previous studies which primarily explore the relationship between social interactions and density. As such, I question the work of Jane Jacobs regarding the importance of neighbourhood walkability for interactions.

This chapter also relates to a literature on neighbourhood effects which considers the impact of spatial sorting on either labour market outcomes (Andersson et al. (2016); De la Roca and Puga (2017)) or improved mental health (Katz et al. (2001); Ludwig et al. (2001)). This second stream of literature largely attributes the positive effects of moving from a poor to a less poor neighbourhood to decreased violence and increased safety on the streets. Furthermore, a literature on social networks links social relationships with improved health outcomes by facilitating access to resources and information (Hawkley and Cacioppo (2010); Fiorillo and Sabatini (2011)), as well as with neighbourhood resiliency in the wake of natural disasters (Sampson (2011)).[2] Jane Jacobs would, at least partially, credit this impact to the built environment; stores, restaurants, schools, and parks (among others) that encourage street interactions and will in turn make people feel more connected with those around them.

To explore the relationship between environmental characteristics and social interactions, this chapter first estimates the cross-sectional relationship but then addresses endogeneity using a first-differences econometric specification that focuses on how changes to neighbourhood characteristics will affect social interactions. First-differenced equations are estimated for three distinct subsets of the population; those who have not moved over the sample period ('stayers') which provides time series variation in the built environment over time, and movers within and between counties which provides variation to the built environment following the move. The environmental characteristics of interest are: access to various neighbourhood common spaces within a one kilometre radius of one's home (referred to as the

[2] Refer to the chapter by Durlauf and Fafchamps (2005) for a summary of the social capital literature. Also, Blume et al. (2010).

"walkability index") and the average county sociability which is measured as the portion of a county falling in either the top 30 percent or the bottom 30 percent of the entire country's social interactions distribution.[3]

I find a strong and positive cross-sectional relationship that is consistent with the work of Jane Jacobs and with previous literature. However, after controlling for observable characteristics and time-invariant unobservables, the physical built environment appears to have no causal impact on social interactions. Interestingly, the *social* environment remains a significant correlate of own sociability. These results are consistent with sorting behaviour; it appears that individuals are sorting into bigger neighbourhoods in terms of density and access to amenities due to unobservable characteristics (e.g. propensities to be social) and that this sorting is generating the cross-sectional relationship. This provides suggestive evidence that rather than changing the social behaviours of residents, walkable neighbourhoods are changing the composition of residents.

The remainder of the chapter is organized as follows: Section 3.2 discusses the dataset and the creation of the social interaction variables of interest. Section 3.3 presents the empirical framework and Section 3.4 then presents the cross-sectional and first differenced results along with various robustness and specification tests. Finally, Section 3.5 concludes.

3.2 Data and Descriptive Statistics

The German Socio-Economic Panel: In order to estimate the effect of the built environment on social interactions I construct a data set from the German Socio-Economic Panel (SOEP). The SOEP is conducted by the German Institute for Economic Research (DIW Berlin). The panel begins in 1984 in West Germany (FRG) and is conducted annually with participation from households and individuals. As early as June 1990, the SOEP expanded to include the states from the former East Germany (GDR).[4] Through remote access to DIW I have information on county of residence which I merge to population data from the Federal Statistical Office.[5] I use this to calculate county densities in 1995, 2001, 2005, and 2011. Table 3.A1 of the Appendix provides descriptive statistics for all variables used throughout this chapter. The final dataset spans 1994 to 2009 and consists of 145,455 observations for 25,806 unique individuals. Care has been taken by DIW to ensure a weighted representation of each county, proportional to its size. In the dataset there is an average movement rate within counties of eight percent per year and between counties of two percent per year. Six percent of the sample is included in all 15 years of the study and the average number of years of participation is seven.

Measurement of the Built Environment

My primary measure of the built environment comes from household access to common spaces, or the

[3] Throughout I also control for county density given the strong theoretical relationship between density and interactions. However, because it is at the county level rather than the town or neighbourhood level, I refrain from making causal statements.

[4] For the purposes of this chapter, I focus solely on those years following reunification and for which questions regarding household neighbourhood characteristics were asked (1994 to 2009). As a robustness check, I do exclude East Germany due to the possibility of differential perspectives on social engagement as compared to the West.

[5] There are currently 439 counties in Germany with 438 of them included in the dataset as of 2004. The average size of a county is 808 km^2, with an average density of $539/km^2$. Federal statistical office: https://www.destatis.de/EN/Homepage.html, and http://www.citypopulation.de

walkability of one's neighbourhood. In 1994, 1999, 2004, and 2009, SOEP conducted an extended household survey that focused on neighbourhood characteristics. They asked how long it takes to walk to each of the following: shops, restaurants, family doctor, kindergarten school, primary school, youth center, old age home, park, sports center, and public transit. The respondent can answer less than 10 minutes, 10 to 20 minutes, greater than 20 minutes, or can not be walked to. My walkability index consists of the total amenities an individual can walk to within 10 minutes (approximately 0.8 kms); this index lies between zero and 10 for each respondent. Less than nine percent of the sample has all amenities within a 10 minute walk from home, and approximately 30 percent has all amenities within a 20 minute walk. Table 3.A2 includes the time series variation in the environmental characteristics for those categorized as movers and stayers over the sample period; the values reported are the average at the county level. The built environment changes slower for stayers than for movers, and slower for movers within counties than for movers between counties. For those who move, I exploit these changes as a shock to the built environment and for those who do not move, I exploit the exogenous changes to the environment around them. The variability in access to these common spaces following a move, as well as the change in access over time for stayers, provides time-series variation in the built environment.

Measurement of Social Capital

In order to carry out my analysis of the social interaction effects of one's environment, I require information on interpersonal relationships as well as information on the frequency of various activities. In each wave of the survey individual respondents are asked how often they participate or engage in various activities with neighbours and family, groups and associations, as well as how often they go to the cinema, eat at restaurants, or utilize various other urban amenities. Fukuyama (1995) offered three broad categories of social capital that are used in empirical works: voluntary community association activity, trust and informal cooperation, and quality of family relations. Consistent with this and previous literature, I categorize the questions on social interactions in the SOEP into four categories: (i) group involvement – annual participation in local politics, volunteer work, attendance at church, and participation in sports; (ii) familial relations – how often do you visit your family; (iii) neighbour interactions – how often do you visit your neighbours; and (iv) general community involvement – how often do you go out for dinner, to the cinema and concerts, to cultural events, and attend social gatherings.[6] From these survey responses I create four social interaction indices to be used as my dependent variables. Table 3.A3 of the Appendix shows the variation in social interactions across respondents; movers (both within and between counties) show more variation in their yearly measures of social interactions than do stayers.

Measurement of Average County Sociability

I also consider the impact of the average sociability of one's neighbourhood on own sociability. As per Galster et al. (2008), to calculate the county sociability mix, I specify the proportion of a county's population that falls into either the top 30 percent (''% in the bottom third') or bottom 30 percent ('%

[6]In the paper by Brueckner and Largey (2008), they divide social interactions into those from neighbourhood contacts and friendship and those from group involvement. They use the following measures of an individuals' neighbourhood contacts and friendships: how often the respondent socializes with neighbours, the number of people the respondent can confide in, the number of close friends, the frequency of socializing with friends in a public place, and the frequency with which friends are invited to the respondent's house. To measure the respondent's group involvement they use: whether the respondent works with neighbours to make neighbourhood improvements, whether they are a member of a hobby-oriented club, the frequency of attendance at club meetings, and the number of non-church groups to which the member belongs. In the paper by Borck (2007), he focuses on questions around trust, close friendships, attitudes towards crime, as well as memberships in unions, professional bodies, staff councils, environmental organizations, and other club0type organizations.

in the top third') of the distribution; the middle 40 percent is the omitted category. The higher the percentage of the county's population falling in the top 30 percent compared to the average, the more relatively socially engaged is the county.[7] Controlling for this average county sociability should capture most unobservable time-varying county characteristics.

3.3 Empirical Strategy

Helsley and Zenou (2011) develop a theory for social interactions in cities considering the networks of the interacting individuals. When agents can choose their location, there is a tendency for those more central in the network to locate more centrally in the city. This would imply an endogenous relationship between location and social interactions. If we allow for peoples' unobservable characteristics to influence the marginal benefit derived form social interactions, the effects estimated cross-sectionally are biased. In estimating the relationship between each type of social interactions and environmental characteristics there is the possibility that an individual's unobserved characteristics, among which is his unobservable propensity to engage in social activities, are driving some of the results.

Ideally I would like to estimate the following equation for individual i in county c at time t:

$$\begin{aligned} SI_{ict} &= \alpha + \beta_1 A_{ct} + \beta_3 R_{ct} + \beta_2 D_{ct} + \Gamma X_{it} + \epsilon_{ict}, \qquad t = 1, ..., T \\ \text{and,} \quad \epsilon_{ict} &= \delta_i + \delta_{it} + \eta_c + \eta_{ct} + \mu_{ict} \end{aligned} \tag{3.1}$$

where, A_{ct} represents neighbourhood walkability and R_{ct} represents relative sociability of the county that individual i resides in at time t. D_{ct} controls for county density, X_{it} is a vector of observable individual characteristics, and ϵ_{ict} is the error term comprised of unobservable time-invariant and time-variant individual characteristics (δ_i and δ_{it}, respectively), unobservable time-invariant and time-variant county characteristics (η_c and η_{ct}, respectively), as well as an idiosyncratic error term (μ_{ict}). My social interactions variable (SI_{ict}) takes one of four values: (i) interactions with groups or associations; (ii) interactions with family members; (iii) interactions with neighbours; and (iv) community involvement.

In a cross-sectional estimation, δ_i, δ_{it}, η_c, and η_{ct} are assumed to be independent of the regressors. However, by omitting them from the estimation equation the coefficient on location characteristics will be biased if either, individuals with a propensity to be social locate in high-density neighbourhoods or within walking distance to a number of amenities, or, if there are unobservable characteristics of a county that affect both social interactions and access to amenities.

I address this bias by using a first-differencing specification that focuses first on the time-series variation in the environment of those who do not move over the sample period ('stayers'), then on both movers within counties and movers between counties.[8]

[7] Average neighbourhood sociability is calculated both with and without respondent i. Results presented in this chapter exclude respondent i.

[8] This approach has been applied in other contexts, see Eid et al. (2008) and Galster et al. (2008) but to the best of my knowledge has been absent from the literature looking at determinants of social interactions. Borck (2007) includes estimates using fixed-effects however, as previously mentioned, he explores a different research question.

3.3.1 Addressing Endogeneity

Consider an individual in periods t and $t-1$. By differencing 3.1 with respect to time, I obtain:

$$\begin{aligned}\mathrm{SI}_{ict} - \mathrm{SI}_{ict-1} &= \alpha - \alpha + \beta_1(A_{ct} - A_{ct-1}) + \beta_2(R_{ct} - R_{ct-1}) + \beta_3(D_{ct} - D_{ct-1}) \\ &+ \Gamma(X_{it} - X_{it-1}) + \epsilon_{ict} - \epsilon_{ict-1}\end{aligned} \tag{3.2}$$

or,

$$\begin{aligned}\Delta\mathrm{SI}_{ict} &= \beta_1\Delta A_{ct} + \beta_2\Delta R_{ct} + \beta_3\Delta D_{ct} + \Gamma\Delta X_{it} + \Delta\epsilon_{ict}, \qquad t = 2, ..., T \\ \text{and,} \quad \Delta\epsilon_{ict} &= \Delta\delta_{it} + \Delta\eta_{ct} + \Delta\mu_{ict}\end{aligned} \tag{3.3}$$

where Δ denotes the time difference operator.

Note that differencing removes time invariant characteristics that are both observable and unobservable. For those who move between periods $t-1$ and t, the variation in their neighbourhood characteristics following a move can be used to analyze the impact of the environment on social engagement.[9]

As can be seen from Equation 3.3 there could still be unobservable time-varying characteristics that I am not controlling for. By exploiting the time series variation in stayers' neighbourhood characteristics, given these individuals do not move, changes to county density and average neighbourhood sociability should be exogenous and uncorrelated with any unobserved time-varying individual characteristics. Furthermore, controlling for average county sociability should be capturing unobservables that are correlated at the county level with both the built environment and social interactions. Because environmental changes occur slowly, for the 'stayers' I look at environmental changes over both a five year and a ten year period.

For 'movers' (both within and between counties), I use changes in the social interactions one year after moving compared to one year prior to moving as well as five years after moving compared to one year prior. For example, if the move occurred at time t, I look at the difference between t+1 and t-1. As has previously been mentioned, the primary difficulty with estimating the environmental impact is the fact that individuals choose where to live and how many amenities to "consume" based on unobservable individual characteristics. If cities are full of people with high social capital, or high social propensities, we should see this in the sorting of highly social individuals into cities. In other words, residents of high-density neighbourhoods should be consuming higher levels of social interactions.

As I will show in the cross-sectional results, there is a strong positive correlation between an individual's social interactions and the county density. Therefore, individuals within high density neighbourhoods do on average have higher levels of annual interactions. If there is sorting behaviour of highly social people into higher-density neighbourhoods, I should find fairly insignificant results in my first-differenced specifications.

[9] An attractive feature of the differencing approach is that it removes the assumption of strict exogeneity of the error term; $E[\epsilon_{it}|\alpha, X_i, X_{it}] = 0$. Weak exogeneity simply requires, $E[\Delta X'_{it}\Delta\epsilon_{it}] = 0$ (Wooldridge, *Econometric Analysis of Cross-Section and Panel Data*, Second Edition, 2010). Similarly, in the presence of heteroskedasticity or serial correlation of the error term, first differencing is efficient. In the following estimations, standard errors are clustered at the county level to control for the possibility of serial correlation in the errors.

