Endogenous Gentrification and Housing Price Dynamics∗

Veronica Guerrieri†
University of Chicago and NBER

Daniel Hartley‡
Federal Reserve Bank of Cleveland

Erik Hurst§
University of Chicago and NBER

June 1, 2010

Abstract

In this paper, we explore differential changes in house prices across neighborhoods within a city to better understand the nature of house price dynamics across cities. First, we document in detail that there is substantial and systematic heterogeneity in house price dynamics within a city during city wide housing price booms and busts. Second, we propose a new model of within city house price dynamics that is consistent with the empirical facts. We assume that there is a positive neighborhood externality: people like to live close to richer neighbors. We show that there is an equilibrium where households fully segregate based on their income. In response to housing demand shocks, the model predicts that the poor neighborhoods on the boundary with the rich ones are the most price elastic. For example, when richer households enter a city, the rich neighborhoods will expand outward turning previous poor neighborhoods into richer neighborhoods. As the amenities in such neighborhoods increase, land prices in the neighborhood will increase accordingly. We refer to this process as gentrification. We then empirically test this new mechanism against other mechanisms that could explain within city house price differences. We find strong support for the existence of endogenous gentrification in explaining housing price dynamics within a city. Finally, we show that even after controlling for other important determinants of land prices, the endogenous gentrification mechanism is still important in explaining cross city differences in house price dynamics.

∗The authors would like to thank seminar participants at the Winter 2010 NBER EFG program meeting, Chicago, MIT, Rochester, Stanford, UCLA, University of Akron, Wharton, Wisconsin, and the Federal Reserve Banks of Atlanta, Boston, Chicago, Cleveland, Minneapolis, and St. Louis. We are particularly indebted to Gary Becker, Hoyt Bleakley, Morris Davis, Matt Kahn, Kevin Murphy, Jesse Shapiro, and Todd Sinai for their detailed comments on previous drafts of this paper. All remaining errors are our own. Guerrieri and Hurst would like to acknowledge financial support from the University of Chicago’s Booth School of Business. The views expressed herein are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of Cleveland or the Federal Reserve System.

†Contact: vguerrie@chicagobooth.edu
‡Contact: daniel.hartley@clev.frb.org
§Contact: Erik.Hurst@chicagobooth.edu
1 Introduction

It has been well documented that there are large differences in house price appreciation across US metropolitan areas.\footnote{See, for example, Davis, Ortalo-Magné, and Rupert (2007), Glaeser, Gyourko, and Nathanson (2008), Van Nieuwerburgh and Weill (2010), and Saiz (2010).} For example, according to the Case-Shiller Price Index, real property prices increased by over 100 percent in Washington DC, Miami, and Los Angeles between 2000 and 2006, while property prices appreciated by less than 10 percent in Charlotte, Denver, and Detroit during the same time period. Across all cities for which a Case-Shiller index exists, the standard deviation in real house price growth between 2000 and 2006 was 53 percent. Such variation is not a recent phenomenon. During the 1990s, the cross city standard deviation in house price growth was 28 percent.

While most of the literature has focused on trying to explain cross MSA differences in house price appreciation, we document that there are also substantial within city differences in house price appreciation. For example, between 2000 and 2006 residential properties in the Harlem neighborhood of New York City appreciated by over 130 percent while residential properties in midtown Manhattan only appreciated by 45 percent. The New York City MSA as a whole appreciated by roughly 80 percent during this time period. Such patterns are common in many cities. Using within city price indices from a variety of sources, we show that the average within MSA standard deviation in house price growth during the 2000 - 2006 period was roughly 20 percent. Similar patterns are also found during the 1990s. As is commonly discussed in the popular press, these large relative movements in property prices within a city during city wide property price booms are often associated with neighborhood gentrification. Returning to the Harlem example, a recent New York Times article discussed how the composition of Harlem residents has been changing rapidly over the period when Harlem house prices were increasing at a relatively faster rate than other New York neighborhoods.\footnote{See the article “No Longer Majority Black, Harlem Is In Transition” from the January 5th, 2010 New York Times.}

Our goals in this paper are fourfold. First, we set out to document a new set of facts about the nature of within city house price movements. We show that there are substantial differences in house price growth within a city during periods of city-wide house price booms and busts. The house price appreciation for the city as a whole is just a composite of the house price movements within all the neighborhoods of the city. Therefore, understanding the movements in house prices across neighborhoods within a city is essential to understand house price movements at the city (or MSA) level.\footnote{Often, in our exposition, we use the terms “city” and “MSA” interchangeably. However, when we conduct the empirical work below, we will be much more explicit about whether or not we are discussing within city price movements or within MSA price movements.} Our primary data source is the Case-Shiller zip code level repeat sales housing price index for a large cross section of U.S. cities. We show that variation in house prices across neighborhoods within a city is almost as large as the variation in house prices across cities. In addition, we show that the there is a systematic pattern to the variation. For example, during city wide housing booms (busts), it is the poorer neighborhoods that experience the highest property price appreciations (depreciations). In particular, it is the poor neighborhoods that are
on the boundary with the rich neighborhoods that are the most price responsive within the city during housing price booms. These facts are stable during the housing price booms that occurred during the 1980s, 1990s, and 2000s.

Our second goal in the paper is to develop a model of neighborhood gentrification that generates within city house price movements consistent with the data. We propose a spatial equilibrium model of a linear city populated by poor and rich households with an elastic aggregate supply of houses. Our key assumption is the presence of a positive neighborhood externality: households’ utility is increasing in the average income of the neighbors. Although, we do not explicitly model the direct mechanism for the externality, we have many potential channels in mind. For example, crime rates are lower in richer neighborhoods. If households value low crime, individuals will prefer to live in wealthier neighborhoods. Likewise, the quality and extent of public goods may be correlated with the income of neighborhood residences. For example, school quality - via peer effects, parental monitoring or direct expenditures - tends to increase with neighborhood income. Finally, if there are increasing returns to scale in the production of desired neighborhood amenities (number and variety of restaurants, easier access to service industries such as dry cleaners, movie theaters, etc.), such amenities will be more common as the income of one’s neighbors increases. Although we do not take a stand on which mechanism is driving the externality, our preference structure is general enough to allow for any story that results in higher amenities being provided in higher income neighborhoods.4

We show that, with limited assumptions, there is an equilibrium with full segregation, where the rich are concentrated all together and the poor live at the periphery. Given the externality, households are willing to pay more to live closer to rich neighbors. Poor households who benefit less from the externality are less willing to pay high rents to live in the rich neighborhoods, so in equilibrium they live farther from the rich. House prices achieve their maximum in the richer neighborhoods and decline as we move away from them, to compensate for the lower level of the externality. For the neighborhoods that are far enough from the rich, there is no externality and house prices are equal to the marginal cost of construction.

A few stark predictions come from this simple model. The main implication is that an unexpected permanent housing demand shock (e.g., an increase in the city average income) will cause the rich to expand into areas previously occupied by the poor. We refer to this phenomenon as gentrification. As this happens, house prices in the gentrified neighborhoods are driven up due to the neighborhood externality. This implies that, during a city-wide house price boom, low price neighborhoods appreciate at a faster rate than high price neighborhoods. In particular, the poor neighborhoods that are in close proximity to the rich neighborhoods are the ones that should respond the most. Also, our mechanism implies that unexpected permanent shocks to housing demand lead to permanent increases in house prices at the city

---

4The importance of urban density in facilitating local consumption externalities has been recently emphasized in the work of Voith (1999), Voith and Gyourko (2002), Glaeser, Kolko, and Saiz (2001), Becker and Murphy (2003) and Rossi-Hansberg, Sartre, and Owens III (2010). Our innovation is to embed these consumption externalities into a model of neighborhood development within a city and show how such preferences affect the reaction of house prices to housing demand shocks both across various types of neighborhoods and at the aggregate level.
level because of the increase in prices within the gentrified neighborhoods. This latter effect occurs even when the size of the city is completely elastic. In other words, our model generates cross city differences in house prices and house price appreciation rates that do not rely on traditional supply constraints (regulation that prevents building, the steepness of the land gradient, natural barriers, etc.) which have been emphasized by other authors recently.\(^5\) Formally, we show that average price growth is affected both by the size of the demand shock and by the particular shape of preferences, technology, and the income distribution within the city.

Our third goal of the paper is to test the implications of our model of neighborhood gentrification against other potential stories to explain differences in house price movements within a city. Historically, three prominent classes of models have been put forth to explain within city differences in land prices. First, cities are viewed as centers of production agglomeration, as in the classic work by Alonso (1964), Mills (1967), and Muth (1969). Households prefer to live close to the jobs located in the center business district, but space around the center business district is limited. This implies that land prices get bid up around the center business district. Individuals could live farther away from the jobs but with higher transportation costs. In these models, it is the proximity to the jobs that drive higher land prices. The second story to explain within city price differences is related to the models put forth by Rosen (1979) and Roback (1982).\(^6\) In these models, one neighborhood within the city is more desirable than another because of a fixed natural amenity (like proximity to the ocean or beautiful vistas) that is valued by individuals. Neighborhoods that are in close proximity to the natural fixed amenity have higher land prices, all else equal. In both these classes of models, a housing demand shock would also increase house prices mostly in the high-price neighborhoods, that is, the ones closer to the jobs or to the fixed amenities. Third, differences across areas within an MSA with respect to regulatory barriers to adjusting housing supply could lead to differences in land prices across those areas, as in Glaeser and Gyourko (2003) and Glaeser, Gyourko, and Saks (2005).

In our empirical work, we show that the neighborhoods that experience higher growth rates in property prices show signs of gentrification. In particular, the low price neighborhoods in close proximity to the initially high price neighborhoods which appreciate the most, experience large increases in median income, substantial declines in the poverty rate, and an influx of new residents. Using a regression framework, we show that all of these results persist even controlling for proximity to the jobs (commuting time of residence), proximity to fixed natural amenities like lakes, rivers and oceans, and differences in regulatory barriers. We are not arguing that these other stories are unimportant in explaining land price differences. We do, however, show that our mechanism persists and is empirically important in explaining the large within-city variation in house prices even after controlling for these other potential mechanisms.