3.4 Results

3.4.1 Baseline OLS Results

The baseline OLS results from the pooled regression for all years and all individuals are presented in Table 3.1. I present the results for the indices of social interactions: interactions with groups (*Group A*), family (*Group B*), neighbours (*Group C*), and the general community (*Group D*). Table 3.2 presents a partial decomposition of these results into dependent variables of particular interest (i.e. local political involvement, volunteer work, time spent with neighbours, and time spent attending social events) and neighbourhood characteristics of interest (i.e. shops, primary schools, and parks). I also include the results using 'distance to' — which ranges from a 10 minute walk to inaccessible (approx. 0.8kms to beyond 1.6kms) — as opposed to the dummy amenity variable. The full decomposition for all possible combinations of interactions and environmental characteristics is presented in Tables 3.A4 and 3.A5 Appendix 3A.

The cross-sectional results for the built environment's impact on all measures of social interactions are overwhelmingly significant and positive. The significant and positive relationship remains even after controlling for socio-demographic characteristics and regional controls.[10] This is consistent with the previous literature that has looked at the correlation between the environment and social engagement. In most cases controlling for county density decreases the magnitude of the coefficient on walkability but it remains positive and statistically significant. Furthermore, when I control for average neighbourhood sociability the coefficient on density becomes insignificant. This suggests that density and sociability are strongly correlated. Furthermore, across all specifications the coefficients on average neighbourhood sociability are significant and positive at the one percent level. This suggests that the distribution of individuals with low, middle, and high levels of sociability within a county is correlated with a resident's own sociability.

In Table 3.2 I highlight the specific impact of shops, primary schools, and parks, on political involvement, volunteer work, time spent with neighbours, and time spent at social community events. For almost all types of interactions I find that each of the environmental features have a positive and significant effect.[11] This is particularly true for access to parks. I find that an individual with a park within walking distance of home (compared to one without) is more likely to visit with neighbours, attend community social events, conduct volunteer work and even participate in local politics. These results support the Jane Jacobs perspective of urban design – walkable neighbourhoods are correlated with community involvement.

These overall strong positive and significant results on access to amenities — and in particular parks and shops — are likely contaminated by endogenous factors. It could be that unobservable characteristics are affecting both an individual's social interactions as well as their location choice. For example, if

[10] To check for reporting bias (i.e. that those who use certain amenities may have a more accurate estimation of the distance from their home) I run the same regressions using county averages. This assumes that the bias is evenly distributed across the county. Overall, the magnitudes of the coefficients decrease slightly but there is no change to the significance of the results. I perform the same exercise on the 'stayers' sample.

[11] Local political involvement is largely unaffected by the environmental characteristics. In the German SOEP, the average time spent in local politics is less than any other measure of social engagement. It would be interesting to look at this relationship in a more politically active population. Although, Glaeser and Gottlieb (2006) found similar results for the relationship between density and civic involvement in the US context.

more social people choose to live near parks, the estimated impact of parks on social interactions will be biased upwards. The next two subsections address this issue using first differences for both the sample of non-movers ('stayers') and the sample of movers.

3.4.2 First Differences of Stayers

Table 3.3 presents the first differenced results for those who have not moved over the sample period ('stayers'). Using the time-series variation in a stayer's environment I difference over five and 10 year intervals, reporting the 10 year intervals – the differences between 1999 and 2009. This first-differencing removes unobservable time-invariant individual characteristics. For each of, interactions with groups (*Group A*), interactions with family (*Group B*) and neighbours (*Group C*), and interactions with the general community (*Group D*), I do not find any significant effect of neighbourhood walkability on social interactions. However, as in the cross-section the significant effect of average county sociability remains; higher average sociability is correlated with more social interactions. This result also holds when I decompose the social interaction indices into their component parts.

Tables 3.A7 and 3.A8 of the Appendix present the breakdown by types of interactions and neighbourhood characteristics for changes over 10 years. In most cross-tabulations of social interactions of environmental characteristics the significant results found in the pooled cross-section disappear after controlling for unobservable characteristics. This implies that the unobservable propensity to socialize is likely contaminating previous results and that the environment does not play as large a role in fostering social engagement as previously estimated.[12] This contradicts the views of Jacobs (1961) focusing on the placement of shops, parks and public schools (among other things) to generate social engagement. Furthermore, whereas Brueckner and Largey (2008) find a negative effect of density on social interactions, I find no significant effect (although an often negative coefficient). While, these differences could be largely due to an analysis on very different populations, it is interesting that both of our attempts to address endogeneity have refuted cross-sectional results.

These results compared to those found in the pooled cross-section provide evidence of sorting behaviour by more social individuals into high-density neighbourhoods. That said, the decision to stay in a neighbourhood is made endogenously. If people have had an unobserved change in their propensity to engage socially that is also correlated with changes in their neighbourhoods (for example, if individuals become more social and lobby for a park nearby) then these results will still suffer from endogeneity. Provided this does not happen frequently, my results should still be capturing the average effect. Furthermore, if I assume that there is a true positive underlying relationship between social interactions and neighbourhood walkability, then any unobserved change to an individual's sociability should be biasing the results upwards (i.e. towards finding a result).

One concern still inherent in these estimates for stayers is that changes in a county's environmental characteristics happen very slowly (beyond 10 years). Therefore, in the next section I look at those who move; first, those who move within a county and second, those who move between counties.

[12] As an alternative check on the non-randomness of location choice, I look specifically at the regression of social interactions on environmental characteristics for children who have entered the SOEP survey upon turning 17 but who have not moved from their parents' house. I have 1193 observations for 1097 individuals. I find no significance on access to amenities or on density, or density squared. This is true for interactions with family, with neighbours, with the community and with groups.

3.4.3 First Differences of Movers

3.4.3.1 Movers within a County

Approximately eight percent of the respondents in the SOEP move within counties per year; 44% of the sample move at least once. I exploit the change in the environment following a move to estimate the impact of environmental characteristics on social interactions. The decision to move is not random and individuals who value social interactions may move closer to amenities or to higher density neighbourhoods. First, I consider one year following a move compared to the year prior to moving.[13] Second, I consider five years following a move compared to one year prior to moving. For brevity, I report the differences one year following a move given that the results do not change when I extend to the differences over five years.

While the stayers are analyzed using own responses to the built environment questions the movers are analyzed using county average responses. The built environment questions were only asked in 1994, 1999, 2004, and 2009 and given that people move outside of these years, a move in say 1995 would not have the individual's own responses to environmental questions until 1999, four years after their actual move and this would drastically reduce the sample size. Neither the pooled cross-section nor the stayers sample changed when I replaced own-responses with county averages so I do not expect this to bias the results.

Table 3.4 presents the main results and is analogous to Table 3.3 for the stayers. The first four columns present the first differences for one year after moving compared to one year prior for Group Interactions, followed by Family, Neighbours, and the Community. For the most part I find no significant impact of access to amenities on any of the social interaction indices, nor on any of their component parts (Table 3.A10 of the appendix). Table 3.A11 of the Appendix presents the complete cross-tabulation each type of interaction and environmental characteristic, analogous to Table 3.A8 for the stayers. I do however find that neighbourhood walkability affects political involvement and interactions with neighbours. The significant results for political involvement is coming largely from proximity to schools. While this is potentially interesting, these effects disappear when looking at movers between counties. Aside from this, only my measures of average neighbourhood sociability remain significant. As with the stayers subsample I find a significant and positive result for those in the top 30 percent of the distribution. While access to amenities and county density generally have no significant impact on social interactions, the sociability of one's neighbours does.

3.4.3.2 Movers between Counties

Approximately two percent of the respondents in the SOEP dataset move between counties per year. Table 3.5 presents the results for movers between counties and are analogous to those in Table 3.3 for the stayers and Table 3.4 for the movers within counties. Once again I find an insignificant impact of neighbourhood walkability on social interactions and when looking at the specific features of the environment (Tables 3.A13 and 3.A14), I find no pattern or significance with any of the environmental variables. I

[13] I exclude the year of the move due to a large number of missing values for social interactions in these years. Similarly, I do not have the month of the move in the current data set.

do however maintain a significant relationship with respect to average neighbourhood sociability; I find positive and significant results for the percentage of individuals in the bottom 30 percent.

Thus far, for the samples of stayers, movers within counties, and movers outside of counties I find no significant impact of neighbourhood walkability or density on social interactions. One point worth addressing in more detail is the magnitude of the standard errors on the first-difference results relative to the cross sectional results. While the estimated coefficients tend towards zero, the standard errors become quite large after differencing. In the next subsection, I will briefly address the confidence intervals around the results.

3.4.4 Standard Errors on First-Differenced Results

Figures 3.A1 (a)-(d) of the Appendix present the graphical representation for the various regressions along with their confidence intervals. In the regressions for interactions with groups, the estimated coefficients in the cross-section and over the stayers differenced sample are statistically different from each other at the 95% confidence level for all values of neighbourhood walkability greater than one. This is true also for interactions with the community and with family. While the standard errors increase, they are still outside the bounds of the cross-sectional results. This is however not true for interactions with neighbors; I cannot distinguish the cross-sectional coefficient from the first-differenced stayers coefficient.

However, when looking at the movers, the coefficient on movers between counties is statistically different from the cross-section coefficient at the 95% confidence level for all values of neighborhoods walkability greater than one. For movers within counties, they are statistically different, for all values of neighborhood walkability greater than two. The movers sample is quite a bit noisier than the stayers, so the 95% confidence intervals around the coefficients for group interactions, family interactions, and community interactions often include the cross-sectional results at the lower levels of neighborhood walkability. However, as the walkability index moves beyond seven, they diverge.

While I am aware that the differenced results are less precise than the cross-sectional results, the fact that all differenced regressions generate similar results across all groups considered is comforting. If these results are in fact driven by sorting based on unobservables I should also find evidence of sorting on observables. In the next section I discuss these characteristics.

3.4.5 Sorting of more Social Individuals into more Walkable Communities

If sorting is driving the cross-sectional results, I should see the same set of characteristics correlated with social interactions that are predictive of who is living in more walkable, centrally located, neighbourhoods. Appendix Table 3.A6 includes the full set of cross-section regression results. In the cross-sectional regressions of social interactions on neighbourhood characteristics, the individual controls provide some insight into the observable characteristics of more social individuals. It is interesting to note that these characteristics are not necessarily consistent across each category of interaction. For example, men are more likely to participate in community activities while less likely to interact with their family or neighbours. Those with higher incomes are more likely to participate in group actives, community activities and interact with their neighbours, but less likely to socialize with their family. Relative to married people, those who are single, separated, or divorced, are more likely to interact with their neighbours and to go out in their community. Relative to those who are not in the labour force, people working full time or people who are retired are more likely to participate in their community, whereas those who are unemployed or on maternity leave are less likely.

Considering which of these observable characteristics also predict who lives in more walkable, centrally located neighbourhoods, Table 3.6 shows the results from regressing the walkability index, county density, and distance to the city center on the observable control characteristics. While men are more likely to go out in their community they are also more likely to live near more amenities, but in less dense counties and further from the city center. Higher income individuals are more likely to live both closer to amenities, in more dense counties, and closer to the city center. Homeowners tend to live further from the city center, as do those on maternity leave (relative to not being in the labour force). Relative to married individuals, those who are single or separated are more likely to live closer to the city center and in more dense counties.

While this is not conclusive evidence of sorting behaviour, it is interesting to note that many of the observable characteristics related to social interactions are also related to living near amenities, or in higher-density locations closer to the city center. As a final set of robustness checks I look at changes in access to amenities for more specific groups of respondents; specific both in terms of their geographic locations and their socio-demographic characteristics.[14] There could be certain people who have more time elasticity such that they can go to shops and parks more often. Alternatively, parents of young children are more likely to use schools and youth centers than those without children. In the next section I consider these possibilities.

3.4.6 Robustness Checks

Up to this point I have estimated the first differences for stayers and for those who move within a county or between counties. In this section I now consider various splits of these original subsamples. In Tables 3.A16 and 3.A18 I include the first set of robustness checks for changes in the interactions of stayers and changes in the interactions of movers between counties, respectively.

First, I expand amenity access from 0.8 kilometers to 1.6 kilometers from home. Next, I restrict the sample of stayers to those in East Germany, those in West Germany, and then I exclude Berlin.[15] This allows me to explore persistent differences that may be present in either region and ensures that the large representation of Berlin in the sample is not driving the results. Overall the results do not change: the relationship between neighbourhood walkability and social interactions is insignificant. Finally, for both the stayers and the movers I then divide the sample into those who experience increases in their access to neighbourhood amenities and those who experience decreases. This is an attempt to look specifically at those who experience the biggest changes in their neighbourhood characteristics. Similarly, I look

[14] As a further specification test I considered looking at partners who move following a job change of their spouse. Unfortunately in the SOEP dataset there are a relatively small percentage of movers which implies an even smaller percentage who move for any one particular reason. When I look at this subset of the sample I am left with less than 100 observations. Therefore, I do not pursue this further at this point.

[15] For the movers, I considered looking at those who move between East and West Germany however, less than 100 people in the sample moved between the two.

at those who faced the biggest top quarter increase in density and those who faced the biggest bottom quarter decrease in density. For the stayers, the only significance I find is with respect to interactions with family however, these results disappear when I consider the same changes for movers.

Next, I focus specifically on changes in access to shops, primary schools, and parks, for each of: respondents less than 40 years old, females only, parents of children less than 16 years old, and respondents who are either retired or on maternity leave. Each of these groups of people have different uses for certain amenities and different elasticities with respect to their time available to use either. Table 3.A17 contains the results for the stayers, Table 3.A19 contains the results for the movers. For the stayers, I find an overwhelming set of insignificant results over all sample stratifications.

However, for movers between counties I find some significance of access to primary schools in increasing family interactions for individuals less than 40, females only, and those with young children. I also find that access to a primary school increases time spent in group interactions for individuals less than 40 and those with young children. This suggests that there are some time varying factors which drive the decision to move and also affect sociability; controlling for these time varying factors should remove this endogenous relationship.

Overall, the original results I present looking at changes in social interactions with groups, family, neighbours, and the general community, are robust to alternative specifications and sample selections. Across all specifications for all groups of people, I find overwhelmingly insignificant results for the impact of environmental characteristics on social interactions. This indicates that the significance found in the pooled cross-section (in both this chapter and in the previous literature) is driven by (i) individuals with a propensity to be social or (ii) individuals experiencing a change that affects both their location decision and sociability who are sorting into neighbourhoods with greater access to amenities.