\(^5\)See, for example, Saiz (2010) and the cites within.
\(^6\)The Rosen (1979) and Roback (1982) models were built to explain cross city variation in housing prices but can be naturally extended to explain within city variation in housing prices.
gentrification in the surrounding neighborhoods. To do this, we follow the procedure used by Bartik (1989) and Blanchard and Katz (1992) in that we use the initial industry mix of residents in the neighborhood to predict how incomes in a given neighborhood should have evolved going forward. Using this variation, we show that house prices in a given neighborhood respond strongly to the instrumented shocks to income of nearby neighborhoods, even after controlling for the instrumented income shock in their own neighborhood. Income shocks experienced by far away neighbors have no effect on house prices in a given neighborhood. Moreover, we show that neighborhoods whose neighbors experienced big income shocks actually gentrified. Again, we show that these results persist even after controlling for the traditional stories of proximity to jobs or proximity to fixed natural amenities.

As noted above, we show theoretically that the gentrification mechanism at the heart of our model can also help to explain property price differences across cities. In the last portion of our paper, we show that even controlling for changes in commuting time and controlling for cross-city differences in housing supply elasticities, changes in income have strong predictive power in explaining cross city differences in housing price growth. We highlight that even among cities where there are little to no barriers to adjusting supply, housing prices respond strongly to changes to income changes holding the distribution of commuting times within the city constant.

At this point, it might be useful to address why one would care whether changes in house prices were driven by the mechanism highlighted in our paper or one of the other mechanisms traditionally used to explain within city and cross city differences in house prices. First, we feel it is economically important to distinguish between the different mechanisms. In particular, depending on the question of interest, the welfare implications will differ substantially depending on which model is used to explain within city house price changes. For example, consider a city with a net outflow of richer residents. Many authors have pointed out that housing costs would fall sharply in a such a city (see, Moretti (2009) and Notowidigdo (2010) for recent examples). In these papers, the fall in house prices is typically considered as something that increases the welfare of the remaining residents because of the cheaper housing stock. However, in these types of papers, there is an implicit assumption that neighborhood amenities remain constant as housing prices decline. To the extent that amenities are endogenous, as in our model, such an assumption would be inappropriate. Therefore, we think it is important to understand local neighborhood externalities when exploring the welfare costs associated with changes in housing demand. Additionally, we feel that it is important to understand why housing prices are changing within a given area if one wants to trace out the causal effect of how changes in house prices affect other variables such as household consumption or household defaults. Many researchers treat such changes in house prices as exogenous. To the extent that the changes in house prices are correlated with changes in neighborhood residents or neighborhood amenities, the identifying assumption in many of these papers will be undermined. Lastly, in terms of local developers or city planners, the existence of neighborhood externalities expands the set of policies they may choose to adopt when pursuing urban development plans.
Finally, our paper contributes to many literatures. First, a few existing studies have looked at within city price movements. Most of these perform their within city analysis by comparing the appreciation rates of “high end” properties to the appreciation rate of “low end” properties. Two papers, however, look specifically at differences in house prices appreciation across zip codes within a metropolitan area. Case and Mayer (1996) look at differential movements in prices within cities of the Boston metro area between 1982 and 1992 while Case and Marynchenko (2002) look at differential trends in prices across different zip codes within the Boston, Chicago, and Los Angeles metro areas during the 1983 to 1993 period. No systematic relationships emerged from these studies. Our work complements this literature by systematically examining within city movements in house prices during housing price booms and housing price busts.

Second, our paper speaks to the large literature on the gentrification of urban areas. Recent work has discussed the role of the following in explaining gentrification: the increased consumption benefits from living in a city (Glaeser, Kolko, and Saiz (2001)), the age of a city’s housing stock (Rosenthal (2008) and Brueckner and Rosenthal (2008)) and direct public policy initiatives via community redevelopment programs (see Busso and Kline (2007) and Rossi-Hansberg, Sartre, and Owens III (2010)). Despite the broad literature on gentrification, very little work emphasizes the importance of spatial dependence - either theoretically or empirically - in predicting the spatial patterns of gentrification. Our work adds to this literature by showing how generic shocks to housing demand within a city can result in the gentrification of neighborhoods. However, there are a few notable exceptions. Brueckner (1977) finds that urban neighborhoods in the 1960s that were in close proximity to rich neighborhoods got relatively poorer between 1960 and 1970 as measured by income growth. Conversely, Kolko (2007) finds the opposite pattern in that during the 1990s, neighborhoods bordering richer neighborhoods grew faster than otherwise similar neighborhoods bordering poorer neighborhoods. Our addition to this literature is that we propose a model that can reconcile these facts and then formally tests the model’s predictions. During periods of declining housing demand in urban areas (like the suburbanization movement during the 1960s), the richer neighborhoods on border of the rich areas will be the first to contract. Conversely, during urban renewals (like what was witnessed during the 1990s), the poor neighborhoods bordering the richer neighborhoods will be the first to gentrify.

Lastly, and most importantly, our work adds to the growing literature showing that cities are not only centers of production agglomeration but also centers of consumption agglomeration. See, for example,
Glaeser, Kolko, and Saiz (2001). Individuals not only like to be around certain types of jobs, they also like to be around certain types of neighbors. The key to our model is that changes in shocks to housing demand like lower interest rates, changes in the skill premium, or even shocks to the productivity of a local industries, endogenously cause the gentrification process to occur. This process, as we show, has effects on the rate of house price appreciation for the city as a whole.

In summary, our paper shows that the existence of neighborhood externalities has important implications for the nature of real estate price dynamics across neighborhoods within a city and across cities. We conclude that such externalities need to be embedded in both theoretical and empirical models designed to explain both time series and cross sectional housing price dynamics.

2 A Motivating Example: Housing Prices in Chicago

Before we present the model, we want to illustrate that the mechanism we are proposing to explain within city and cross city differences in housing price dynamics is plausible. We do so by focusing on housing prices within the city of Chicago. As noted above, there are two prominent classes of models that the prior literature has focused on to explain differences in land prices within a city: land prices are higher in location closer to the business center and land prices are higher in neighborhoods featuring a fixed natural amenity. In this section, we show that these two explanations account for only some of the variation in house prices within Chicago.

We choose Chicago as our motivating example for three reasons. First, Chicago is a flat featureless plain that is bordered to the east by Lake Michigan. Aside from the lake, there is little other geographical variation that defines different parts of Chicago with respect to desirable fixed natural amenities. This allows us to easily control for proximity to the exogenous natural amenity which could also determine land prices within a city. Second, Chicago has sharp differences in housing/land prices between its north and south sides. As we show below, the large variation in land prices between the north and south sides is hard to reconcile with the commuting cost and fixed amenity explanations. Finally, we were able to easily collect the universe of housing price transactions for Chicago for the 2000 - 2008 period. For most of our empirical work in later sections, we will have to rely on price indices that were already created at the sub city (e.g. zip code) level by Case-Shiller and Zillow. We describe these data in Section 4. However, given that we have the underlying micro data for Chicago, we can create price indices at lower levels of aggregation which correspond more closely to the concept of a neighborhood.

We were able to access the entire transaction level housing price data for Chicago because the Chicago Tribune manages an online database which records the underlying deed data for all transacted properties

10The Chicago river flows from the southwest of Chicago merging into Lake Michigan in Chicago’s downtown (Loop) neighborhood. Highways and railroads also transverse Chicago. While these transportation networks often serve as neighborhood barriers, they comprise very little of the land mass within Chicago. In our empirical work below, we will be looking at housing prices within a broad Chicago neighborhood. Most of our identification is going to come from the fact that Chicago’s north side has much higher land prices than Chicago’s south side despite similar distances to the center business district and similar distances to Chicago’s lake front.
We downloaded the data directly from the Chicago Tribune web site. We were able to link the Chicago house price data to attributes about the house using information from the Cook County Tax Assessor’s office. Given this data, we computed our own hedonically adjusted price indices for all Chicago community districts. Chicago community districts are a collection of census tracts within Chicago. In terms of size, the Chicago community areas are smaller than zip codes and better match the concept of a neighborhood within Chicago. There are 77 officially defined Chicago community districts. We refer to this data as the Chicago Tribune Index.

The data appendix provides a detailed discussion of how we compute our price indices using this data. Our price indices do not rely on repeat sales for identification. Instead, we create a hedonically adjusted price index using property level observables that we merged in from the Cook County Tax Assessor’s office. To ensure we have enough property transactions, we compute our price index at the yearly level (as opposed to either the quarterly or monthly level used by Case Shiller and Zillow). As we show in the online robustness appendix, if we use our methodology and recompute our price index at the zip code level, it correlates highly with both the Case Shiller index and the Zillow index for Chicago over the same time period.

Figure 1A shows the variation in year 2000 house prices (in year 2000 dollars) across Chicago community districts. The figure groups Chicago neighborhoods into quintiles by the median predicted house price within the community areas using the hedonically adjusted Chicago Tribune data. The darker areas on the map indicate neighborhoods with higher median housing prices. Housing prices on the north side of Chicago are systematically much higher on average than housing prices on the south side of Chicago. In particular, the housing prices in the north-eastern part of Chicago which border Lake Michigan (top right) are much higher than property prices in the south-eastern part of Chicago which also borders Lake Michigan. Figure 1B shows the variation in year 2000 mean household income (in year 2000 dollars) across neighborhoods within Chicago. The mean household income data comes from the 2000 U.S. Census. Like Figure 1A, we group neighborhoods into quintiles based on median income. Figure 1B shows that incomes are also higher on the north side of Chicago relative to the south side with the highest incomes occurring in the north eastern neighborhoods.

Table 1 explores the importance of the traditional mechanisms discussed above in explaining house price differences across Chicago neighborhoods. In the first column of Table 1, we report the simple relationship between the log of housing prices in the Chicago neighborhoods (hedonically adjusted as

---

11 This data is the same underlying data that Case-Shiller and Zillow use to compute their city level and zip code level price indices.
13 As a result, our price indices are more akin to the price indices computed by Zillow as opposed to the price indices computed by Case Shiller. We discuss both of these price indices in greater detail in Section 4.
14 The results of this table are mostly illustrative. We realize that our measure of house prices used in this table is imperfect in that it does not control well for structure attributes. In our main empirical work below, we are going to use the Case-Shiller price index data to measure the change in house prices holding structure attributes fixed. Furthermore, if we redo our analysis at the zip code level and use the Zillow house price measures for average home value in the zip code, the results remain unchanged. We show these specifications in our online Robustness Appendix.
described above) on measures of average neighborhood commuting times and distance to the Chicago lake front. For the latter, we simply use the log distance (in miles) to the Chicago lake front. For commuting costs, we use two measures. First, we use the log of the average commuting time to work (in minutes) as reported by the households within the community area from the 2000 U.S. Census. Our second measure of commuting costs is the distance to the “Chicago Loop” which is the name for the center business district of Chicago. We define the Chicago Loop as the centroid of the loop community area. As seen from column 1 of Table 1, the traditional measures of within city house price differences (proximity to jobs and fixed amenities), by themselves, explain 59 percent of the variation in land prices across Chicago. Places with shorter commuting times and places closer to Lake Michigan have higher house prices.

The results in column 1 do not prove that these mechanisms by themselves are important. Small amenity differences across neighborhoods could lead to large differences in land prices across neighborhoods if households have a preference to live near richer households and richer households sort into the more desirable neighborhoods (see Becker and Murphy (2003)). In columns 2 and 3, we include the log mean income of household residents (from the 2000 U.S. Census) by itself as a regressor (column 2) and to the controls for commuting costs and distance to Lake Michigan (column 3). The results in these columns show that differences in income, by itself, explains roughly half the variation in housing prices across neighborhoods. Moreover, even when income is added to the controls for commuting costs and distance to Lake Michigan, it has a large effect in explaining house price differentials across neighborhoods. The incremental $R^2$ of including income to the controls in column 1 is roughly 10 percentage points. The point estimate of the coefficient suggests that an elasticity of neighborhood average house prices to neighborhood average income of 0.68. These results underlie the relationships shown in Figure 1 A and B. The north side of Chicago, which is much richer than the south side of Chicago, has higher land prices even though commuting times are roughly similar and the average distance to Lake Michigan is similar.

It is worth noting that the residents of the neighborhoods on the north side of Chicago are more likely to be white compared to the residents of the south side of Chicago. It is well documented that racial segregation is high within Chicago (see Card et al 2008). This is potentially important given that some authors have posited that preferences of whites to live away from blacks can cause differences in land prices within a neighborhood. Given that income and race are so highly correlated, it is potentially hard to tease out a preference for the race of one’s neighbors and a preference for the income of one’s neighbors. In the work that follows, we do two things to test that the income of one’s neighbors is an

---

15 The U.S. Census provides the fraction of households who report commuting times which reside in preset bins. For example, the fraction of household who report their commuting time is between 0-4 minutes, 5-9 minutes, etc. With respect to computing the average commuting time across all households in the zip code, we use the midpoint of each bin. This assumption is not too problematic given the range of each bin is quite small. Our results are also robust to using the median commuting time (as opposed to the mean). For a full discussion of how we converted the actual Census data into the mean time commuting within each zip code, see the Data Appendix.

16 See, for example, Muth (1969), Courant (1974), Rose-Ackerman (1975), Kain and Quigley (1975), and Courant and Yinger (1977). For a more recent example, see Bayer, Fang, and McMillan (2009). Likewise, both Schelling (1969) and Card, Mas, and Rothstein (2008) also show how the desirability to have neighbors of a given race could lead to racial segregation within a city.
important determinant of land prices above and beyond the race of one’s neighbors. First, in all of our empirical work in subsequent sections, we will control directly for the race of one’s neighbors when testing for whether or not the income of one’s neighbors matter. As seen from column (4) of Table 1, housing prices in Chicago are still higher in richer neighborhoods even after controlling for the racial composition of the neighborhood as well as commuting costs and the distance to Lake Michigan. Nothing we find in this paper suggests that a preference for the racial composition of one’s neighbors is not important. Instead, we pursue the converse and show that even after controlling for race, higher income neighborhoods have higher housing prices. Our second approach is to test directly for the importance of income in determining land prices comes. In Section 6, we instrument directly for predicted neighborhood changes in income (using the initial industry mix of the neighborhood) and examine the change in land prices in surrounding neighborhoods to that predicted income shock. This method lets us isolate directly shocks to income and their effect on land prices, controlling for the racial composition of the neighborhoods.

While the Chicago discussion is only meant to offer a motivating example, we still feel this is illuminating about the potential mechanisms that explain the variations in land prices within cities. The standard stories of differences in commuting costs and differences in distance to a fixed natural amenity (which is the distance to Lake Michigan for Chicago) explain only some of the variation in the level of house prices within Chicago. The north side of Chicago is much richer and it has much higher property prices despite similar proximity to Lake Michigan and similar commuting times relative to the south side of Chicago.17 It is hard to explain the sharp differences in land prices within Chicago without relying on some other story. As we have shown above, whatever that story is, it seems correlated with the income of both the residents within the neighborhood and the income of the residents in nearby neighborhoods. In the later sections, we take a much more systematic approach to test the implications of the model that we set forth.

3 Model

In this section, we develop a spacial equilibrium model of housing prices across neighborhoods within a city. The key ingredient of the model is a positive neighborhood externality: people like to live next to individuals with higher income. As noted in the introduction, we do not take a stand on the underlying micro foundations of such preferences, although we have many potential stories in mind. Additionally, when presenting the model, we abstract from both the notion of commuting costs and of fixed geographical amenities. We make this choice not because we think these traditional mechanisms are unimportant, but to highlight the implications of the our mechanism.

17 We do not have anything to say as to why richer households located on the north side of Chicago relative to the south. Some have proposed that the stock yards on the south side of Chicago created a disamenity (a bad odor) which made land less desirable on the south side relative to the north side in the early to mid 20th century. The stock yards closed forty years ago. Our mechanism could explain why differences in land prices could persist within Chicago after the elimination of such a negative amenity.
3.1 Set up

Time is discrete and runs forever. We consider a city populated by two types of infinitely-lived households: a continuum of rich households of measure $N^R$ and a continuum of poor households of measure $N^P$. Each period households of type $s$, for $s = R, P$, receive an exogenous endowment of consumption goods equal to $y^s$, with $y^P < y^R$.

The city is represented by the real line and each point on the line $i \in (-\infty, +\infty)$ is a different location.$^{18}$ Agents are fully mobile and can choose to live in any location $i$. Denote by $n^s_t(i)$ the measure of households of type $s$ who live in location $i$ at time $t$ and by $h^s_t(i)$ the size of the house they choose. In each location, there is a maximum space that can be occupied by houses normalized to 1,$^{19}$ that is,

$$n^R_t(i) h^R_t(i) + n^P_t(i) h^P_t(i) \leq 1 \text{ for all } i, t.$$ 

Moreover, market clearing requires

$$\int_{-\infty}^{+\infty} n^s_t(i) \, di = N^s \text{ for } s = R, P. \tag{1}$$

The key ingredient of the model is that there is a positive location externality: households like to live in areas where more rich households live. Each location $i$ has an associated neighborhood, given by the interval centered at $i$ of fixed radius $\gamma$. Let $H_t(i)$ denote the total space occupied by houses of rich households in the neighborhood around location $i$,$^{20}$ that is,

$$H_t(i) = \int_{i-\gamma}^{i+\gamma} h^R_t(j) \, n^R_t(j) \, dj. \tag{2}$$

Households have non-separable utility in non-durable consumption $c$ and housing services $h$. The location externality is captured by the fact that households enjoy more their consumption if they live in locations with higher $H_t(i)$. The utility of an household of type $s$ located in location $i$ at time $t$ is given by

$$u^s (c, h, H_t(i)),$$

where $u(.)$ is weakly concave in $c$ and $h$. For tractability, we assume that $u$ takes the following functional form: $u^s (c, h, H) = c^\alpha h^\beta (A + H)^{\delta^s}$, where $\alpha$, $\beta$, and $\delta^s$ are non negative scalars. Moreover, we assume that $\delta^R \geq \delta^P$, so that rich households who generate the externality benefit from it at least as much as poor households.

---

$^{18}$We choose to model the city as a line because it simplifies our analysis. The main implications of our model extend to a circular city as in Lucas and Rossi-Hansberg (2002).

$^{19}$Our notion of space is uni-dimensional: if there is need for more space to construct houses the city has to expand horizontally. We could enrich the model with a bi-dimensional notion of space, by allowing a more flexible space constraint in each location. For example, we could imagine some form of adjustment cost to construct in each location, so that in reaction to a demand shock the city can expand both in the horizontal and in the vertical dimension (our model being the extreme case with infinite adjustment cost). Our conjecture is that our mechanism would go through with this extension and we believe this is an interesting direction for future work.

$^{20}$An alternative is to define the neighborhood externality $H_t(i)$ as the measure of rich households living in the neighborhood around location $i$. However, this would make the model less tractable without affecting the substance of the mechanism.
On the supply side, there is a representative firm who can build housing in any location \( i \in (-\infty, +\infty) \). There are two types of housing: rich houses (type \( R \)) and poor houses (type \( P \)). Each type of household only demands houses of his own type. The marginal cost of building houses of type \( s \) is equal to \( C_s \), with \( C_R \geq C_P \). If the firm wants to convert houses of type \( \bar{s} \) into houses of type \( s \), he has to pay \( C_s - C_{\bar{s}} \).

The (per square foot) price of a house for households of type \( s \) in location \( i \) at time \( t \) is equal to \( p_s^t(i) \).

Hence there is going to be construction in any empty location \( i \) as long as \( p_s^t(i) \geq C_s \). Moreover, if the firm wants to construct a house of type \( s \) in a location occupied by a house of type \( \bar{s} \), he has to pay the converting cost and the additional cost of convincing households of type \( \bar{s} \) to leave. Hence, there is going to be construction of houses of type \( s \) in any location occupied by agents of type \( \bar{s} \) if \( p_s^t(i) \geq C_s - C_{\bar{s}} + p_{\bar{s}}^t(i) \).