3.5 Discussion and Conclusion

The optimal design of a community aimed at facilitating social interactions and community engagement has a great deal of policy relevance. Jane Jacobs criticized urban planners for not catering to the residents who need walkable neighbourhoods with access to amenities because these are necessary for social interactions. Robert Putnam criticized urban sprawl as destructive to social and civic engagement. In this chapter I empirically estimate the relationship between urban form and social interactions considering the effect of neighbourhood walkability (i.e. access to amenities) and average sociability.

In a pooled cross-section, results corroborate previous literature – social interactions are significantly and positively correlated with neighbourhood walkability and density. Recognizing that the location decisions of individuals are not random and will be influenced by unobservable attributes, I use a first-differences specification. These unobservable propensities may cause people to sort into more social neighbourhoods, or neighbourhoods that have more street level interactions. I consider three different subsets of the population – stayers, movers within counties, and movers between counties.

The significant effects found in the cross-section disappear for neighbourhood walkability and county density. This suggests that the cross-sectional results of previous research are subject to endogeneity bias and that there is sorting by relatively social individuals into high-density neighbourhoods or cities.

I find some evidence that this sorting is correlated with life-cycle changes such as having young children that may affect both residential decisions and social relationship. I also find that in all specifications the relationship with average county sociability remains consistently positive and significant. This provides suggestive evidence that the composition of individuals in a neighbourhood matter more than the physical built features of the environment at generating interactions. That said, if more walkable neighbourhoods are attracting more socially inclined people this is likely a desirable outcome. Furthermore, there are any number of benefits derived from a more walkable community including safety, a sense of connectedness, improved health, and aesthetic appeal, which should not be disregarded. So, while I may not find a significant relationship between the built environment and social interactions specifically, there are a number of other positive benefits whose true relationship requires further investigation.

Tables

Table 3.1: Baseline OLS Results: cross-sectional relationship

	Group A				**Group B**			
Dependent Variable	Group Interactions (1994-2009)				Family Interactions (1994-2009)			
	(1)	(2)	(3)	(4)†	(1)	(2)	(3)	(4)†
Independent Variables								
Walkability	0.017*		0.025***	0.025***	0.012***		0.016***	0.013***
	(0.009)		(0.007)	(0.005)	(0.003)		(0.003)	(0.002)
Density		0.076	-0.224**	0.012		0.017	-0.006	-0.033**
(/1,000)		(0.110)	(0.085)	(0.042)		(0.048)	(0.046)	(0.017)
Density^2		-0.001	0.003	-0.001		-0.001	-0.001	0.001**
(/100,000)		(0.003)	(0.003)	(0.001)		(0.001)	(0.001)	(0.000)
Average Sociability:								
% in top third				3.759***				1.174***
				(0.127)				(0.054)
% in bottom third				-2.329***				-1.455***
				(0.083)				(0.029)
	Group C				**Group D**			
Dependent Variable	Neighbor Interactions (1994-2009)				Community Interactions (1994-2009)			
	(1)	(2)	(3)	(4)†	(1)	(2)	(3)	(4)†
Independent Variables								
Walkability	0.022***		0.015***	0.010***	0.071***		0.040***	0.033***
	(0.003)		(0.003)	(0.002)	(0.009)		(0.007)	(0.005)
Density		0.183***	0.090***	-0.028*		0.481***	0.314***	0.006
(/1000)		(0.001)	(0.001)	(0.000)		(0.105)	(0.082)	(0.049)
Density^2		-0.004***	-0.002**	0.001		-0.008***	-0.005*	-0.000
(/100,000)		(0.001)	(0.001)	(0.000)		(0.003)	(0.002)	(0.001)
Average Sociability:								
% in top third				0.948***				2.621***
				(0.088)				(0.118)
% in bottom third				-1.409***				-2.809***
				(0.036)				(0.120)
Individual Controls			•	•			•	•
Regional Controls				•				•
Observations	145,455	145,455	140,710	140,710	145,455	145,455	140,710	140,710
Individuals	25,806	25,806	24,745	24,745	25,806	25,806	24,745	24,745

Notes: The above table presents the baseline OLS regressions for Social Interactions with each of *Groups*(Panel A), *Family* (Panel B), *Neighbors* (Panel C), and *the Community* (Panel D) as a function of Neighborhood Walkability, Density, and Average Neighborhood Sociability. Point estimates are reported and standard errors are in parenthesis and are clustered at the county level; * significant at 10 percent, **significant at 5 percent. ***significant at 1 percent. All regressions include year fixed effects; † include controls for state fixed effects.

Table 3.2: Baseline OLS Results: by neighborhood characteristics

Independent Variable	Walkability Index	*Neighborhood Characteristics* (access to (0/1 dummy)) Parks	Schools	Shops	(distance to (0.8 to 1.6 km)) Parks	Schools	Shops
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent Variables							
Group Interactions	0.025***	0.114***	0.157***	0.069***	-0.429**	-0.736***	-0.420**
	(0.01)	(0.03)	(0.03)	(0.03)	(0.13)	(0.14)	(0.16)
Political Involvement	0.041	0.013**	0.006	0.000	-0.023	-0.021	-0.024
	(0.03)	(0.01)	(0.01)	(0.01)	(0.03)	(0.03)	(0.03)
Volunteer Work	0.182**	0.038**	0.044**	0.007	-0.093	-0.253***	-0.148*
	(0.07)	(0.01)	(0.01)	(0.01)	(0.07)	(0.07)	(0.08)
Neighbor Interactions	0.010***	0.029**	0.034**	0.079***	-0.531***	-0.217	-0.301*
	(0.01)	(0.01)	(0.01)	(0.01)	(0.16)	(0.16)	(0.17)
Community Interactions	0.033***	0.158***	0.148***	0.103***	-0.395**	-0.604***	-0.589***
	(0.01)	(0.03)	(0.03)	(0.03)	(0.14)	(0.12)	(0.14)
Social Events	0.203***	0.063***	0.032**	0.015	-0.243***	-0.059	-0.071
	(0.06)	(0.01)	(0.01)	(0.01)	(0.07)	(0.06)	(0.06)
	Observations: 140,710				Observations: 116,535		
	Individuals: 24,745				Individuals: 21,212		

Notes: The above table presents a sample of the baseline OLS regression results broken down by neighborhood characteristics and social interactions of particular interest. Columns (2), (3), and (4) present the results for having access to each of parks, schools, and shops within a 10 minute walk of home, as a yes/no dummy variable. Columns (5), (6), and (7) present the results for the approximate distance to each of parks, schools, and shops. Standard errors are reported in parenthesis and are clustered at the county level; * significant at 10 percent, **significant at 5 percent. ***significant at 1 percent. All regressions include year and state fixed effects, individual and regional controls.

Table 3.3: First Differences of Stayers: 10 year time variation

	Group A				**Group B**			
Dependent Variable	Group Interactions (1999-2009)				Family Interactions (1999-2009)			
	(1)	(2)	(3)	(4)†	(1)	(2)	(3)	(4)†
Independent Variables								
Walkability	-0.014		-0.017	-0.013	-0.010		-0.013	-0.015
	(0.02)		(0.02)	(0.02)	(0.01)		(0.01)	(0.01)
Density		0.162	-0.171	0.737		0.733	0.954	1.191
(/1,000)		(1.44)	(1.43)	(1.26)		(0.79)	(0.80)	(0.94)
Density^2		0.002	0.010	-0.006		-0.006	-0.009	-0.014
(/100,000)		(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	
Average Sociability:								
% in top third				2.950***				0.914**
				(0.60)				(0.38)
% in bottom third				-1.623***				-0.935**
				(0.45)				(0.19)

	Group C				**Group D**			
Dependent Variable	Neighbor Interactions (1999-2009)				Community Interactions (1999-2009)			
	(1)	(2)	(3)	(4)†	(1)	(2)	(3)	(4)†
Independent Variables								
Walkability	0.008		0007	0.006	0.004		0.004	0.005
	(0.01)		(0.01)	(0.01)	(0.02)		(0.02)	(0.02)
Density		-0.799	-0.797	0.266		-0.724	-1.524	-0.759
(/1000)		(0.65)	(0.66)	(0.69)		(1.27)	(1.29)	(1.35)
Density^2		0.007	0.008	-0.009		0.005	0.020	0.011
(/100,000)		(0.01)	(0.01)	(0.01)		(0.02)	(0.02)	(0.02)
Average Sociability:								
% in top third				1.503**				1.209**
				(0.58)				(0.53)
% in bottom third				-0.813***				-2.182***
				(0.21)				(0.45)
Individual Controls			•	•			•	•
Regional Controls				•				•
Observations	2,546	2,546	2,464	2,464	2,546	2,546	2,464	2,464
Individuals	2,546	2,546	2,464	2,464	2,546	2,546	2,464	2,464

Notes: The above table presents the first differenced results for the stayers subset of the population. First-differences are taken over 10 years of panel data, looking at changes to the built environment between 1999 and 2009. Results are shown for interactions with each of *Groups*(Panel A), *Family* (Panel B), *Neighbors* (Panel C), and *the Community* (Panel D) as a function of Neighborhood Walkability, Density, and Average Neighborhood Sociability. Point estimates are reported and standard errors are in parenthesis and are clustered at the county level; * significant at 10 percent, **significant at 5 percent. ***significant at 1 percent. All regressions include year fixed effects; † include controls for state fixed effects.

Table 3.4: First Differences of Movers *within* Counties: 1 year after move

	Group A				**Group B**			
Dependent Variable	Group Interactions (1994-2009)				Family Interactions (1994-2009)			
	(1)	(2)	(3)	(4)†	(1)	(2)	(3)	(4)†
Independent Variables								
Avg. Walkability	0.005		-0.006	-0.009	0.022		0.017	0.014
	(0.03)		(0.03)	(0.03)	(0.02)		(0.02)	(0.02)
Density		0.380	0.042	0.145		1.074	0.820	0.831
(/1,000)		(1.41)	(1.56)	(1.39)		(0.83)	(0.90)	(0.85)
Density^2		-0.007	-0.001	-0.003		-0.020	-0.015	-0.015
(/100,000)		(0.03)	(0.03)	(0.03)		(0.02)	(0.02)	(0.02)
Average Sociability:								
% in top third				1.627***				0.751**
				(0.37)				(0.24)
% in bottom third				-1.471***				-0.972***
				(0.27)				(0.12)

	Group C				**Group D**			
Dependent Variable	Neighbor Interactions (1994-2009)				Community Interactions (1994-2009)			
	(1)	(2)	(3)	(4)†	(1)	(2)	(3)	(4)†
Independent Variables								
Avg. Walkability	0.038**		0.035**	0.025*	0.065**		0.056*	0.035
	(0.02)		(0.02)	(0.02)	(0.03)		(0.03)	(0.03)
Density		-0.264	-0.136	0.139		0.786	1.206	1.282
(/1000)		(0.74)	(0.80)	(0.84)		(0.83)	(0.88)	(1.07)
Density^2		0.006	0.003	-0.002		-0.014	-0.023	-0.025
(/100,000)		(0.01)	(0.02)	(0.02)		(0.02)	(0.02)	(0.02)
Average Sociability:								
% in top third				1.502***				1.756***
				(0.25)				(0.35)
% in bottom third				-1.126***				-1.828***
				(0.12)				(0.26)
Individual Controls			•	•			•	•
Regional Controls				•				•
Observations	5,279	5,279	5,059	5,059	5,279	5,279	5,059	5,059
Individuals	4,675	4,675	4,489	4,489	4,675	4,675	4,489	4,489

Notes: The above table presents the first differenced results for the subset of the population who move within their county. First-differences are taken as the difference one year after moving compared to one year prior to moving, straddling the year of the move (i.e. (t+1) - (t-1)). Results are shown for interactions with each of *Groups*(Panel A), *Family* (Panel B), *Neighbors* (Panel C), and *the Community* (Panel D) as a function of Average Neighborhood Walkability, Density, and Average Neighborhood Sociability. Point estimates are reported and standard errors are in parenthesis and are clustered at the county level; * significant at 10 percent, **significant at 5 percent. ***significant at 1 percent. All regressions include year fixed effects; † include controls for state fixed effects.

Table 3.5: First Differences of Movers *between* Counties: 1 year after move

	Group A				**Group B**			
Dependent Variable	Group Interactions (1994-2009)				Family Interactions (1994-2009)			
	(1)	(2)	(3)	(4)†	(1)	(2)	(3)	(4)†
Independent Variables								
Avg. Walkability	0.003		0.023	0.022	0.028		0.020	0.012
	(0.03)		(0.04)	(0.04)	(0.02)		(0.02)	(0.02)
Density		-0.055	-0.068	-0.028		0.014	0.039	0.077
(/1,000)		(0.12)	(0.14)	(0.14)		(0.08)	(0.08)	(0.09)
Density^2		0.001	0.001	0.001		-0.001	-0.001	-0.001
(/100,000)		(0.01)	(0.01)	(0.01)		(0.02)	(0.02)	(0.02)
Average Sociability:								
% in top third				0.788				0.577
				(0.53)				(0.44)
% in bottom third				-1.193**				-0.801***
				(0.42)				(0.22)

	Group C				**Group D**			
Dependent Variable	Neighbor Interactions (1994-2009)				Community Interactions (1994-2009)			
	(1)	(2)	(3)	(4)†	(1)	(2)	(3)	(4)†
Independent Variables								
Avg. Walkability	-0.018		-0.021	-0.029	0.053		0.036	0.020
	(0.02)		(0.02)	(0.02)	(0.03)		(0.04)	(0.04)
Density		0.016	0.041	0.018		0.063	-0.065	-0.203*
(/1000)		(0.05)	(0.07)	(0.07)		(0.13)	(0.12)	(0.12)
Density^2		-0.001	-0.002	-0.001		-0.001	0.001	0.005
(/100,000)		(0.01)	(0.01)	(0.01)		(0.01)	(0.01)	(0.01)
Average Sociability:								
% in top third				0.036				1.654***
				(0.63)				(0.53)
% in bottom third				-0.687***				-0.950**
				(0.22)				(0.43)
Individual Controls			•	•			•	•
Regional Controls				•				•
Observations	1,384	1,384	1,320	1,320	1,384	1,384	1,320	1,320
Individuals	1,309	1,309	1,248	1,248	1,309	1,309	1,248	1,248

Notes: The above table presents the first differenced results for the subset of the population who move *between* counties. First-differences are taken as the difference one year after moving compared to one year prior to moving, straddling the year of the move (i.e. (t+1) - (t-1)). Results are shown for interactions with each of *Groups*(Panel A), *Family* (Panel B), *Neighbors* (Panel C), and *the Community* (Panel D) as a function of Average Neighborhood Walkability, Density, and Average Neighborhood Sociability. Point estimates are reported and standard errors are in parenthesis and are clustered at the county level; * significant at 10 percent, **significant at 5 percent. ***significant at 1 percent. All regressions include year fixed effects; † include controls for state fixed effects.