Finally, there is a continuum of risk-neutral competitive intermediaries who own the houses and rent them to the households. The intermediaries are introduced for tractability. If we allowed the households to own their houses, nothing would change in steady state, but the analysis of a demand shock would be more complicated.\(^{21}\) The (per square foot) rent for a house of type \( s \) in location \( i \) at time \( t \) is denoted by \( R_s^t(i) \).

As long as the rent in location \( i \) at time \( t \) is positive, the intermediaries find it optimal to rent all the houses in that location. Also, for simplicity, assume that houses do not depreciate. Competition among intermediaries requires that for each location \( i \) the following arbitrage equations hold:

\[
p_s^t(i) = R_s^t(i) + \left( \frac{1}{1 + r} \right) p_{s, t+1}^s(i) \text{ for all } t, i, s. \tag{3}
\]

### 3.2 Equilibrium

An equilibrium is a sequence of rent and price schedules \( \{ R_R^t(i), R_P^t(i), p_R^t(i), p_P^t(i) \}_{i \in R} \) and of allocations \( \{ n_R^t(i), n_P^t(i), h_R^t(i), h_P^t(i) \}_{i \in R} \) such that households maximize utility, the representative firm maximizes profits, intermediaries maximize profits, and markets clear.

Because of full mobility, the household’s maximization problem reduces to a series of static problems. The problem of households of type \( s \) at time \( t \) is simply

\[
\max_{c, h, i \in I_t^s} c^\alpha h^\beta [A + H_t(i)]^{\delta s},
\]

s.t. \( c + hR_s^t(i) \leq y^s \),

where households take as given the function \( H_t(i) \), the rent schedule \( R_s^t(i) \), and the set \( I_t^s \) of locations where houses for type-\( s \) households are available. Hence, conditional on choosing to live in location \( i \) at time \( t \), the optimal house size is

\[
h_s^t(i) = \frac{\beta}{\alpha + \beta R_s^t(i)} y^s \text{ for all } t, s, i \in I_t^s. \tag{4}
\]

Households choose to live in bigger houses in neighborhoods where the rental price is lower and, conditional on a location, richer households choose bigger houses. Given that households are fully mobile, it must be

\(^{21}\)When the economy is hit by a positive demand shock, we will show that house prices appreciate by different amount in different locations. If households own their houses this would introduce an additional source of heterogeneity in wealth which would complicate the analysis.
that at each point in time, the equilibrium rents in different locations make them indifferent. In particular, agents of type \( s \) have to be indifferent among living in different locations where houses of their type are available at time \( t \), that is, in all \( i \in I^s_t \). Then it must be that
\[
U^s_t(i) \equiv \alpha^s \beta \left( \frac{y^s}{\alpha + \beta} \right)^{\alpha + \beta} \frac{[A + H_t(i)]^s}{R^s_t(i)^\beta} = \bar{U}^s_t \text{ for all } t, s, i \in I^s_t.
\]
This, in turns, requires that
\[
R^s_t(i) = K^s [A + H_t(i)]^\frac{s}{s} \text{ for all } t, s, i \in I^s_t,
\]
for some constant \( K^s \). This expression is intuitive, as rents must be higher in locations with a stronger externality. Moreover, rich households who are more affected by the location externality, given \( \delta R \geq \delta P \), are willing to pay higher rents for higher degrees of externality.

**Proposition 1.** If \( \delta R \geq \delta P \), there exists an equilibrium with full segregation.

Let us construct an equilibrium with full segregation, where the rich households are concentrated in the city center, while the poor households live at the periphery of the city.\(^{22}\) As a normalization, let us choose point 0 as the center of the city. It follows that \( I^R_t = [-I_t, I_t] \) and \( I^P_t = [-\bar{I}_t, -I_t) \cup (I_t, \bar{I}_t] \), for some \( \bar{I}_t > I_t > 0 \). In this model, both the size of rich neighborhoods, \( I_t \), and the size of the city, \( \bar{I}_t \), are equilibrium objects. Given that such an equilibrium is symmetric in \( i \), from now on, we can restrict attention to \( i \geq 0 \).

Given that rich households live in locations where there are no poor, it must be that \( h^R_t(i) n^R_t(i) \) is either equal to 1 or to 0 and is equal to 1 for all \( i \in [0, I_t] \). Then, we can easily derive the function \( H_t(.) \) as follows:
\[
H_t(i) = \begin{cases} 
2\gamma & \text{for } i \in [0, I_t - \gamma] \\
\max \left\{ \gamma + I_t - i, 0 \right\} & \text{for } i \in (I_t - \gamma, I_t].
\end{cases}
\]
That is, neighborhoods close to the city center are fully developed and enjoy the maximum degree of externality, while the farther a location is from the center the smaller the strength of the externality. Figure 2 shows the externality \( H_t(i) \) for a given \( t \) as a function of the location. If \( \bar{I}_t > I_t + \gamma \), there are going to be locations at the margin of the city where the externality has zero effect. From now on, we assume that the measure of poor households, \( N^P \), is sufficiently large so that \( \bar{I}_t > I_t + \gamma \).

Combining (6) and (7), we obtain
\[
K^R = R^R_t(I_t) (A + \gamma)^{-\frac{\beta P}{\beta}} \text{ and } K^P = R^P_t(\bar{I}_t) A^{-\frac{\beta P}{\beta}},
\]
so that we can rewrite the rent schedules as
\[
R^R_t(i) = R^R_t(I_t) \left( 1 + \frac{\min \{ \gamma, I_t - i \}}{A + \gamma} \right)^{\frac{\beta P}{\beta}} \text{ for } i \in [0, I_t],
\]
\[
R^P_t(i) = R^P_t(\bar{I}_t) \left( 1 + \frac{\max \{ \gamma + I_t - i, 0 \}}{A} \right)^{\frac{\beta P}{\beta}} \text{ for } i \in (I_t, \bar{I}_t].
\]
\(^{22}\)If there was a location with construction of type \( s \) and no type \( s \) households living there, the intermediaries would be willing to decrease the rent to 0 inducing households of type \( s \) to move into that location.
\(^{23}\)There may be other equilibria with full segregation and more centers of agglomeration. As long as these centers are far enough from each other, the implications in terms of house prices are isomorphic to our equilibrium with one center.
From the optimizing behavior of the representative firm, it must be that the price of a poor house at the boundary of the city is equal to the marginal cost \( C^P \). Moreover, the price of a rich house at the boundary of the rich neighborhoods must be equal to the price of a poor house, which is the compensation needed to vacate poor households living there, plus the additional cost of transforming a poor house in a rich one. This implies that \( p^P_t (\bar{I}_t) = C^P \) and \( p^R_t (I_t) = p^P_t (I_t) + C^R - C^P \). In equilibrium prices are constant over time and hence arbitrage conditions (3) require that for each location \( i \in \mathcal{I}_t^s \) prices satisfy

\[
p^s_t (i) = \frac{1 + r}{r} R^s_t (i) \quad \text{for all } t, i, s.
\] (11)

Combining this conditions we obtain

\[
R^P_t (\bar{I}_t) = \frac{r}{1 + r} C^P \quad \text{and} \quad R^R_t (I_t) = R^P_t (I_t) + \frac{r}{1 + r} (C^R - C^P),
\] (12)

where, from (6) and (8), we have

\[
R^P_t (I_t) = \frac{r}{1 + r} C^P \left( \frac{A + \gamma}{A + \max \{ \gamma + I_t - \bar{I}_t, 0 \}} \right)^{\frac{\alpha}{\beta}}.
\] (13)

Combining these last two expressions with (9), (10), and (11) allows us to determine the rent and the price schedules as a function of \( I_t \) and \( \bar{I}_t \) only. Figure 2 also shows the shape of the price schedule as a function of the location.

In our full segregation equilibrium, the rich households are concentrated in the city center, while the poor are located at the periphery. Moreover, equilibrium prices reflect the fact that locations that are further away from the center of the rich enclave and which are closer to the space occupied by poor households are less appealing. In particular, prices are the highest in the center of the rich neighborhoods. As we move away from the center, prices start declining because the space in the neighborhood occupied by rich households goes down. This segregation equilibrium is sustained by the fact that the poor are unwilling to lower their non housing consumption by paying higher rent to get the larger neighborhood externality.

To complete the characterization of the equilibrium, we need to determine the size of the city, \( \bar{I}_t \), and the size of the rich neighborhoods, \( I_t \). Using market clearing (1) together with the optimal housing size (4) and the fact that \( \mathcal{I}_t^R = [\bar{I}_t, I_t] \) and \( \mathcal{I}_t^P = [-\bar{I}_t, -I_t] \cup (I_t, \bar{I}_t) \), we obtain the following expressions for \( I_t \) and \( \bar{I}_t \):

\[
I_t = \gamma + (A + 2\gamma)^{-\frac{2\rho}{\alpha + \delta}} \left\{ \frac{\alpha}{\alpha + \delta} \frac{N^R y^R}{2K^R} - \frac{\beta}{\beta + \delta_R} \left[ (A + 2\gamma) \frac{3 + \delta_R}{\beta} - (A + \gamma) \frac{3 + \delta_R}{\beta} \right] \right\}
\] (14)

\[
\bar{I}_t = I_t + \gamma + A^{-\frac{2\rho}{\alpha + \delta}} \left\{ \frac{\alpha}{\alpha + \delta} \frac{N^P y^P}{2K^P} - \frac{\beta}{\beta + \delta_P} \left[ (A + \gamma) \frac{3 + \delta_P}{\beta} - A \frac{3 + \delta_P}{\beta} \right] \right\}.
\] (15)

As intuition suggests, the rich neighborhoods are more developed when \( N^R \) (the number of rich people) or \( y^R \) (the income of rich people) are higher, and when the marginal cost of construction \( C^R \) or the interest rate \( r \) are lower. Moreover, the city overall is bigger when the rich neighborhoods are more developed,
when there are more poor households or poor are richer, higher $N^P$ or $y^P$, and when the marginal cost of construction $C^P$ or the interest rate are lower.