Table 3.6: Characteristics of Individuals Living in Walkable Neighborhoods

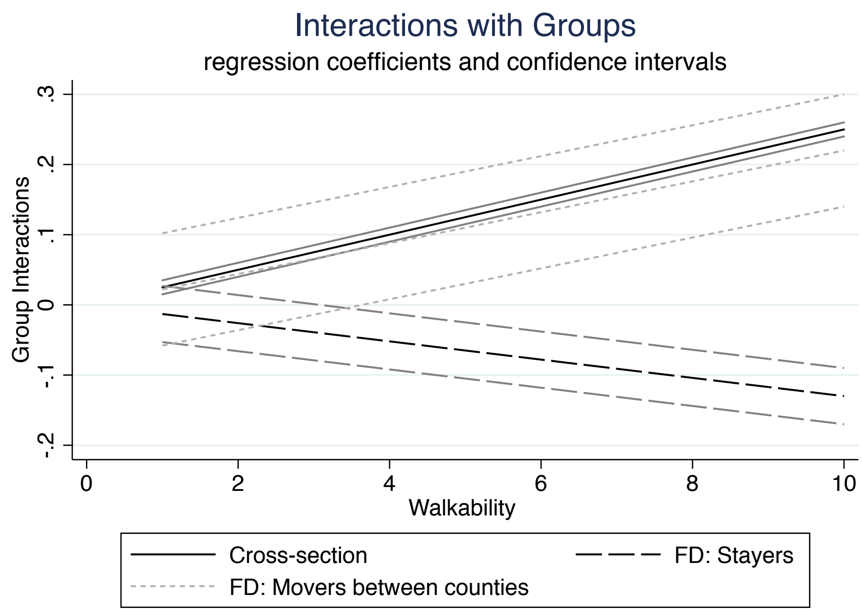
	Dependent Variables		
	Walkability (Index)	Density (/1000)	Distance to city center
Independent Variables			
male	0.062**	-0.029**	0.021*
	(0.02)	(0.01)	(0.01)
age	0.025**	0.007	-0.005
	(0.01)	(0.004)	(0.004)
young children	-0.093	-0.023	0.040
	(0.07)	(0.02)	(0.03)
income	0.217**	0.115***	-0.107**
	(0.07)	(0.03)	(0.04)
education	0.013	0.026***	-0.034***
	(0.01)	(0.01)	(0.01)
home-owner	-0.719***	-0.370***	0.572***
	(0.08)	(0.06)	(0.06)
marital status (omitted category: married)			
single	0.132	0.117**	-0.090**
	(0.09)	(0.04)	(0.05)
separated	0.121	0.079**	-0.149**
	(0.12)	(0.04)	(0.06)
divorced	0.153*	0.054**	-0.071
	(0.09)	(0.02)	(0.04)
labor force status (omitted category: not in labor force)			
working	0.085	-0.050**	0.038
	(0.07)	(0.02)	(0.03)
unemployed	0.160*	0.003	0.003
	(0.08)	(0.02)	(0.04)
maternity leave	0.063	-0.123**	0.113**
	(0.12)	(0.05)	(0.06)
retired	-0.034	0.008	-0.004
	(0.08)	(0.03)	(0.04)
constant	2.174**	-0.742*	4.298***
	(0.83)	(0.44)	(0.52)
Observations	140,710	140,710	140,710
Individuals	24,745	24,745	24,745

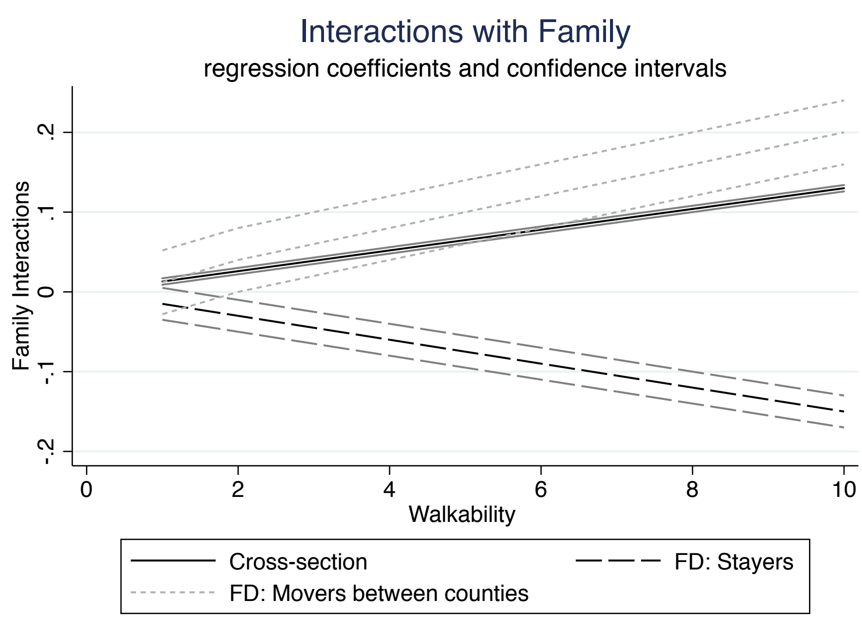
Notes: The above table presents OLS coefficients from regressions of the walkability index, county density, and distance to city center on observable individual characteristics. Point estimates are reported and standard errors are in parenthesis and are clustered at the county level; * significant at 10 percent, **significant at 5 percent. ***significant at 1 percent. All regressions include year and state fixed effects.

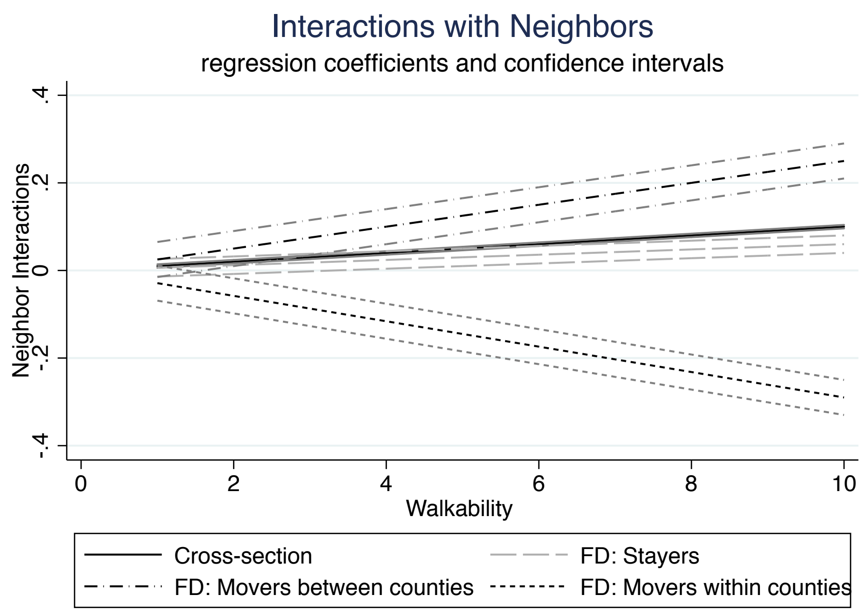
Appendix 3.A

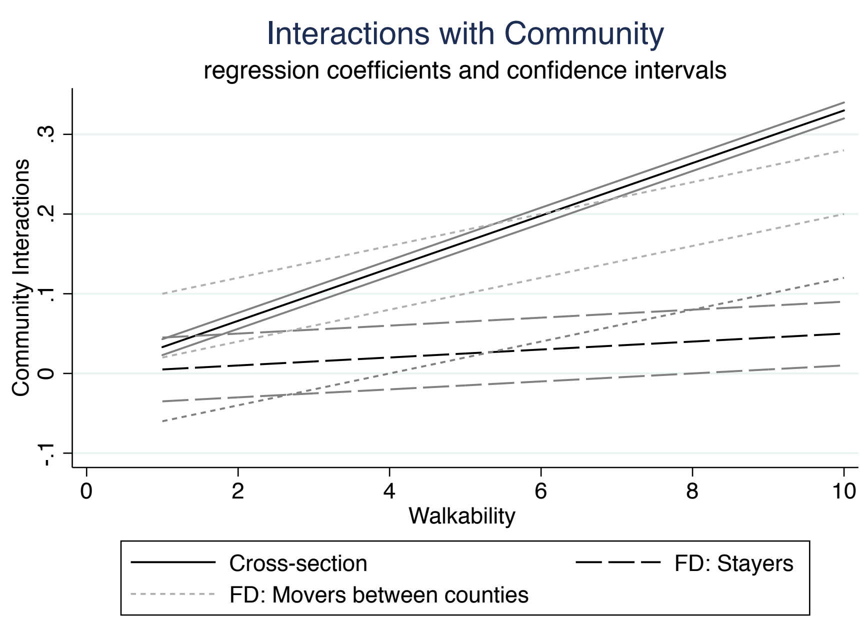
Figures

Figure 3.A1: Standard Errors on Social Interactions






Interactions with Neighbors
regression coefficients and confidence intervals
Neighbor Interactions
-.4
-.2
0
.2
.4
0
2
4
6
8
10
Walkability
Cross-section
FD: Stayers
FD: Movers between counties
FD: Movers within counties


Interactions with Community
regression coefficients and confidence intervals
Community Interactions
-.1
0
.1
.2
.3
0
2
4
6
8
10
Walkability
Cross-section
FD: Stayers
FD: Movers between counties

Tables

Table 3.A1: Descriptive Statistics

Variable	Definition	Mean	SD	Min.	Max.
Dependent Variables					
Total Interaction	participation in any activity, 1994-2009	12.77	4.45	0	33
Group Interactions	group participation, 1994-2009	2.75	2.36	0	16
politics	participate in local politics (d/y)	0.16	0.49	0	4
volunteer	perform volunteer work (d/y)	0.56	1.00	0	4
church	attend church (d/y)	0.82	0.99	0	4
sports	participate in sports (d/y)	1.22	1.31	0	4
Neighbor Interactions	visiting neighbors (d/y), 1994-2009	2.23	0.91	0	4
Family Interactions	visiting family (d/y), 1994-2009	2.33	0.95	0	4
Community Interactions	community presence, 1994-2009	5.45	2.28	0	14
drinks	going out for dinner/drinks (d/y)	1.62	0.96	0	4
cinema	going to cinema/concerts (d/y)	0.87	0.84	0	4
cultural	attending cultural events (d/y)	0.82	0.72	0	4
social	attending social gatherings (d/y)	2.14	0.82	0	4
Environmental (Independent) Variables					
AM 0.8km	amenities within 0.8km (≈ 10 min walk)	5.04	2.97	0	10
AM 1.2km	amenities within 1.2km (≈ 10-20 min walk)	7.77	2.47	0	10
AM 1.6km	amenities within 1.6km (>20 min walk)	8.71	2.05	0	10
park	distance to park (km)	0.97	0.26	0	1.6
kindergarten	distance to kg school (km)	1.01	0.26	0	1.6
primary school	distance to primary school (km)	1.03	0.27	0	1.6
shops	distance to shops (km)	0.96	0.24	0	1.6
pubs	distance to restaurants (km)	0.92	0.23	0	1.6
doctor	distance to family doctor (km)	0.91	0.29	0	1.6
youth center	distance to youth center (km)	1.11	0.30	0	1.6
old age home	distance to old age home (km)	1.14	0.30	0	1.6
sports center	distance to sports center (km)	1.08	0.29	0	1.6
transit	distance to public transit (km)	0.86	0.15	0	1.6
cdensity	county density (county population/sq km)	795	1005	6	5600
city center	= 1 if house in city center, 0 otherwise	0.11	0.29	0	1

Continued on next page...

... table continued

Variable	Definition	Mean	Min.	Max.
Additional Control Variables				
male	= 1 if male, 0 if female	0.48	0	1
age	age of the respondent	46.58	17	100
age^2	age squared			
child16	= 1 if children less than 16 in household	0.32	0	1
married	= 1 if married, 0 otherwise	0.61	0	1
separated	= 1 if serparated 0 otherwise	0.02	0	1
single	= 1 if single 0 otherwise	0.24	0	1
divorced	= 1 if divorced 0 otherwise	0.07	0	1
widowed	= 1 if widowed 0 otherwise	0.07	0	1
lfs: working	= 1 if fully employed, 0 otherwise	0.57	0	1
lfs: unemployed	= 1 if unemployed, 0 otherwise	0.06	0	1
lfs: maternity leave	= 1 if on mat leave, 0 otherwise	0.02	0	1
lfs: in training/school	= 1 if in school, 0 otherwise	0.04	0	1
lfs: retired	= 1 if retired, 0 otherwise	0.16	0	1
lfs: other	= 1 if working pt, in military, etc.	0.15	0	1
log annual income	log of annual income	10.45	2.5	13.8
educ	number of years of education	11.84	7	18
owner	= 1 if owns home, 0 otherwise	0.48	0	1
german	= 1 if German born	0.86	0	1
move	= 1 if moved county between two periods	0.02	0	1
move within	= 1 if moved within a county between two periods	0.076	0	1
East Germany	= 1 if county is in E. Germany, 0 otherwise	0.23	0	1
median income	median income of the county	36,769		
State1-State16	set of dummy variables controlling for state		0	1
Low SI Groups	% in bottom third of SI Groups dist'n	0.36	0	1
High SI Groups	% in top third of SI Groups dist'n	0.25	0	1
Low SI Family	% in top third of SI Family dist'n	0.47	0	1
High SI Family	% in bottom third of SI Family dist'n	0.10	0	1
Low SI Neighbors	% in top third of SI Neighbors dist'n	0.51	0	1
High SI Neighbors	% in bottom third of SI Neighbors dist'n	0.05	0	1
Low SI Comm	% in bottom third of SI Community dist'n	0.28	0	1
High SI Comm	% in top third of SI Community dist'n	0.26	0	1

Observations: 145,455
Individuals: 25,806

Table 3.A2: County Level Environmental Variation

Variable	**Full Sample**		**Stayers**		**Movers (between)**		**Movers (within)**	
differences over:	1 year	1 year	5 years	10 years	1 year	5 years	1 year	5 years
AM 0.8km								
average (county)	1.20	0.40	0.83	0.95	1.38	1.25	0.67	0.94
increase	(1.4)	(0.6)	(0.7)	(0.8)	(1.2)	(1.1)	(1.0)	(0.8)
average (county)	1.36	0.39	0.85	1.00	1.41	1.43	0.70	1.09
decrease	(1.6)	(0.6)	(0.7)	(0.8)	(1.2)	(1.2)	(0.8)	(0.9)
County Density								
average (county)	34.95	31.67	44.19	71.72	1070.61	879.40	90.85	118.18
increase	(252.5)	(212.3)	(267.0)	(343.6)	(1051.4)	(1002.9)	(392.5)	(443.8)
average (county)	25.21	24.63	43.66	40.79	1071.08	980.25	158.43	179.73
decrease	(225.8)	(218.4)	(289.9)	(62.7)	(1074.4)	(1094.4)	(534.4)	(531.9)
Observations:	119,646	74,309	9,009	2,546	1,190	950	6,398	3,283

Notes: Time series variation in the environmental characteristics for those categorized as movers and stayers over the sample periods; mean changes over 1 year, 5 years, and 10 years. Standard deviations are in parenthesis.

Table 3.A3: Variation in Social Interactions

Variable	**Full Sample**		**Stayers**		**Movers (between)**		**Movers (within)**	
differences over:	1 year	1 year	5 years	10 years	1 year	5 years	1 year	5 years
Group Int's	-0.02	-0.03	-0.09	-0.02	-0.04	0.15	0.10	0.21
	(0.7)	(0.7)	(2.1)	(2.1)	(1.8)	(2.1)	(1.8)	(2.0)
Family Int's	-0.01	-0.01	-0.04	-0.08	-0.03	-0.06	0.05	-0.02
	(0.3)	(0.2)	(0.9)	(1.0)	(1.1)	(1.1)	(1.0)	(1.0)
Neighbor Int's	-0.02	-0.02	-0.09	-0.19	-0.12	-0.14	-0.09	-0.15
	(0.2)	(0.2)	(0.8)	(1.0)	(1.0)	(1.0)	(1.0)	(1.0)
Community Int's	-0.02	-0.01	-0.18	-0.32	-0.25	-0.57	-0.29	-0.37
	(0.6)	(0.5)	(1.5)	(1.9)	(2.0)	(2.0)	(1.8)	(2.1)
Observations:	119,646	74,309	9,009	2,546	1,910	950	6,398	3,283
Individuals:	20,104	11,746	6,483	2,546	1,658	791	5,287	2,698

Notes: Time series variation in social interactions for those categorized as movers and stayers over the sample periods; mean changes over 1 year, 5 years, and 10 years. Standard deviations are in parenthesis.

Table 3.A4: Baseline OLS Results: by type of interaction

Panel A				
Dependent Variables				
Group Interactions:	politics	volunteer	church	sports
Independent Variables				
Walkability	0.003**	0.008**	0.002	0.013***
	(0.001)	(0.002)	(0.003)	(0.003)
Density	-0.006	-0.62**	-0.063	0.143***
(/1,000)	(0.017)	(0.023)	(0.040)	(0.035)
Density^2	-0.001	0.001**	0.001	-0.003**
(/100,000)	(0.001)	(0.000)	(0.001)	(0.001)
Average Sociability:				
% in top third	0.485***	1.545***	1.024***	0.705***
	(0.049)	(0.076)	(0.115)	(0.080)
% in bottom	-0.075**	-0.233***	-0.954***	-1.067***
	(0.028)	(0.050)	(0.105)	(0.080)

Panel B				
Dependent Variables				concert/
Community Interactions:	cultural	social	eat/drink	cinema
Independent Variables				
Walkability	0.007***	0.008***	0.014***	0.004**
	(0.002)	(0.002)	(0.002)	(0.002)
Density	0.032*	0.017	-0.026	-0.017
(/1,000)	(0.019)	(0.035)	(0.029)	(0.016)
Density^2	-0.001*	-0.001	0.001	0.001
(/100,000)	(0.000)	(0.001)	(0.001)	(0.001)
Average Sociability:				
% in top third	0.574***	0.354***	0.900***	0.793***
	(0.076)	(0.073)	(0.063)	(0.043)
% in bottom	-0.619***	-0.771***	-0.828***	-0.592***
	(0.039)	(0.062)	(0.056)	(0.036)

Individual Controls	•	•	•	•
Regional Controls	•	•	•	•
Observations	140,710	140,710	140,710	140,710
Individuals	24,745	24,745	24,745	24,745

Notes: The above table presents the baseline OLS regression results broken down into the specific types of interactions. In the main results, politics, volunteer work, sports, and church attendance are grouped together into *Group Interactions*; cultural, social events, going out to eat/drink, and going to the concert/cinema are grouped together into *Community Interactions*. Standard errors are reported in parenthesis and are clustered at the county level; *significant at 10 percent, **significant at 5 percent. ***significant at 1 percent. All regressions include year and state fixed effects, individual and regional controls.

Table 3.A5: Baseline OLS Results: full cross-tabulation

Independent Variables		*Neighborhood Amenities*									
	Walkability	Shops	Restau-	Doctor	Kinder-	Primary	Youth	Old-age	Park	Sports	Transit
			rants			garten	School	Center	Home		Center
Group A											
Group	0.025***	0.069***	0.062*	0.065*	0.157***	0.096**	0.162***	0.110**	0.114***	0.103***	0.057
Interactions	(0.01)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)
Politics	0.003**	0.000	0.013*	0.006	0.006	-0.001	0.022**	0.017**	0.013**	0.018**	0.002*
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Volunteer	0.008**	0.007	0.011	0.020	0.044***	0.026*	0.065***	0.033**	0.038*	0.029**	0.019
	(0.00)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)
Church	0.002	0.026	0.037**	0.004	0.001	-0.002	0.036**	0.003	-0.000	-0.003	-0.006
	(0.002)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Sports	0.013***	0.035**	0.001	0.034**	0.106***	0.072***	0.039**	0.057**	0.063***	0.058***	0.042**
	(0.00)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Group B											
Family	0.013***	0.038**	0.059***	0.034**	0.051***	0.042**	0.054***	0.044**	0.084***	0.071***	0.055**
Interactions	(0.00)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)
Group C											
Neighbor	0.010***	0.029**	0.032**	0.028**	0.034**	0.016	0.038**	0.027**	0.079***	0.056***	0.025
Interactions	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
Group D											
Community	0.033***	0.103***	0.084***	0.085***	0.148***	0.102***	0.134***	0.145***	0.158***	0.150***	0.157***
Interactions	(0.01)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)
Cultural	0.007***	0.028***	0.120	0.012	0.029**	0.024**	0.021**	0.036**	0.045***	0.020**	0.041***
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Social	0.008***	0.015	0.027**	0.015	0.032**	0.017	0.027**	0.027**	0.063***	0.050***	0.041**
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
Eat/drink	0.014***	0.048***	0.045***	0.051***	0.060***	0.045***	0.057***	0.063***	0.034**	0.057***	0.054***
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
Cinema/Concert	0.004**	0.012	-0.001	0.007	0.026**	0.017**	0.028**	0.019*	0.016*	0.022**	0.021*
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Observations	140,710	140,710	140,710	140,710	140,710	140,710	140,710	140,710	140,710	140,710	140,710
Individuals	24,745	24,745	24,745	24,745	24,745	24,745	24,745	24,745	24,745	24,745	24,745

Notes: The above table presents the baseline OLS regression results broken down into each surveyed neighborhood characteristic. In the main regression results, these amenities are aggregated into a walkability index, *Walkability*. Standard errors are reported in parenthesis and are clustered at the county level; *significant at 10 percent, **significant at 5 percent. ***significant at 1 percent. All regressions include year and state fixed effects, individual and regional controls.

Table 3.A6: Baseline OLS Results: cross-sectional relationship
(coefficients on individual controls (cont'd from Table 3.1))

	Dependent Variables			
	Groups	Family	Neighbours	Community
Independent Variables				
male	0.036	-0.090***	-0.057***	0.088***
	(0.03)	(0.01)	(0.01)	(0.02)
age	0.024***	-0.016***	-0.040***	-0.082***
	(0.01)	(0.003)	(0.003)	(0.01)
young children	0.053	0.021	-0.028*	-0.561***
	(0.04)	(0.02)*	(0.02)	(0.03)
income	0.287***	-0.099***	0.037**	0.745***
	(0.03)	(0.02)	(0.01)	(0.03)
education	0.161***	-0.017***	0.008***	0.145***
	(0.01)	(0.003)	(0.002)	(0.01)
home-owner	0.535***	0.021	0.043**	0.171***
	(0.04)	(0.02)	(0.01)	(0.03)
marital status (omitted category: married)				
single	-0.047	-0.231***	0.170***	0.736***
	(0.05)	(0.02)	(0.02)	(0.05)
separated	-0.193*	-0.158***	0.126**	0.478***
	(0.10)	(0.04)	(0.04)	(0.10)
divorced	-0.266***	-0.189***	0.110***	0.449***
	(0.06)	(0.03)	(0.03)	(0.05)
labor force status (omitted category: not in labor force)				
working	0.070*	-0.104***	-0.151***	0.419***
	(0.04)	(0.02)	(0.02)	(0.07)
unemployed	-0.198***	-0.185***	-0.121***	-0.238***
	(0.05)	(0.03)	(0.02)	(0.05)
maternity leave	-0.399***	0.071**	-0.025	-0.266***
	(0.07)	(0.03)	(0.03)	(0.07)
retired	0.163***	0.055**	0.078***	0.310***
	(0.05)	(0.03)	(0.02)	(0.05)
constant	-5.125***	2.703***	2.192***	-4.517***
	(0.41)	(0.22)	(0.18)	(0.44)
Observations	140,710	140,710	140,710	140,710
Individuals	24,745	24,745	24,745	24,745

Notes: The above table presents the baseline OLS regressions for Social Interactions with Groups, Family, Neighbors, and the Community, as a function of features of the built environment (columns (4) of Table 4) and individual controls. Point estimates are reported and standard errors are in parenthesis and are clustered at the county level; * significant at 10 percent, **significant at 5 percent. ***significant at 1 percent. All regressions include year and state fixed effects.

Table 3.A7: First Differences of Stayers: by type of interaction

Panel A				
Dependent Variable				
{Group Interactions:}	politics	volunteer	church	sports
Independent Variables				
Walkability	-0.002	-0.006	0.002	-0.007
	(0.01)	(0.01)	(0.01)	(0.01)
Density	0.159	0.031	-0.427	0.975
(/1,000)	(0.42)	(0.59)	(0.62)	(0.75)
Density^2	-0.003	-0.001	0.011	-0.013
(/100,000)	(0.01)	(0.01)	(0.01)	(0.01)
Average Sociability:				
% in top third	0.188	1.237***	0.490*	1.035**
	(0.12)	(0.33)	(0.25)	(0.35)
% in bottom third	-0.164	-0.097	-0.636**	-0.726**
	(0.10)	(0.25)	(0.21)	(0.26)
Panel B				
Dependent Variable				concert/
{Community Interactions:}	cultural	social	eat/drink	cinema
Independent Variables				
Walkability	0.003	0.006	-0.006	0.002
	(0.01)	(0.01)	(0.01)	(0.01)
Density	0.429	-0.761	0.460	-0.888*
(/1,000)	(0.45)	(0.50)	(0.88)	(0.49)
Density^2	-0.006	0.012*	-0.011	0.016**
(100,000)	(0.01)	(0.01)	(0.01)	(0.01)
Average Sociability:				
% in top third	0.329*	0.148	0.191	0.542**
	(0.19)	(0.22)	(0.27)	(0.20)
% in bottom decile	-0.356**	-0.320**	-1.067***	-0.440**
	(0.16)	(0.16)	(0.23)	(0.16)
Individual Controls	•	•	•	•
Regional Controls	•	•	•	•
Observations	2,468	2,468	2,468	2,468
Individuals	2,468	2,468	2,468	2,468

Notes: The above table presents the first-differenced results for the 'stayers' subset of the population broken down into the specific type of social interactions and differenced over 10 years of panel data (1999 - 2009). In the main results, politics, volunteer work, sports, and church attendance are grouped together into *Group Interactions*; cultural, social events, going out to eat/drink, and going to the concert/cinema are grouped together into *Community Interactions*. Standard errors are reported in parenthesis and are clustered at the county level; *significant at 10 percent, **significant at 5 percent. ***significant at 1 percent. All regressions include year and state fixed effects, individual and regional controls.