Finally, to complete the construction of the equilibrium, we have to check that the households choose their location optimally, that is, we have to check that the rich would not prefer to move to a poor neighborhood and vice versa. More precisely, we need to prove that

$$U^R_t(i) \leq \bar{U}^R_t \text{ for all } i \in [I_t, \bar{I}_t]$$

$$U^P_t(i) \leq \bar{U}^P_t \text{ for all } i \in [0, I_t]$$

where $U^s_t(i)$ is defined in expression (6). In the Appendix, we show that both these conditions are satisfied if $\delta^R \geq \delta^P$, completing the proof of the Proposition.

### 3.3 Demand shock

We are now interested in analyzing how house prices, both at an aggregate and at a disaggregate level, react to shocks to the demand for housing. We will do so, by focusing on the equilibrium with full segregation we have constructed in the previous section.

In equilibrium, the aggregate price level is given by

$$P_t = \frac{2}{I_t} \int_0^{I_t} p^R_t(i) \, di + \frac{2}{I_t - \bar{I}_t} \int_{I_t}^{\bar{I}_t} p^P_t(i) \, di,$$

where, from the analysis in the previous section,

$$p^R_t(i) = C^P \left(1 + \frac{\gamma}{A}\right)^{\frac{\delta^P}{\beta}} + C^R - C^P \left(1 + \min \left\{ \frac{\gamma, I_t - i}{A + \gamma} \right\} \right)^{\frac{\delta^P}{\beta}} \text{ for } i \in [0, I_t], \quad (16)$$

$$p^P_t(i) = C^P \left(1 + \frac{\max \{\gamma + I_t - i, 0\}}{A}\right)^{\frac{\delta^P}{\beta}} \text{ for } i \in (I_t, \bar{I}_t], \quad (17)$$

with $I_t$ and $\bar{I}_t$ given by (14) and (15).

For concreteness, we analyze the economy’s reaction to an income shock, but the price dynamics are equivalent if we consider a shock to the interest rate (lower $r$) or an influx of richer households to the city (i.e., higher $N^R$). Imagine that at time $t + 1$ the economy is hit by an unexpected and permanent increase in income. Let us assume that the income endowment of both rich and poor increase proportionally, that is, $y^s_{t+1} = \phi y^s_t$ with $\phi > 1$ for $s = P, R$. We now show that the aggregate level of house prices permanently increases and prices in locations with a higher initial price level typically react less than prices in locations where houses are cheaper to start with and which are closer to the expensive neighborhoods. As income increases, both rich and poor wants to buy bigger houses and the city starts expanding, that is, both $I_t$ and $\bar{I}_t$ increase and the rich households start expanding in poor neighborhoods. This is what we refer to as endogenous gentrification. The house prices in gentrified neighborhood are driven up due to our externality.
Let us define the function $g_t(\cdot) : [C^P, \bar{p}] \mapsto [1, \infty)$, where $g_t(p)$ denotes the average gross growth rate between time $t$ and $t+1$ in locations where the initial price is equal to $p$, that is,

$$g(p) = E_{t+1}\left[ \frac{p_t+1(i)}{p_t(i)} | p_t(i) = p \right].$$

The next proposition shows that after an unexpected permanent demand shock, the aggregate price level permanently increases and the price growth rate is higher in locations with initial lower price level, whenever prices are higher than the minimum level.\(^{24}\)

**Proposition 2.** Imagine that at time $t+1$ the economy is hit by an unexpected and permanent increase in income, that is, $y_{s}^{t+1} = \phi y_s^t$ with $\phi > 1$ for $s = P, R$. Then there is a permanent increase in the aggregate price level $P_t$, and

$$E_{t+1}\left[ \frac{p_t+1(i)}{p_t(i)} | p_t(i) = \bar{p} \right] < E_{t+1}\left[ \frac{p_t+1(i)}{p_t(i)} | p_t(i) < \bar{p} \right].$$

Moreover, if the shock is large enough, $g_t(p)$ is non-increasing in $p$ for all $p > C^P$.

Figure 3 illustrates the reaction of house prices to a positive demand shock (a proportional increase in income) in different locations. Given that the city is symmetric, the figure represents only the positive portion of the real line. One can notice that both the size of the city $I_t$ and the size of the rich neighborhoods $I_t$ expand, while prices remain constant at the two extremes: in the richest locations in the center of the city and far enough from the rich neighborhoods. Most important, prices strictly increase in the rich neighborhoods not fully developed yet and, even more, in the poor neighborhoods physically closer to the rich city center. Clearly, this makes the aggregate level of prices in the city increase permanently. Figure 4 shows the price growth rate as a function of the initial price level in different locations together with the OLS regression of price appreciation on initial price level. If the shock is big enough and positive, our model delivers a negative relationship between initial prices and the subsequent housing price growth. Moreover, according to our model, this negative relationship is driven by the locations with low initial price level that are closer to the rich locations. After a demand shock these are the locations that are going to be gentrified, that is, where the rich will move in, increasing the degree of the externality and hence driving prices up.

The next proposition shows the main implication of our model: among the locations with initial level of price equal to $C^P$, the ones that appreciate the most are close to the city center where the rich live. The proof of the proposition is in the Appendix.

**Proposition 3.** Imagine that at time $t+1$ the economy is hit by an unexpected and permanent increase in income, that is, $y_{t}^{s+1} = \phi y_t^s$ with $\phi > 1$ for $s = P, R$. Then

$$\frac{d(p_t+1(i)/p_t(i))}{di} < 0 \text{ for } p_t(i) = C^P.$$
In the work that follows, this proposition will lie at the heart of our empirical work. Among the poor neighborhoods, it is the poor neighborhoods in close proximity to the richer neighborhoods that should appreciate the most during a period of a city wide housing demand shock. This proposition also underlies the variation in appreciation rates among the poorer neighborhoods. Lastly, we will also test for the mechanism leading to the house price appreciation. The poorer neighborhoods next to the richer neighborhoods experience large price increases because they gentrify. Richer households expand into the neighborhood thereby increasing the desirability of being in those neighborhoods. The empirical work that follows shows strong support for all of these predictions.

Finally, we will show that in the data the slope of the regression line in Figure 4 across cities is steeper the higher is the average growth rate of prices at the city level. We consider two different stories which can rationalize this relationship within our model. First, it could be that different cities are hit by demand shocks of different size. Second, it could be that different type cities are hit by a common demand shock. In both cases, our model generate a positive relationship between the degree of gentrification and the size of the city-wide price boom.

### Proposition 4

*Imagine that at time $t + 1$ the economy is hit by an unexpected and permanent increase in income, that is, $y_{t+1}^s = \phi y_t^s$ with $\phi > 1$ for $s = P, R$. Then the growth rate in the aggregate price level is larger the larger is the increase $\phi$ and, if the shock is large enough, $d^2 g_t(p)/d\phi \geq 0$ for all $p > C^P$ where the derivative is well-defined.*

This proposition shows that if two identical cities are hit by demand shocks of different sizes, the one hit by the larger shock is going to feature both higher aggregate price growth rate and more price convergence due to a higher degree of gentrification.

### Proposition 5

*Consider two cities, $A$ and $B$, where $y_A^R \geq y_B^R$ and $y_A^P \geq y_B^P$ with at least one strict inequality. If at time $t + 1$ they are both hit by an unexpected and permanent increase in income of the same proportion and large enough, then the growth rate in the aggregate price level is larger in city $A$ and $g'_A(p) \leq g'_B(p)$ for all $p > C^P$ where the derivative is well-defined.*

This proposition shows that in two cities with different levels of ex-ante income hit by the same demand shock, house prices react differently. In particular, if the shock is large enough, the initially richer city is the one that features both higher aggregate price growth rate and higher within city house price convergence. In the last section of the paper, we will explore the extent to which our mechanism explains cross city differences in housing price appreciation rates during the 1990s and 2000s.

### 4 Data on Within City House Prices

In the rest of the paper, we will explore the empirical implications of the theory outlined in the proceeding section. In this section, we discuss the primary data sources we use to measure house prices (and house price changes) across cities and across neighborhoods within cities.
For the remainder of the paper, our primary measure of house prices across neighborhoods within a city during different time periods comes from the Case-Shiller zip code level price indices. The Case-Shiller indices are calculated from data on repeat sales of pre-existing single-family homes. The benefit of the Case-Shiller index is that it provides consistent constant-quality price indices for localized areas within a city or metropolitan area over long periods of time. Most of the Case-Shiller zip code-level price indices go back in time through the late 1980s or the early 1990s. The data was provided to us at the quarterly frequency and our most recent data is for the fourth quarter of 2008. As a result, for each metro area, we have quarterly price indices on selected zip codes within selected metropolitan areas going back roughly 20 years.

There are a few things that we would like to point out about the Case-Shiller indices. First, the Case-Shiller zip code level indices are only available for certain zip codes in certain metropolitan areas. We focus on the 30 center cities for which the Case-Shiller zip code indices exist. Our sample includes only those cities where an index is provided for at least 14 zip codes in either 1990 and 2000 or 2000 and 2006.

Second, we only use information for the zip codes where the price indices were computed using actual transaction data for properties within the zip code. Some of the zip code price indices computed as part of the available Case-Shiller data use imputed data or data from some of the surrounding zip codes. We exclude all such zip codes from our analysis. As a result, the Case-Shiller zip codes that we use in our analysis do not cover the universe of zip codes within a city or metropolitan area. For example, only about 50 percent of the zip codes in the city of Chicago have housing price indices computed using actual transaction data. The fraction in other cities (like Charlotte) is closer to 100 percent. The zip codes within the cities that tend to have either missing or imputed zip code housing price indices are the zip codes where there are very few housing transactions or where most of the housing transactions are for non-single family homes. As we show below, the Case-Shiller zip code price indices correlates very highly with the zip code price indices computed by other methods.

Third, the Case-Shiller indices have the goal of measuring the change in land prices by removing structure fixed effects using their repeat sales methodology. However, this methodology only uncovers changes in land prices if the attributes of the structure remain fixed over time. If households change the attributes of the structure via remodeling or through renovations, the change in the house prices uncovered by a repeat sales index will be a composite of changes in land prices and of improvements to the housing structure. Those who compute the Case-Shiller index are aware of such problems and, albeit

---

25 The zip code indices are not publicly available. Fiserv, the company overseeing the Case-Shiller index, provided them to us for the purpose of this research project. The data are the same as the data provided to other researchers studying local movements in housing prices. See, for example, Mian and Sufi (2009)

26 These cities include: Akron, Atlanta, Baltimore, Charlotte, Chicago, Cincinnati, Columbus (OH), Dayton, Denver, Fresno, Jacksonville, Las Vegas, Los Angeles, Memphis, Miami, New York, Oakland, Orlando, Philadelphia, Phoenix, Portland (OR), Raleigh, Sacramento, San Diego, San Jose, Scottsdale, Seattle, St. Paul, Tampa, and Toledo.

27 New York City is an outlier. Only 9 zip codes within New York City had prices indices computed using actual (non-imputed) data. To overcome this problem, we used a supplementary data set to explore the patterns in New York City. We discuss this data below.
imperfectly, take steps to minimize the effect of potential remodeling and renovations when computing the index. In particular, the index puts a lower weight on repeated sales transactions where the change in price is likely to reflect changes in the housing structure, that is, when the change in price was either disproportionately large or disproportionately small. Additionally, the index excludes all properties where the property type changed (i.e., a single family home is converted to condos) and it excludes all properties where the home sells within six months after a purchase. These properties tend to follow the redevelopment of the property. Also, all repeated sales transactions are weighted based on the time interval between first and second sales. Sales pairs with longer time intervals are given less weight than sales pairs with shorter intervals. The assumption is that if a sales pair interval is longer, then it is more likely that a house may have experienced a physical change.

Given the discussion above, we see that there are two main limitations to the Case-Shiller index. First, the Case-Shiller index does not always exist for all zip codes within a city. Second, the index may not be perfectly capturing changes in land prices because it cannot perfectly control for unobserved renovations or remodeling. As a result, one may wonder whether our results are biased in some way from these limitations.

We take both concerns seriously and to address them we augment our analysis with data from a variety of other sources. To address the first issue, as discussed above, we have compiled transaction level data on the universe of all residential housing transactions for Chicago. We can compute reliable zip code estimates for Chicago if we group transactions at a yearly frequency as opposed to a quarterly frequency. As we show in the online appendix, the results for Chicago using the universe of zip codes is nearly identical to the results for Chicago that we get using the sub-sample of zip codes for which we have a Case Shiller index. Additionally, to explore the patterns within New York City (NYC) in much greater detail, we use the Furman Center repeat sales index which covers all of NYC. The Furman data uses NYC community districts as its level of aggregation as opposed to zip codes. There are 59 community districts in NYC which represent clusters of several neighborhoods. The Furman data for NYC extend back to 1974. The benefit of the Furman data is that it gives us extensive coverage of all NYC neighborhoods over a long time period. Like the Case-Shiller data, it uses deed data to compute a repeat sales price indices. However, given the larger level of aggregation, it is better able to compute reliable price indices. Again, the patterns we find including the universe of neighborhoods in NYC is similar to the patterns we document for other cities using the Case Shiller indices.

To address the second issue, we augment our results using the Zillow Home Value Indices at the zip code level. Zillow does not use a repeat sales methodology. Instead, Zillow uses the same underlying deed data as the Case-Shiller index, but creates a hedonically adjusted price index. The Zillow index uses

---

28 For more information on the construction of the Case-Shiller indices see the Standard and Poor’s website which documents their home price index construction methodology.

29 See http://furmancenter.org/

detailed information about the property, collected from public records, including the size of the house, the number of bedrooms, the number of bathrooms, etc. To the extent that the average characteristics of the home are changing over time, the Zillow index will capture such changes. It is reassuring to note that there is a high degree of correlation with respect to price changes between the three price indices. For example, restricting the sample to only the zip codes for which we have a Case-Shiller price index, the correlation in the growth in housing prices at the zip code level between 2000 and 2006 using the Case-Shiller data and then using the Zillow data was 0.94. Restricting our attention to Chicago, the correlation in housing price growth between 2000 and 2006 for the Case-Shiller data and Chicago Tribune data was 0.74.

In most of the analysis below, we will use Case-Shiller house price indices within cities, rather than within metropolitan areas. The reason for this is that we want to hold constant political economy factors (like the level of taxes and public goods) which could differ across cities within a metropolitan area. By focusing on within-city movements, we can hold such factors as the general level of taxes and public goods fixed. As mentioned above, we restrict our attention to those cities where we have data for at least 14 zip codes within the city during the time period we are studying. This makes the sample sizes change slightly across different analyses. Using the Case-Shiller data and restricting our analysis to within only the primary Case-Shiller city (within each metro area), we have data on 697 zip codes during the 1990-2000 period and we have data on 689 zip codes during the 2000-2006 period. In the online Robustness Appendix, we explore the behavior of zip codes within the entire MSA (not just within the primary city). In those cases, sample sizes are much higher.

Even though we have price indices at the zip code computed by Case Shiller and Zillow, we show the main results in the paper using only the Case Shiller data. We prefer the Case-Shiller data because of its reliance on repeat sales (something that Zillow does not do) and because the Case-Shiller data is explicit about which zip codes it thinks it does not have enough variation to reliably compute a price index. However, in the online appendix that accompanies this paper, we redo all the main analysis found in the subsequent sections using the Zillow data instead of the Case Shiller data. When we restrict the Zillow data to cover the same zip codes for which we have the Case Shiller data, the results using the Zillow data is nearly identical to the results using the Case Shiller data. This is not surprising given the high correlation between the two series, as reported above.

In the last section of the paper, where we attempt to explain differences in house price appreciation across cities, we use the Federal Housing Finance Agency (FHFA) metro level housing price indices instead of the Case-Shiller metro area indices. The Case-Shiller metro indices are limited to only 20 metro areas while the FHFA metro area indices cover over 200 cities. The FHFA index is also a repeat sales index and it includes properties of all different types (single family homes, condos, town homes, etc.), but it restricts the properties to only ones that are purchased with conventional mortgages.\footnote{Additional information about the construction of these housing price indices can be found at http://www.fhfa.gov/}. The reason we
switch to the FHFA index when exploring the cross city patterns is that the FHFA index covers a much
greater number of cities. The time series patterns for the FHFA index and the Case-Shiller index are
nearly identical at the city level for the cities where both indices are computed.

Finally, all data in the paper are reported in real 2000 prices, unless otherwise indicated. Likewise,
all growth rates are in real terms. We use the CPI-U (all items less shelter) to convert the nominal house
prices into real house prices.

5 Within City House Price Variation and Gentrification

In this section, we more systematically explore the predictions of our model in three steps. First, we show
descriptive features about the housing price dynamics across different neighborhoods within a city. Second,
we show that the neighborhoods within a city that appreciated the most during city wide housing price
booms are the poorer neighborhoods that directly abutted the richer neighborhoods. Third, we show that
these poor neighborhoods that appreciated the most during city wide housing price booms show signs of
gentrification.

5.1 Descriptive Patterns For Within City Price Movements

To start, we document the variability in housing price dynamics within a city during a given housing price
boom or bust. While much work has documented the variation in house prices across cities, little work has
been done systematically documenting the variation in house prices within a city. Table 2 shows the degree
of between- and within-city variation in house price appreciation in 2000-2006 (raw 1) and 1990-2000 (raw
2). When computing within city variation, we compute the variation in house price growth across zip
codes, using different measures of house price appreciation. For comparison, Columns 1 and 2 focus on
cross MSA variation in house prices. Column 1 uses data from the FHFA MSA level house price indices,
while Column 2 uses data from the Case-Shiller MSA level house price indices. As seen from Table 2,
there is a large variation in price appreciations across MSAs during both the 1990s and the 2000s. This is
consistent with the well documented facts discussed in Davis, Ortalo-Magné, and Rupert (2007), Glaeser,
Gyourko, and Nathanson (2008), Van Nieuwerburgh and Weill (2010), and Saiz (2010).

The last three columns of Table 2 show within city, cross-zip-code variation in house price appreciation
for the same time periods. For columns 3 and 4, we use data from the Case Shiller indices and show the
results for the main city within the MSA (column 3) and the MSA as a whole (column 4). In column 5, we
show the within city standard deviation in house prices using the Zillow data, which is only available for
2000-2006. The table shows that within city variation was roughly half as large as the cross city variation
during the 2000-2006 period and was of the same order of magnitude as the cross city variation during the
1990-2000 period. In both periods, the within city variation in house price appreciation across zip codes
was substantial. Understanding what drives the differences in appreciation rates within a city can help us
to explain the variation in house prices across cities.
According to our model, and as stated in Proposition 2, when there is a large increase in housing demand within a city, neighborhoods with a low price level should appreciate more on average than neighborhoods with a high price level. Moreover, Propositions 4 and 5 show that the bigger the city-wide demand shock, the larger the price appreciation of poor neighborhoods, on average, relative to richer neighborhoods. Also, according to Proposition 3, the variation in housing price appreciations should be higher among poorer neighborhoods relative to richer neighborhoods. The results in Figure 5 and Table 3 show support for these predictions during the 1980s, the 1990s and the 2000s.

Figure 5 plots the relationship between the house price level in year 2000 and the real growth in house prices between 2000 and 2006 for all Chicago community areas. In this figure, house prices are measured using the hedonic index we constructed with the Chicago Tribune data. Figure 5A shows that there is a sharp negative relationship between initial price level and the subsequent price appreciation for Chicago neighborhoods between 2000 and 2006, which is consistent with our model. Also, there is more variation in price appreciation for the community areas with initially low prices relative to those with initially high prices. In particular, among the neighborhoods in the top quartile of the housing price distribution in 2000, the standard deviation in house price growth between 2000 and 2006 was 14 percent. For the neighborhoods in the bottom quartile of the house price distribution in 2000, the standard deviation in house price growth between 2000 and 2006 was 26 percent. The differences in the standard deviation in house price growth between the rich and the poor neighborhoods was significant at the 1 percent level.