Table 3.A8: First Differences of Stayers: full cross-tabulation

Independent Variables		*Neighborhood Amenities*									
	AM 0.8km	Shops	Restau- rants	Doctor	Kinder-	Primary garten	Youth School	Old-age Center	Park Home	Sports	Transit Center
Group A											
Group	-0.013	-0.115	-0.052	-0.095	-0.148	-0.099	0.059	-0.096	-0.018	-0.102	0.000
Interactions	(0.02)	(0.09)	(0.09)	(0.10)	(0.09)	(0.10)	(0.09)	(0.09)	(0.08)	(0.08)	(0.09)
Politics	-0.002	0.005	-0.035*	-0.002	-0.005	-0.037	-0.002	-0.015	0.006	-0.015	0.003
	(0.01)	(0.02)	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.02)	(0.03)
Volunteer	-0.006	-0.048	-0.065	-0.059	-0.079*	-0.071	-0.075	-0.028	0.004	0.021	0.027
	(0.01)	(0.05)	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)	(0.05)
Church	0.002	-0.043	0.048	-0.035	-0.013	-0.003	0.002	0.042	0.013	-0.021	-0.044
	(0.01)	(0.04)	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)	(0.04)	(0.05)
Sports	-0.007	-0.029	0.001	0.002	-0.050	0.012	-0.017	-0.095*	-0.041	-0.088*	0.014*
	(0.01)	(0.05)	(0.06)	(0.06)	(0.05)	(0.06)	(0.05)	(0.06)	(0.05)	(0.05)	(0.06)
Group B											
Family	-0.015	0.002	-0.063	-0.068	-0.059	-0.077	-0.035	0.016	0.028	-0.059	-0.010
Interactions	(0.01)	(0.05)	(0.05)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.04)	(0.06)
Group C											
Neighbor	0.006	-0.011	-0.086*	-0.018	0.051	0.025	-0.006	0.059	0.025	0.029	0.041
Interactions	(0.01)	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)	(0.06)
Group D											
Community	0.005	0.011	-0.034	-0.005	0.019	-0.008	0.036	0.062	0.001	-0.115	-0.057
Interactions	(0.02)	(0.09)	(0.09)	(0.10)	(0.09)	(0.10)	(0.09)	(0.09)	(0.08)	(0.08)	(0.11)
Cultural	0.003	0.045	-0.004	-0.033	-0.002	-0.011	0.004	-0.034	0.019	-0.032	0.045
	(0.01)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)
Social	0.006	-0.051	-0.000	0049	0.030	0.056	0.053	0.063	-0.018	-0.026	0.003
	(0.01)	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)
Eat/drink	-0.006	-0.011	-0.017	-0.012	0.027	-0.050	-0.031	-0.000	-0.022	-0.032	-0.015
	(0.01)	(0.04)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.05)
Cinema/Concert	0.002	0.029	-0.013	-0.009	-0.035	-0.003	0.009	0.033	0.022	-0.025	0.024
	(0.01)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)
Observations	2,468	2,468	2,468	2,468	2,468	2,468	2,468	2,468	2,468	2,468	2,468
Individuals	2,468	2,468	2,468	2,468	2,468	2,468	2,468	2,468	2,468	2,468	2,468

Notes: The above table presents the first-differenced results for the 'stayers' subset of the population broken down into each surveyed neighborhood characteristic and differenced over 10 years of the panel data (1999 - 2009). In the main regression results, these amenities are aggregated into a walkability index, *Walkability*. Standard errors are reported in parenthesis and are clustered at the county level; *significant at 10 percent, **significant at 5 percent. ***significant at 1 percent. All regressions include year and state fixed effects, individual and regional controls.

Table 3.A9: First Differences of Stayers: 10 year time variation (coefficients on individual controls (cont'd from Table 3.3))

	Dependent Variables			
	Groups	Family	Neighbor	Community
Independent Variables				
age	0.157***	0.054**	0.027*	0.021
	(0.04)	(0.02)	(0.02)	(0.04)
young children	0.165	-0.025	0.067	-0.046
	(0.11)	(0.06)	(0.05)	(0.10)
income	-0.109	-0.105*	-0.063	0.136
	(0.13)	(0.06)	(0.06)	(0.12)
education	-0.005	-0.015	-0.002	-0.050
	(0.04)	(0.02)	(0.02)	(0.12)
home-owner	0.172	0.147*	0.014	0.131
	(0.15)	(0.08)	(0.09)	(0.15)
marital status (omitted category: married)				
single	-0.142	0.231	0.386**	0.395
	(0.40)	(0.23)	(0.16)	(0.34)
separated	0.417	0.007	0.008	0.012
	(0.47)	(0.22)	(0.20)	(0.47)
divorced	-0.075	-0.067	0.047	-0.225
	(0.25)	(0.22)	(0.15)	(0.28)
labor force status (omitted category: not in labor force)				
working	-0.367**	-0.127**	-0.098*	-0.205**
	(0.12)	(0.06)	(0.06)	(0.09)
unemployed	-0.298***	-0.069	0.055	-0.200
	(0.15)	(0.09)	(0.09)	(0.16)
maternity leave	-1.530**	0.205	0.187	-1.539**
	(0.56)	(0.26)	(0.23)	(0.47)
retired	0.083	0.041	0.068	0.105
	(0.12)	(0.06)	(0.06)	(0.11)
Observations	2,464	2,464	2,464	2,464
Individuals	2,464	2,464	2,464	2,464

Notes: The above table presents the first-differenced results for the 'stayers' subset of the population. Results are shown for interactions with Groups, Family, Neighbors, and the Community, as a function of features of the built environment (columns (4) of Table 6) and individual controls. Point estimates are reported and standard errors are in parenthesis and are clustered at the county level; * significant at 10 percent, **significant at 5 percent. ***significant at 1 percent. All regressions include year and state fixed effects.

Table 3.A10: First Differences of Movers *within* Counties: by type of interaction

Panel A				
Dependent Variable	Change in:			
{Group Interactions}	politics	volunteer	church	sports
Independent Variables				
Avg. Walkability	0.019**	0.002	-0.017	-0.012
	(0.01)	(0.01)	(0.01)	(0.02)
Density	-0.189	0.591	-0.728**	0.470
(/1,000)	(0.63)	(0.68)	(0.34)	(1.15)
Density^2	0.004	-0.012	0.014**	-0.009
(/100,000)	(0.01)	(0.01)	(0.01)	(0.02)
Average Sociability:				
% in top third	0.279**	0.880***	0.066	0.402*
	(0.09)	(0.20)	(0.12)	(0.21)
% in bottom third	0.002	0.057	-0.446***	-1.085***
	(0.06)	(0.13)	(0.10)	(0.18)
Panel B	Change in:			
Dependent Variable				concert/
{Community Interactions}	cultural	social	eat/drink	cinema
Independent Variables				
Avg. Walkability	0.011	0.011	-0.009	0.022*
	(0.01)	(0.01)	(0.02)	(0.01)
Density	0.796**	-0.127	0.215	0.397
(/1,000)	(0.36)	(0.41)	(0.61)	(0.59)
Density^2	-0.015*	0.004	-0.004	-0.009
(/100,000)	(0.01)	(0.01)	(0.01)	(0.01)
% in top third	0.382**	0.219	0.538**	0.618***
	(0.12)	(0.13)	(0.14)	(0.11)
% in bottom third	-0.413	-0.272**	-0.699***	-0.445***
	(3.50)	(3.75)	(7.73)	(2.25)
Individual Controls	•	•	•	•
Regional Controls	•	•	•	•
Observations	5,059	5,059	5,059	5,059
Individuals	4,489	4,489	4,489	4,489

Notes: The above table presents the first-differences results for the 'movers within counties' subset of the population broken down into the specific type of social interactions. Differences are over 1 year after moving compared to 1 year prior (i.e. (t+1) - (t-1)). In the main results, politics, volunteer work, sports, and church attendance are grouped together into *Group Interactions*; cultural, social events, going out to eat/drink, and going to the concert/cinema are grouped together into *Community Interactions.* Standard errors are reported in parenthesis and are clustered at the county level; *significant at 10 percent, **significant at 5 percent. ***significant at 1 percent. All regressions include year and state fixed effects, individual and regional controls.

Table 3.A11: First Differences of Movers *within* Counties: full cross-tabulation

Independent Variables		*Neighborhood Amenities*									
	AM 0.8km (avg)	Shops	Restau-rants	Doctor	Kinder-	Primary garten	Youth School	Old-age Center	Park Home	Sports	Transit Center
Group A											
Group	-0.009	-0.038	0.084	-0.000	0.022	-0.051	0.160*	0.010	-0.097	0.026	-0.042
Interactions	(0.03)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.09)	(0.10)	(0.06)	(0.06)	(0.14)
Politics	0.019**	0.009	0.009	-0.010	0.029*	0.016	0.012	-0.013	0.022	0.030*	-0.005
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.04)
Volunteer	0.002	-0.015	0.055	-0.020	0.003	-0.017	0.012	-0.012	-0.005	0.039	0.076
	(0.01)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.07)
Church	-0.017	-0.030	-0.008	-0.008	-0.024	-0.026	0.021	0.021	-0.021	-0.021	-0.012
	(0.01)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.04)	(0.03)	(0.03)	(0.03)	(0.05)
Sports	-0.012	-0.002	-0.029	0.038	0.014	-0.023	0.115**	0.014	-0.093**	-0.021	-0.100
	(0.02)	(0.04)	(0.05)	(0.04)	(0.05)	(0.05)	(0.05)	(0.07)	(0.05)	(0.05)	(0.09)
Group B											
Family	0.015	-0.061	0.028	0.001	0.025	0.055	0.061	0.152**	0.017	0.050	-0.043
Interactions	(0.01)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	(0.05)	(0.04)	(0.04)	(0.09)
Group C											
Neighbor	0.025*	0.059	0.010	-0.017	0.055	0.067**	0.043	-0.008	-0.012	0.055	-0.049
Interactions	(0.02)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.05)	(0.03)	(0.03)	(0.07)
Group D											
Community	0.035	0.000	0.050	0.099	0.028	0.031	0.139	0.127	-0.055	0.049	0.158
Interactions	(0.03)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.09)	(0.07)	(0.07)	(0.07)	(0.16)
Cultural	0.011	-0.043	-0.017	0.023	0.012	-0.006	-0.028	-0.066	0.000	0.011	0.037
	(0.01)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.07)
Social	0.011	0.033	0.031	0.048*	0.032	0.029	0.069	0.104**	-0.010	0.019	-0.018
	(0.01)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)	(0.05)	(0.05)	(0.03)	(0.03)	(0.07)
Eat/drink	-0.009	0.023	0.051	0.006	-0.018	0.008	0.064	0.062	-0.049	-0.058	0.079
	(0.02)	(0.03)	(0.04)	(0.04)	(0.04)	(0.03)	(0.04)	(0.05)	(0.03)	(0.04)	(0.09))
Cinema/Concert	0.022*	-0.013	-0.015	0.022	0.002	-0.000	-0.034	0.027	0.004	0.077**	0.060
	(0.01)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.06)
Observations	5,059	5,059	5,059	5,059	5,059	5,059	5,059	5,059	5,059	5,059	5,059
Individuals	4,489	4,489	4,489	4,489	4,489	4,489	4,489	4,489	4,489	4,489	4,489

Notes: The above table presents the first-differenced results for the 'movers within counties' subset of the population broken down into each surveyed neighborhood characteristic. Differences are taken over 1 year after moving comported to 1 year prior (i.e. (t+1) - (t-1)). In the main regression results, these amenities are aggregated into a walkability index, *Walkability*. Standard errors are reported in parenthesis and are clustered at the county level; *significant at 10 percent, **significant at 5 percent. ***significant at 1 percent. All regressions include year and state fixed effects, individual and regional controls.

Table 3.A12: First Differences of Movers *within* Counties: 1 year after move (coefficients on individual controls (cont'd from Table 3.4))

	Dependent Variables			
	Groups	Family	Neighbor	Community
Independent Variables				
age	-0.003	-0.037	-0.017	-0.204*
	(0.08)	(0.05)	(0.06)	(0.12)
young children	-0.400**	0.013	-0.042	-0.738***
	(0.15)	(0.09)	(0.09)	(0.14)
income	-0.144	-0.136**	-0.069	0.166
	(0.12)	(0.06)	(0.06)	(0.10)
education	0.001	0.052	0.002	-0.065
	(0.06)	(0.04)	(0.04)	(0.06)
home-owner	0.144	-0.036	-0.004	-0.005
	(0.11)	(0.07)	(0.07)	(0.11)
marital status (omitted category: married)				
single	-0.028	0.067	0.140	0.809***
	(0.20)	(0.10)	(0.10)	(0.21)
separated	0.029	-0.253	0.178	0.446*
	(0.38)	(0.17)	(0.15)	(0.26)
divorced	-0.450	-0.295*	0.279*	0.793**
	(0.29)	(0.16)	(0.15)	(0.25)
labor force status (omitted category: not in labor force)				
working	-0.211	-0.036	-0.084	-0.000
	(0.51)	(0.12)	(0.12)	(0.21)
unemployed	-0.243	0.062	-0.129	0.333
	(0.27)	(0.15)	(0.15)	(0.24)
maternity leave	-0.382	0.249	-0.302*	-0.433
	(0.38)	(0.16)	(0.18)	(0.36)
retired	0.266	0.492**	0.019	0.308
	(0.35)	(0.20)	(0.18)	(0.26)
Observations	1,320	1,320	1,320	1,320
Individuals	1,248	1,248	1,248	1,248

Notes: The above table presents the first-differenced results for the 'movers between counties' subset of the population. Results are shown for interactions with Groups, Family, Neighbors, and the Community, as a function of features of the built environment (columns (4) of Table 6) and individual controls. Point estimates are reported and standard errors are in parenthesis and are clustered at the county level; * significant at 10 percent, **significant at 5 percent. ***significant at 1 percent. All regressions include year and state fixed effects.