Table 3 shows similar results for a large selection of cities and metro areas during different time periods and using different measures of housing price appreciation. The key findings in this table are that the results shown for Chicago in Figure 5 hold in other cities during the 2000-2006 period, hold in other cities during the 1990s and 1980s and hold in reverse during periods of housing price busts. Specifically, Table 3 shows the mean growth rate in property prices over the indicated time period for neighborhoods in different quartiles of the initial house price distribution within the city or metro area. The last column shows the p-value of the difference in house price appreciation rates between the properties that were initially in the top (column 1) and bottom (column 4) quartiles of the housing price distribution within the city or metro area. In all cases, the initial level of housing prices used to define the quartiles in period \(t\) is defined using the average level of reported house price for the neighborhood from the corresponding U.S. Census (i.e., 2000, 1990, or 1980 depending on the time period studied).

We can conclude three things from the results in Table 3.\(^{32}\) First, the results for Chicago are robust no matter what price index we used. The results in Figure 5 persist even if we use the Zillow data or the Case Shiller data. No matter what index we use, between 2000 and 2006, initially low priced neighborhoods appreciated at much higher rates than initially high priced neighborhoods. Moreover, the results do not change when we look at only the zip codes within the city of Chicago (rows 1 - 3) or when we look at zip

\(^{32}\)In the robustness appendix, we show the underlying scatter plots (akin to Figure 3 for Chicago) for all the cities/MSAs discussed in Table 3.
codes within the Chicago MSA as a whole (row 4).

The second result from Table 3 is that the patterns in Chicago are found in all other cities. For example, within the Los Angeles, Washington DC, New York City, Boston and San Francisco MSAs, house prices in initially poor neighborhoods appreciated at substantially higher rates than initially rich neighborhoods during the 2000 - 2006 period. Although we have not shown the results in the table, but the results persist within Manhattan using the detailed community level price indices from the Furman Center. For example, neighborhoods on the upper west and east sides of Manhattan were both initially poorer and appreciated at much higher rates than neighborhoods in lower and midtown Manhattan. As noted in the introduction, properties in Harlem appreciated at over 130 percent while properties in midtown Manhattan appreciated at roughly 45 percent. Also, as seen from Table 3, these within city patterns are not limited to the recent period. During the 1990s, Denver and Portland experienced large housing price booms and within these cities, it was the poor neighborhoods that appreciated at much higher rates than the richer neighborhoods. Additionally, during the 1980s, New York and Boston experienced large housing price booms. Within these cities, it was the poor neighborhoods that appreciated at much higher rates than the richer neighborhoods during this time period. These results suggest that what we are finding is not an artifact of explanations put forth to explain the recent housing price boom.

The third result to be seen from Table 3 is that the opposite patterns occur during housing price busts. San Francisco and Boston experienced housing price busts during the 1990s. Within these cities, it was the poorer neighborhoods that contracted the most during the city wide busts. These results suggest that the poorer neighborhoods are more price elastic during both housing price booms and busts.

Like the Chicago results discussed above, the variation in house price changes for these cities is much higher among the initially lower priced neighborhoods compared to the initially higher priced neighborhoods. For example, within the New York metro area during the 2000 - 2006 period, the standard deviation of house price appreciations within the neighborhoods that were in the bottom quartile of the initial house price distribution was 29 percent (using the Case Shiller index). For comparison, the standard deviation of house prices within the neighborhoods that were in the top quartile of the initial house price distribution was only 5 percent. The differences were significant at the less than 1 percent level. Similar patterns hold in most other cities. It is this variation among low priced neighborhoods that we will exploit to directly test the mechanism at the heart of our model later in this section.

One additional relationship that we show in the online Robustness Appendix is that the gap in the appreciation rates between initially high priced and initially low priced neighborhoods is larger when the city wide property price changes were higher. When the city as a whole had a large property price boom, initially low housing price neighborhoods appreciated at much greater rates than initially high price neighborhoods. However, when the city as a whole had a smaller property price boom, there was little differences in appreciation rates between the initially high priced and initially low priced neighborhoods. For example, within Charlotte - which had only a city wide house price appreciation of less than 10
percent between 2000 and 2000 - property prices in both the top and bottom third of the initial house price distribution appreciated at the same rate. This is consistent with the implications of the model outlined above. When there is little or no city wide demand shock for housing, there is little systematic wide scale gentrification that occurs.

5.2 Housing Price Dynamics and Proximity to Rich Neighborhoods

One of the main predictions of our model is that the variance in house price changes among poorer neighborhoods during housing price booms can be predicted by their proximity to the initially richer neighborhoods. Proposition 3 shows that within the set of neighborhoods with initially low prices, the ones that appreciate the most will be the ones that are closest to the initially high priced neighborhoods. We now use a regression framework to systematically explore the relationship between housing price growth among poorer neighborhoods and their proximity to rich neighborhoods during city wide housing price booms. Also, we want to see whether our proximity results hold up even after controlling for initial differences between the neighborhoods and for the other potential mechanisms, discussed above, that can cause different neighborhoods to appreciate at different rates during a city wide property price boom.

Specifically, pooling all of our data together, we estimate:

$$\Delta P_{i,j}^{t+k}/P_{i,j}^t = \alpha + \beta \ln(Dist_{i,j}^t) + \Gamma X_{i,j}^t + \Psi Z_{i,j}^t + \mu_j + \epsilon_{i,j}^{t+k}$$

where $\Delta P_{i,j}^{t+k}/P_{i,j}^t$ is the growth in housing prices, as measured by the Case-Shiller index, between period $t$ and $t+k$ within neighborhood $i$ in city $j$. The variable of interest in this regression is $\beta$, the coefficient on $\ln(Dist_{i,j}^t)$ which measures the log of the distance (in miles) to the nearest zip code in city that resides in the top quartile of zip codes with respect to housing prices in period $t$. In the online Robustness Appendix, we show the results from the same regression using the Zillow data.

The vector $X_{i,j}^t$ includes a series of variables designed to control for initial differences across the zip codes. These controls include the log of average household income of residents in neighborhood $i$ in period $t$, the log of the average initial house price of residents in neighborhood $i$ in period $t$, and the fraction of the residents in neighborhood $i$ in period $t$ that are black. We also include a series of dummies to control for the age of the housing stock within the zip code in period $t$. All of the $X$ controls came from either the 1990 or 2000 U.S. Census, depending on the specification. By including these controls, we are asking whether two otherwise similar neighborhoods have different house price appreciations during city wide housing booms based on the proximity to the high priced neighborhoods. We also control for the fact that as shown by Rosenthal (2008), the age of the housing stock could be an important determinant of predicting which neighborhoods will subsequently gentrify. To control for the age of the housing stock, we include 4 dummy variables. For the 2000 - 2006 regressions these are: the fraction of the housing stock built less than 5 years ago, the fraction of the housing stock built between 5 and 19 years ago, the fraction built between 20 and 39 years ago, and the fraction built between 40 and 59 years ago. The omitted group is the fraction of the housing stock built more than 60 years ago. For the 1990 - 2000 regressions the age of structure variables are the same except that the last dummy is the fraction built between 40 and 49 years ago and the omitted group is the fraction of the housing stock built more than 50 years ago. The slight difference is due to differences between the 1990 and 2000 Census.
different sized housing price booms may affect different types of neighborhoods differently. To do this, we interacted the initial income and the initial house price within the zip code with the size of the city wide housing boom. These interaction terms have little effect on our estimates. Given that we also include city fixed effects, $\mu_j$, all of our identification comes from within city variation.

Lastly, we also include a vector $Z$ which is designed to control for the other mechanisms which can explain why house prices move differentially within a city. Specifically, we control for the average commuting time within neighborhood $i$ during period $t$. This is designed to measure how long it takes for residents to commute to the jobs within the city. Holding distance to the jobs constant, we are asking whether there are difference in property price appreciations across neighborhoods based upon the proximity to the high priced neighborhoods. Also, we control for the distance to fixed natural amenities like lakes, rivers, and oceans that either are part of the city or border the city. If there are no lakes, rivers, or oceans that border or are part of the city, the distance variables are set to zero.$^{34}$

Before proceeding, let us mention that the identifying assumption behind our empirical analysis is that after controlling for commuting times, distance to lakes, rivers, and oceans, the race of neighborhood residents, and the age composition of the housing stock, there is nothing else that affects the desire of households to live in a given neighborhood aside from the amenities that are generated from having richer neighbors. This, however, does beg the question as to why richer residents, conditional on our set of controls, choose to congregate within one neighborhood within a city as opposed to another. One potential explanation would be that the rich congregated in an area originally because of an amenity that used to exist but that amenity has sense faded away. This would be the case for a city like Chicago which had the stock yards on the south side which could have caused richer residents to located to the north side. For the universe of cities in our data, it is hard to do a systematic analysis of all potential within city differences in fixed amenities. However, we are reassured by the fact that adding controls for the fixed amenities that we do observe does little to change our estimates. If the observables for which we do control are not having a significant effect on our results, it is less likely that any unobservables for which we do not control would have a significant effect on our results.

Table 4 shows the results of the above regressions on two samples. The first is for house price changes between 2000 and 2006 (columns 1 and 2) and the second is for house prices changes between 1990 and 2000 (columns 3, 4, and 5). In columns 1 and 3, we estimate the above equation without the vector of $Z$ controls. In columns 2, 4 and 5, we include the $Z$ controls. In both samples, we restrict our analysis to only those zip codes within the bottom half of the city’s house price distribution. In essence, we are asking whether it is the low price neighborhoods that are close to the high price neighborhoods that appreciate the most during housing price booms, all else equal. During the 2000-2006 period, all cities in our sample experienced a housing price boom. During the 1990-2000 sample, some of the cities did not experience

$^{34}$For our 1990-2000 sample, we also control for the change in commuting times within the neighborhood between 1990-2000. We cannot control for the change in commuting time for the 2000-2006 because we are not able to observe commuting times at the zip code level in 20006.
a housing price boom. Moreover, in some cities, prices actually contracted slightly. In column 5, we re-estimate the specification in column 4 for the 1990-2000 period interacting the distance to high priced neighborhoods with whether the city had a positive house price appreciation or whether the city had a zero or negative house price appreciation.