Table 3.A13: First Differences of Movers *between* Counties: by type of interaction

Panel A				
Dependent Variable		Change in:		
{Group Interactions}	politics	volunteer	church	sports
Independent Variables				
Avg. Walkability	0.002	0.006	0.001	0.012
	(0.01)	(0.02)	(0.01)	(0.02)
Density	-0.003	-0.021	-0.008	0.004
(/1,000)	(0.03)	(0.06)	(0.06)	(0.09)
Density^2	0.000	-0.000	0.000	-0.000
(/100,000)	(0.01)	(0.01)	(0.01)	(0.01)
Average Sociability:				
% in top third	0.258	0.355	-0.101	0.275
	(0.16)	(0.23)	(0.19)	(0.33)
% in bottom third	-0.033	-0.143	-0.537**	-0.489*
	(0.10)	(0.17)	(0.17)	(0.29)

Panel B		Change in:		
Dependent Variable			concert	
{Community Interactions}	cultural	social	eat/drink	cinema
Independent Variables				
Avg. Walkability	-0.006	-0.015	0.023	0.018
	(0.02)	(0.02)	(0.02)	(0.01)
Density	-0.084	-0.042	-0.051	-0.026
(/1,000)	(0.05)	(0.07)	(0.06)	(0.06)
Density^2	0.002	0.001	-0.001	0.001
(/100,000)	(0.01)	(0.01)	(0.01)	(0.01)
% in top third	0.290	0.434	0.440*	0.490**
	(0.27)	(0.26)	(0.26)	(0.22)
% in bottom third	-0.398*	-0.043	-0.317	-0.192
	(0.20)	(0.23)	(0.21)	(0.18)

Individual Controls	•	•	•	•
Regional Controls	•	•	•	•
Observations	1,320	1,320	1,320	1,320
Individuals	1,248	1,248	1,248	1,248

Notes: The above table presents the first-differences results for the 'movers between counties' subset of the population broken down into the specific type of social interactions. Differences are over 1 year after moving compared to 1 year prior (i.e (t+1) - (t-1)). In the main results, politics, volunteer work, sports, and church attendance are grouped together into *Group Interactions*; cultural, social events, going out to eat/drink, and going to the concert/cinema are grouped together into *Community Interactions*. Standard errors are reported in parenthesis and are clustered at the county level; *significant at 10 percent, **significant at 5 percent. ***significant at 1 percent. All regressions include year and state fixed effects, individual and regional controls.

Table 3.A14: First Differences of Movers *between* Counties: full cross-tabulation

Independent Variables	*Neighborhood Amenities*										
	AM 0.8km (avg)	Shops	Restau- rants	Doctor	Kinder-	Primary garten	Youth School	Old-age Center	Park Home	Sports	Transit Center
Group A											
Group	0.022	0.033	-0.210*	0.006	0.026	0.228**	0.176	0.177	-0.010	0.044	-0.139
Interactions	(0.04)	(0.10)	(0.11)	(0.10)	(0.10)	(0.10)	(0.15)	(0.16)	(0.10)	(0.10)	(0.20)
Politics	0.002	0.018	-0.016	-0.014	0.030	0.027	0.009	0.019	-0.024	-0.030	-0.147**
	(0.01)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.07)
Volunteer	0.006	0.000	-0.076	-0.018	-0.008	0.108**	0.093	0.133*	-0.021	0.060	0.062
	(0.02)	(0.05)	(0.05)	(0.05)	(0.04)	(0.05)	(0.07)	(0.07)	(0.04)	(0.05)	(0.10)
Church	0.001	-0.018	-0.043	-0.032	-0.005	0.008	0.015	-0.010	0.015	-0.023	-0.030
	(0.01)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.05)	(0.05)	(0.04)	(0.04)	(0.09)
Sports	0.012	0.032	-0.075	0.070	0.009	0.085	0.059	0.035	0.021	0.038	-0.025
	(0.02)	(0.06)	(0.07)	(0.06)	(0.06)	(0.07)	(0.09)	(0.10)	(0.06)	(0.07)	(0.15)
Group B											
Family	0.012	-0.040	-0.078	-0.019	0.030	0.133**	0.068	-0.008	0.015	0.014	-0.025
Interactions	(0.02)	(0.05)	(0.07)	(0.06)	(0.06)	(0.05)	(0.07)	(0.07)	(0.07)	(0.05)	(0.15)
Group C											
Neighbor	-0.029	-0.058	-0.106	-0.031	-0.033	0.028	0.054	0.011	-0.039	-0.057	-0.030
Interactions	(0.02)	(0.06)	(0.07)	(0.06)	(0.05)	(0.05)	(0.07)	(0.07)	(0.06)	(0.05)	(0.11)
Group D											
Community	0.020	-0.059	-0.060	0.227**	0.017	0.147	0.374**	0.283*	-0.114	0.078	0.056
Interactions	(0.04)	(0.10)	(0.11)	(0.10)	(0.10)	(0.10)	(0.13)	(0.15)	(0.09)	(0.09)	(0.24)
Cultural	-0.006	0.045	-0.074	0.030	-0.037	-0.010	0.048	-0.021	-0.040	-0.010	0.159*
	(0.02)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.06)	(0.06)	(0.04)	(0.05)	(0.09)
Social	-0.015	-0.096*	-0.011	-0.016	-0.031	0.049	0.042	0.023	0.012	0.030	-0.041
	(0.02)	(0.05)	(0.06)	(0.05)	(0.05)	(0.05)	(0.07)	(0.07)	(0.05)	(0.05)	(0.11)
Eat/drink	0.023	-0.036	0.039	0.102*	0.029	0.076*	0.184**	0.173**	-0.060	0.025	0.012
	(0.02)	(0.05)	(0.06)	(0.05)	(0.04)	(0.05)	(0.06)	(0.07)	(0.05)	(0.05)	(0.11)
Cinema/Concert	0.018	0.029	-0.014	0.112**	0.056	0.034	0.100*	0.108*	-0.027	0.034	-0.074
	(0.01)	(0.04)	(0.05)	(0.04)	(0.05)	(0.05)	(0.06)	(0.06)	(0.04)	(0.05)	(0.08)
Observations	1,320	1,320	1,320	1,320	1,320	1,320	1,320	1,320	1,320	1,320	1,320
Individuals	1,248	1,248	1,248	1,248	1,248	1,248	1,248	1,248	1,248	1,248	1,248

Notes: The above table presents the first-differenced results for the 'movers between counties' subset of the population broken down into each surveyed neighborhood characteristic. Differences are taken over 1 year after moving comported to 1 year prior (i.e. (t+1) - (t-1)). In the main regression results, these amenities are aggregated into a walkability index, *Walkability*. Standard errors are reported in parenthesis and are clustered at the county level; *significant at 10 percent, **significant at 5 percent. ***significant at 1 percent. All regressions include year and state fixed effects, individual and regional controls.

Table 3.A15: First Differences of Movers between counties: 1 year after move (coefficients on individual controls (cont'd from Table 3.5))

	Dependent Variables			
	Groups	Family	Neighbor	Community
Independent Variables				
age	-0.003	-0.037	-0.017	-0.204*
	(0.08)	(0.05)	(0.06)	(0.12)
young children	-0.400**	0.013	-0.042	-0.738***
	(0.15)	(0.09)	(0.09)	(0.14)
income	-0.144	-0.136**	-0.069	0.166
	(0.12)	(0.06)	(0.06)	(0.10)
education	0.001	0.052	0.002	-0.065
	(0.06)	(0.04)	(0.04)	(0.06)
home-owner	0.144	-0.036	-0.004	-0.005
	(0.11)	(0.07)	(0.07)	(0.11)
marital status (omitted category: married)				
single	-0.028	0.067	0.140	0.809***
	(0.20)	(0.10)	(0.10)	(0.21)
separated	0.029	-0.253	0.178	0.446*
	(0.38)	(0.17)	(0.15)	(0.26)
divorced	-0.450	-0.295*	0.279*	0.793**
	(0.29)	(0.16)	(0.15)	(0.25)
labor force status (omitted category: not in labor force)				
working	-0.211	-0.036	-0.084	-0.000
	(0.51)	(0.12)	(0.12)	(0.21)
unemployed	-0.243	0.062	-0.129	0.333
	(0.27)	(0.15)	(0.15)	(0.24)
maternity leave	-0.382	0.249	-0.302*	-0.433
	(0.38)	(0.16)	(0.18)	(0.36)
retired	0.266	0.492**	0.019	0.308
	(0.35)	(0.20)	(0.18)	(0.26)
Observations	1,320	1,320	1,320	1,320
Individuals	1,248	1,248	1,248	1,248

Notes: The above table presents the first-differenced results for the 'movers between counties' subset of the population. Results are shown for interactions with Groups, Family, Neighbors, and the Community, as a function of features of the built environment (columns (4) of Table 6) and individual controls. Point estimates are reported and standard errors are in parenthesis and are clustered at the county level; * significant at 10 percent, **significant at 5 percent. ***significant at 1 percent. All regressions include year and state fixed effects.

Table 3.A16: First Differences of Stayers: alternative specifications

	(1) Full Sample	(2) East only	(3) West only	(4) No Berlin	(5) Increase in AM	(6) Decrease in AM	(7) Top Increase in Density	(8) Top Decrease in Density
Panel A								
Dependent Variable	Change in Group Interactions							
Independent Variables								
Walkability	-0.013	-0.023	-0.012	-0.017	0.018	-0.024	-0.011	0.049
	(0.02)	(0.04)	(0.02)	(0.02)	(0.03)	(0.02)	(0.04)	(0.05)
Panel B								
Dependent Variable	Change in Family Interactions							
Independent Variable								
Walkability	-0.015	0.004	-0.018*	-0.014	-0.009	-0.024*	0.009	0.010
	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.04)	(0.03)
Panel C								
Dependent Variable	Change in Neighbor Interactions							
Independent Variable								
Walkability	0.006	0.014	0.005	0.006	0.016	0.001	0.023	-0.008
	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)	(0.01)	(0.04)	(0.02)
Panel D								
Dependent Variable	Change in Community Interactions							
Independent Variable								
Walkability	0.005	-0.014	0.014	0.003	-0.016	0.038	0.060	0.039
	(0.02)	(0.03)	(0.02)	(0.02)	(0.03)	(0.02)	(0.05)	(0.05)
Observations	2,464	677	1,787	2,395	1,051	1,405	222	327
Individuals	2,464	677	1,787	2,395	1,051	1,405	222	327

Notes: The above table presents the first-differenced results for the 'stayers' subset of the population, considering different samples characterized by specific regional changes. Differences are taken over 10 years of the panel data (1999 - 2009) and consider heterogenous effects across different regions in Germany. Standard errors are reported in parenthesis and are clustered at the county level; *significant at 10 percent, **significant at 5 percent. ***significant at 1 percent. All regressions include year and state fixed effects, individual and regional controls.

Table 3.A17: First Differences of Stayers: subsets of the population

Independent Variables	*Neighborhood Characteristics*											
	Shops				Primary School				Park			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Panel A												
Dependent Variable												
Group	-0.160	-0.026	0.117	-0.144	-0.352	0.046	-0.426	-0.114	0.324	-0.010	-0.152	0.101
Interactions	(0.41)	(0.11)	(0.30)	(0.15)	(0.33)	(0.12)	(0.35)	(0.17)	(0.35)	(0.10)	90.29)	(0.13)
Panel B												
Dependent Variable												
Family	-0.263	-0.011	0.062	-0.098	-0.013	-0.099*	-0.029	-0.134**	-0.171	0.008	-0.105	0.054
Interactions	(0.20)	(0.06)	(0.15)	(0.08)	(0.24)	(0.05)	(0.16)	(0.07)	(0.19)	(0.05)	(0.12)	(0.07)
Panel C												
Dependent Variable												
Neighbor	-0.133	0.006	-0.023	-0.064	-0.123	0.022	0.042	-0.017	0.196	0.012	-0.045	0.060
Interactions	(0.28)	(0.06)	(0.15)	(0.08)	(0.18)	(0.05)	(0.14)	(0.07)	(0.19)	(0.05)	(0.12)	(0.07)
Panel D												
Dependent Variable												
Community	0.118	-0.002	-0.198	-0.107	0.421	0.013	0.212	-0.074	0.103	-0.041	-0.114	-0.001
Interactions	(0.52)	(0.11)	(0.30)	(0.14)	(0.47)	(0.11)	(0.41)	(0.16)	(0.37)	(0.09)	(0.29)	(0.13)
Observations	105	1,316	212	1,006	105	1,316	212	1,006	105	1,316	212	1,006
Individuals	105	1,316	212	1,006	105	1,316	212	1,006	105	1,316	212	1,006
less than 40	•				•				•			
female only		•				•				•		
children			•				•				•	
retired/mat				•				•				•

Notes: The above table presents the first-differenced results for the ‘stayers’ subset of the population, considering further subsamples - individuals less than 40, females only, those with young children and those who are retired or on maternity leave. Differences are taken over 10 years of the panel data (1999 - 2009) and consider heterogenous effects across different regions in Germany. Standard errors are reported in parenthesis and are clustered at the county level; *significant at 10 percent, **significant at 5 percent. ***significant at 1 percent. All regressions include year and state fixed effects, individual and regional controls.

Table 3.A18: First Differences of Movers *between* Counties: alternative specifications

	(1) Full Sample	(2) Foreign	(3) German	(4) Increase in AM	(5) Decrease in AM	(6) Top Increase in Density	(7) Top Decrease in Density
Panel A							
Dependent Variable	Change in Group Interactions						
Independent Variables							
Walkability	0.022	0.230**	-0.004	-0.014	-0.027	0.099	0.036
	(0.04)	(0.11)	(0.04)	(0.09)	(0.07)	(0.10)	(0.11)
Panel B							
Dependent Variable	Change in Family Interactions						
Independent Variables							
Walkability	0.012	-0.022	0.012	0.026	0.070*	-0.032	-0.065*
	(0.02)	(0.06)	(0.02)	(0.04)	(0.04)	(0.06)	(0.04)
Panel C							
Dependent Variable	Change in Neighbor Interactions						
Walkability	-0.029	-0.043	-0.025	-0.061	-0.006	0.009	-0.068
	(0.02)	(0.05)	(0.02)	(0.04)	(0.04)	(0.05)	(0.04)
Panel D							
Dependent Variable	Change in Community Interactions						
Independent Variables							
Walkability	0.020	0.189	-0.001	-0.051	-0.033	-0.139	-0.012
	(0.04)	(0.13)	(0.04)	(0.08)	(0.08)	(0.10)	(0.07)
Observations	1,320	186	1,134	611	704	280	375
Individuals	1,248	179	1,069	602	692	280	375

Notes: The above table presents the first-differenced results for the 'movers between counties' subset of the population, considering different samples characterized by specific regional changes. Differences are one year after moving compared to one year prior. and consider heterogenous effects across different regions in Germany. Standard errors are reported in parenthesis and are clustered at the county level; *significant at 10 percent, **significant at 5 percent. ***significant at 1 percent. All regressions include year and state fixed effects, individual and regional controls.

Table 3.A19: First Differences of Movers *between* Counties: subsets of the population

Independent Variables	*Neighborhood Characteristics*											
	Shops				Primary School				Park			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Panel A												
Dependent Variable												
Group	0.168	-0.010	0.321*	-0.297	0.243*	0.189	0.607***	-0.286	-0.024	-0.173	-0.096	0.509
Interactions	(0.14)	(0.13)	(0.19)	(0.34)	(0.14)	(0.13)	(0.17)	(0.33)	(0.14)	(0.13)	(0.21)	(0.34)
Panel B												
Dependent Variable												
Family	-0.054	-0.037	-0.038	-0.101	0.110*	0.141**	0.154*	-0.031	0.0222	0.084	-0.020	0.364
Interactions	(0.07)	(0.07)	(0.09)	(0.18)	(0.07)	(0.07)	(0.08)	(0.17)	(0.08)	(0.08)	(0.11)	(0.22)
Panel C												
Dependent Variable												
Neighbor	-0.042	0.122*	-0.172*	0.033	0.026	-0.052	0.002	-0.005	-0.030	-0.005	-0.048	-0.288
Interactions	(0.07)	(0.07)	(0.09)	(0.24)	(0.06)	(0.06)	(0.08)	(0.20)	(0.08)	(0.08)	(0.14)	(0.25)
Panel D												
Dependent Variable												
Community	-0.029	-0.146	0.005	0.118	0.120	0.218*	0.242	0.434	-0.198	0.027	-0.013	-0.375
Interactions	(0.13)	(0.12)	(0.19)	(0.30)	(0.13)	(0.12)	(0.17)	(0.34)	(0.14)	(0.12)	(0.16)	(0.31)
Observations	712	680	465	165	712	680	465	165	712	680	465	165
Individuals	682	646	446	160	682	646	446	160	682	646	446	160
less than 40	•				•				•			
female only		•				•				•		
children			•				•				•	
retired/mat				•				•				•

Notes: The above table presents the first-differenced results for the 'movers between counties' subset of the population, considering further subsamples - individuals less than 40, females only, those with young children and those who are retired or on maternity leave. Differences are taken one year after moving compared to one year prior. Standard errors are reported in parenthesis and are clustered at the county level; *significant at 10 percent, **significant at 5 percent. ***significant at 1 percent. All regressions include year and state fixed effects, individual and regional controls.

Bibliography

Andersson, Martin, Johan Klaesson, and Johan P. Larsson (2016), "How local are spatial density externalities? Neighbourhood effects in agglomeration economies." *Regional studies*, 50, 1082–1095.

Arzaghi, Mohammad and J. Vernon Henderson (2008), "Networking off Madison Avenue." *The Review of Economic Studies*, 75, 1011–1038.

Bartik, Timothy (1991), "Who benefits from state and local economic development policies?" *W.E. Upjohn Institute for Employment Research*.

Baum-Snow, Nathaniel and Fernando Ferreira (2015), "Causal inference in urban and regional economics." In *Handbook of Regional and Urban Economics* (Gilles Duranton, J. Vernon Henderson, and William C. Strange, eds.), volume 5 of *Handbook of Regional and Urban Economics*, 3 – 68, Elsevier.

Baum-Snow, Nathaniel and Daniel Hartley (2016), "Accounting for central neighborhood change, 1980-2010." *working paper*.

Bayer, Patrick, Fernando Ferreira, and Robert McMillan (2007), "A unified framework for measuring preferences for schools and neighborhoods." *Journal of Political Economy*, 115, 588–638.

Beaudry, Paul, Mark Doms, and Ethan Lewis (2010), "Should the personal computer be considered a technological revolution? Evidence from U.S. metropolitan areas." *Journal of Political Economy*, 118, 988–1036.

Blume, Lawrence E., William A. Brock, Steven N. Durlauf, and Yannis M. Ioannides (2010), "Identification of social interactions." *Institute for Advanced Studies, Vienna*.

Borck, Rainald (2007), "Consumption and social life in cities: Evidence from Germany." *Urban Studies*, 44, 2105–2121.

Bouston, Leah P. and Robert A. Margo (2013), "A silver lining to white flight? White suburbanization and african-american homeownership, 1940-1980." *Journal of Urban Economics*, 74, 71–80.

Brueckner, Jan K. and Ann G. Largey (2008), "Social interaction and urban sprawl." *Journal of Urban Economics*, 64, 18–34.

Brueckner, Jan K. and Stuart S. Rosenthal (2009), "Gentrification and neighborhood housing cycles: Will America's future downtowns be rich?" *The Review of Economics and Statistics*, 91, 725–743.

Brueckner, Jan K, Jacques-Francois Thisse, and Yves Zenou (1999), "Why is central Paris rich and downtown Detroit poor? An amenity-based theory." *European economic review*, 43, 91–107.

Burgess, Ernest (1925), "The growth of the city." In *The City*, University of Chicago Press.

Card, David, Alexandre Mas, and Jesse Rothstein (2008), "Tipping and the dynamics of segregation." *The Quarterly Journal of Economics*, 123, 177–218.

Costa, Dora L. and Matthew E. Kahn (2000), "Power couples: Changes in the locational choice of the college educated, 1940-1990." *The Quarterly Journal of Economics*, 115, 1287–1315.

Costa, Dora L. and Matthew E. Kahn (2003), "Understanding the american decline in social capital, 1952-1998." *Kyklos, International Review for Social Sciences*, 56, 17–46.

Couture, Victor and Jessie Handbury (2016), "Urban revival in America, 2000 to 2010." *working paper*.

De la Roca, Jorge and Diega Puga (2017), "Learning by working in big cities." *The Review of Economic Studies*, 84, 106–142.

Durlauf, Steven N. and Marcel Fafchamps (2005), "Social capital." In *Handbook of Economic Growth* (Philippe Aghion and Steven N. Durlauf, eds.), volume 1 of *Handbook of Economic Growth*, 1639 – 1699, Elsevier.

Edlund, Lena, Cecilia Machado, and Michaela Sviatchi (2015), "Bright minds, big rent: Gentrification and the rising returns to skill." *NBER Working Paper 21729*.

Eid, Jean, Henry G. Overman, Diego Puga, and Matthew A. Turner (2008), "Fat city: Questioning the relationship between urban sprawl and obesity." *Journal of Urban Economics*, 63, 385–404.

Eisensee, Thomas and David Strömberg (2007), "News droughts, news floods, and US disaster relief." *The Quarterly Journal of Economics*, 122, 693–728.

Ellen, Ingrid Gould, Keren Mertens Horn, and Davin Reed (2016), "Has falling crime invited gentrification?" In *Research Symposium on Gentrification and Neighborhood Change*, Federal Reserve Bank of Philadelphia.

Ferreira, Fernando and Joseph Gyourko (2011), "Anatomy of the beginning of the housing boom: U.S. neighborhoods and metropolitan areas, 1993-2009." *NBER Working Paper 17374*.

Ferreira, Fernando and Joseph Gyourko (2012), "Heterogeneity in neighborhood-level price growth in the United States, 1993-2009." *American Economic Review: Papers and Proceedings*, 102, 134–140.

Fiorillo, D. and F. Sabatini (2011), "Quality and quantity: The role of social interactions in individual health." *Social Science and Medicine*, 73, 1644–1652.

French, Sarah, Lisa Wood, Sarah Foster, Billie Giles-Corti, Lawrence Frank, and Vincent Learnihan (2013), "Sense of community and its association with the neighborhood built environment." *Environment and Behaviour*, 20, 1–21.

Fukuyama, Francis (1995), "Trust: The social virtues and the creation of prosperity." New York: Free Press.

Galster, George, Roger Andersson, Sako Musterd, and Timo M. Kauppinen (2008), "Does neighborhood income mix affect earnings of adults? New evidence from sweden." *Journal of Urban Economics*, 63, 858–870.

Gibbons, Steve, Henry Overman, and Eleonara Patacchini (2015), "Spatial methods." In *Handbook of Regional and Urban Economics* (Gilles Duranton, J. Vernon Henderson, and William C. Strange, eds.), volume 5 of *Handbook of Regional and Urban Economics*, Elsevier.

Glaeser, Edward L. and Joshua D. Gottlieb (2006), "Urban resurgence and the consumer city." *Urban Studies*, 43, 1275–1299.

Glaeser, Edward L and Joseph Gyourko (2005), "Urban decline and durable housing." *Journal of Political Economy*, 113, 345–375.

Glaeser, Edward L., David Laibson, and Bruce Sacerdote (2002), "An economic approach to social capital." *The Economic Journal*, 112, 437–458.

Guerrieri, Veronica, Daniel Hartley, and Erik Hurst (2012), "Within-city variation in urban decline: The case of Detroit." *American Economic Review: Papers and Proceedings*, 102, 120–126.

Guerrieri, Veronica, Daniel Hartley, and Erik Hurst (2013), "Endogenous gentrification and housing price dynamics." *Journal of Public Economics*, 100, 45–60.

Gyourko, Joseph, Christopher Mayer, and Todd Sinai (2013), "Superstar cities." *American Economic Journal: Economic Policy*, 5, 167–199.

Hawkley, Louise C. and John T. Cacioppo (2010), "Loneliness matters: A theoretical and empirical review of consequences and mechanisms." *Annals of Behavioural Medicine*, 40, 218–227.

Helsley, Robert W. and William C. Strange (2007), "Urban interactions and spatial structure." *Journal of Economic Geography*, 7, 119–138.

Helsley, Robert W. and Yves Zenou (2011), "Social networks and interactions in cities." *IZA Discussion Paper*, DP: 5506.

Hitzschke, Stephan (2011), "The optimal size of german cities an efficiency analysis perspective." *Darmstadt Discussion Papers in Economics*.

Hoyt, Homer (1933), "One hundred years of land values in Chicago." *University of Chicago Press*.

Jacobs, Jane (1961), *The Death and Life of Great American Cities*. Random House.

Katz, Lawrence F., Jeffrey R. Kling, and Jeffrey B. Liebman (2001), "Moving to opportunity in Boston: Early results of a randomized mobility experiment." *The Quarterly Journal of Economics*, 116, 607–654.

Lee, Sanghoon and Jeffrey Lin (2017), "Natural amenities, neighborhood dynamics, and persistence in the spatial income distribution." *Review of Economic Studies*, forthcoming.

Leyden, Kevin M. (2003), "Social capital and the built environment: The importance of walkable neighborhoods." *American Journal of Public Health*, 93, 1546–1551.

Ludwig, Jens, Greg J. Duncan, and Paul Hirschfield (2001), "Urban poverty and juvenile crime: Evidence from a randomized housing-mobility experiment." *The Quarterly Journal of Economics*, 116, 655–679.

Maas, Jolanda, Robert A. Verheij, Peter P. Groenewegen, Sjerp de Vries, and Peter Spreeuwenberg (2006), "Green space, urbanity and health: How strong is the relation?" *Journal of Epidemiology and Community Health*, 60, 587–592.

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

Mayock, Tom (2016), "Wages, housing prices, and commutes." *Real Estate Economics*, 44, 258–300.

Moretti, Enrico (2011), "Local labor markets." *Handbook of Labor Economics*, 4, 1237–1313.

Moretti, Enrico (2013), "Real wage inequality." *American Economic Journal: Applied Economics*, 5, 65–103.

NYU Furman Center (2014,2015,2016), "State of New York City's housing and neighborhoods." NYU Furman Center, NYU School of Law.

Olson, Mancur (1965), *The Logic of Collective Action*. Harvard University Press.

O'Sullivan, Arthur (2009), "Schelling's model revisited: Residential sorting with competitive bidding for land." *Regional Science and Urban Economics*, 29, 397–408.

Pitkin, Bill (2001), "Theories of neighborhood change: Implications for community development policy and practice." *Los Angeles*.

Putnam, Robert D. (2000), *Bowling Alone*. Simon and Schuster Paperbacks.

Roback, Jennifer (1982), "Wages, rents, and the quality of life." *Journal of Political Economy*, 90, 1257–1278.

Rosen, Sherwin (1979), "Wage-based indexes of urban quality of life." *Current issues in urban economics*, 3, 490.

Rosenthal, Stuart S. (2008), "Old homes, externalities, and poor neighborhoods: A model of urban decline and renewal." *Journal of Urban Economics*, 63, 816–840.

Rosenthal, Stuart S. and Stephen L. Ross (2015), "Change and persistence in the economic status of neighborhoods and cities." In *Handbook of Regional and Urban Economics* (Gilles Duranton, J. Vernon Henderson, and William C. Strange, eds.), volume 5 of *Handbook of Regional and Urban Economics*, 1047–1120, Elsevier.

Ross, C.E., J.R. Reynolds, and K.J. Geis (2000), "The contingent meaning of neighborhood stability for residents' psychological well-being." *American Sociological Review*, 65, 581–597.

Saiz, Albert (2010), "The geographic determinants of housing supply." *The Quarterly Journal of Economics*, 125, 12531296.

Sampson, Robert J. (2011), "Neighborhood effects, causal mechanisms, and the social structure of the city." *Analytical Sociology and Social Mechanisms*, 227–250.

Schelling, Thomas C. (1971), "Dynamic models of segregation." *Journal of Mathematical Sociology*, 1, 143–186.

Serafinelli, Michel (2017), "'Good firms, worker flows and local productivity." *working paper*.

Sethi, Rajiv and Rohini Somanathan (2004), "Inequality and segregation." *Journal of Political Economy*, 112, 1296–1321.

Wood, Lisa, Tya Shannon, Max Bulsara, Terri Pikora, Gavin McCormack, and Billie Giles-Corti (2008), "The anatomy of the safe and social suburb: An exploratory study of the built environment, social capital and residents' perceptions of safety." *Health and Place*, 14, 15–31.