Table 4 shows that during both the 2000 to 2006 period and the 1990 to 2000 period, proximity to high price neighborhoods is a strong predictor of a housing price appreciation among otherwise similar low priced zip codes during periods of city wide housing booms. The estimates are not very sensitive to including controls for distance to jobs or distance to natural amenities like oceans, lakes, or rivers. Again, this is encouraging in that if the measures of fixed amenities for which we are able to control for do not affect the coefficient, it may be the case that the unobserved differences in fixed amenities that we are unable to control for would not affect the coefficient.

The size of the estimates on distance to high price neighborhoods in Table 4 are economically meaningful. For example, for the specification in columns 2 and 5, a doubling of distance to high price neighborhoods (from 1 to 2 miles, from 2 to 4 miles, etc.) decreases the property price appreciation in the zip code by 5 percent. In other words, zip codes that are within 1 mile of the high priced neighborhoods increased by 10 percent more than neighborhoods that were 4 miles away from the high priced neighborhoods. The average distance to high priced neighborhoods within our sample was 3.9 miles. It should be noted, that there is not much of an effect of distance - in either direction - on low priced neighborhoods during periods when city wide housing prices were constant or falling. This is not surprising given that even when housing prices were falling in these cities, it was not by a large amount between the 1990 to 2000 period. Our model predicts that when housing demand is roughly constant, there should not be a relationship between distance to rich neighborhood and housing price appreciations.

### 5.3 House Price Changes and Gentrification

The results in Table 4 show that during housing price booms it is the poor neighborhoods that are in close proximity to rich neighborhoods that appreciate the most. In this sub-section, we examine more deeply the mechanism of our model. Our model predicts that the poor neighborhoods next to the rich neighborhoods are the ones that appreciate the most because they are the ones where rich households move after a housing demand shock. This implies that neighborhoods that experience higher house price appreciation should also show signs of gentrification.

To analyze whether neighborhoods that experienced a rapid growth in prices also experienced signs of gentrification, we estimate the following regression:

\[
\frac{\Delta P_{i,j}^{t+k}}{P_{i,j}^{t}} = \alpha + \beta Y_{i,j}^{t+k} + \Gamma X_{i,j}^{t} + \mu_j + t_{i,j}^{t+k}
\]

where \(Y_{i,j}^{t+k}\) is some measure of gentrification in neighborhood \(i\) of city/metro area \(j\) during the period \(t\) to \(t+k\) and where \(\Delta P_{i,j}^{t+k}/P_{i,j}^{t}, X_{i,j}^{t}, \) and \(\mu_j\) are defined as above. The regression asks, conditional on
controlling for the initial $X$ vector, whether or not zip codes that showed signs of gentrification experience higher house price growth than other zip codes. For this regression, we use the time period 1990 to 2000. We do this so that we can get our measures of gentrification using the 1990 and 2000 U.S. Censuses. We use three measures of $Y$ (in three different regressions). First, we explore the percentage change in income within the zip code between 1990 and 2000. Second, we use the percentage point change in the poverty rate within the zip code between 1990 and 2000. Lastly, we measure the percent change in the average tenure of residents within the zip code. We are interested in assessing whether neighborhoods that got richer, had declines in poverty rates, or had new people move into them experienced higher house price growth, all else equal.

The results are shown in Table 5. Each column of Table 5 represents a different regression with a different measure of gentrification, $Y$, on the right side. As seen from the table, zip codes that experienced higher income growth, a larger decline in poverty rates, and had a bigger decline in the housing tenure of residents all experienced larger property price increases within the zip code. The magnitudes are also economically large. A twenty percent increase in income within the zip code is predicted to result in a 6.4 percent increase in housing prices within the zip code.\footnote{Given that the mean increase in house price growth across the zip codes during this period was 26 percent, this relationship between income growth and house price growth is quantitatively large. Similar patterns are seen with the change in the poverty rate and the percentage change in the average tenure of the residents. Consistent with our theory, neighborhoods that showed signs of gentrification within a city, all else equal, experienced larger property price increases within the city.}

Let us mention that our model has also stark implications about the variation in rental prices within a city. Getting data on rental prices is much more difficult than getting data on housing prices. Morris Davis, when discussing our paper at an Winter 2010 NBER EFG conference, purchased rental data for the Chicago, San Francisco and Charlotte metro areas. He replicated the main empirical findings in our paper using the rental price data. A discussion of these results can be found in Davis (2010). For our systematic analysis, we prefer using the housing data because of our ability to get disaggregated price indices at finer geographic levels within the city. As Davis shows, such fine geographic disaggregation is not possible with the rental data. The data exists only for large groupings of zip codes. However, it is comforting that our empirical results go through when using rental data.

6 Instrumenting For Neighborhood Income Shocks

In the previous section we have established empirical relationships that are consistent with our theory of endogenous gentrification. However, we could not make any claim about causation, given that we were
silent on the shocks behind the housing demand change. In this section, we try to measure directly whether an exogenous shock to income in one neighborhood affects property prices and measures of gentrification in neighboring areas. As before, we also rule out that these effects are being driven by changes in commuting times or proximity to fixed natural amenities.

6.1 Industry Wage Shocks

To measure exogenous shocks to income for each neighborhood, we are going to use variation in national earnings by industry between 1990 and 2000. Using that, we will predict the expected change in income for each neighborhood in our sample based upon the industry composition of residents of each zip code in 1990. Our identifying assumption is that the change to earnings for each industry at the national level is orthogonal to anything else that would drive house prices in the local neighborhoods included in our analysis. This approach of imputing exogenous income shocks for local economies has been used extensively by others in the literature (see, for example, Bartik (1989) and Blanchard and Katz (1992)).

Specifically, we estimate the following regression:

\[
\frac{\Delta P_{i,t+k}}{P_{i,t}} = \delta_0 + \delta_1 \text{OwnIncShock}_{i,t+k}^{i,j} + \delta_4 \text{NeighborIncShock}_{i,t+k}^{i,j} + \Gamma X_{i,j}^{i,j} + \Psi Z_{i,j}^{i,j} + \mu_j + \epsilon_{i,j,t+k}
\]

where \(\Delta P_{i,t+k}/P_{i,t}^{i,j}\), \(X_{i,t}^{i,j}\), \(Z_{i,t}^{i,j}\), and \(\epsilon_{i,j,t+k}\) are defined as above. The variable \(\text{OwnIncShock}_{i,t+k}^{i,j}\) denotes the predicted income growth for zip code \(i\) in city \(j\) between \(t\) (1990) and \(t+k\) (2000) based on the industry mix of residents in zip code \(i\) in 1990, while the variable \(\text{NeighborIncShock}_{i,t+k}^{i,j}\) is the predicted income growth of the neighboring zip codes between 1990 and 2000 based on the industry mix of their neighbors. We define a zip code’s neighbors based on proximity to the zip code. We discuss this process in more detail below. We will estimate these regressions on the full sample of zip codes within the primary cities (excluding suburbs) from the Case Shiller data.

Before proceeding, an additional discussion of how we computed own and neighbor income shocks is needed. Using the one percent samples from the 1990 and 2000 IPUMS data, we defined income growth for each two-digit industry between 1990 and 2000. Our measure of income was individual earnings. The only restrictions we placed on the data was that the individual had to be employed and over the age of 16. These two restrictions were needed given that the industry breakdown for the Census zip code aggregates are for employed individuals over the age of 16. There is a large variation in income growths between 1990 and 2000 across the industries. For example, the “Personal Services” industry had a real appreciation of annual earnings of 35.4% (followed by “Business and Repair Services” and “Finance, Insurance and Real Estate” at 30.7% and 27.2%, respectively) while the “Transportation” industry had a real appreciation of annual earnings of only 6.5%. The average industry had a real appreciation in earnings of 17.3% between 1990 and 2000 with a standard deviation of 7.9%.

Using the growth rate in industry earnings over the 1990s, we can compute the predicted increase in
earnings for each zip code between 1990 and 2000 based on their industry mix in 1990. Specifically, we
multiply the industry growth rate in earnings by the fraction of people in each neighborhood working in
those industries. Given the Case-Shiller indices are at the zip code level, we conduct our analysis at the
zip code level. Most zip codes contain individuals working in most industries. However, despite that, there
is still variation in predicted incomes across neighborhoods. Within the zip codes we analyze (the 572 zip
codes in major cities covered by Case-Shiller), the median predicted income change is 18.0 percent with a
standard deviation of 1.0 percent. The 5th percentile of the predicted income change across zip codes is
16.0 percent while the 95th percentile is 19.3 percent. While the variation is not large when we aggregate
to the level of zip code, some variation does exist.

For the zip codes in our data, predicted income changes based on the industry mix does predict actual
income changes. Regressing actual income change in the zip code between 1990 and 2000 on the predicted
income change based on initial industry composition and city fixed effects yields a coefficient on actual
income changes of 4.37 with a standard error of 0.90 and an adjusted R-squared of 0.43. The incremental
adjusted R-squared for the predicted income change above and beyond the city fixed effects is 0.04. Again,
while there are lots of differences in income changes across zip codes, the predicted income change measure
based on industry mix does have some predictive power. The first stage F-stat of the predicted income
measure is 18.0.

For the change in income of one’s neighbors, the key is defining which are the neighboring zip codes.
We do this in a few ways. Throughout all of our analysis, we define neighbors as being those other zip
codes that are spatially close to a given zip code. We measure “spatially close” in terms of distance to the
mid point of the zip code. In the results below, we define tiers of neighbors: the 10 closest zip codes, the
next 10 closest zip codes, and the 10 closest zip codes after that. These tiers comprise the universe of zip
codes within a city for all cities in our data.

The results of estimating the above equation are shown in Table 6. In the first column, we estimate
the equation only including the initial house price in the neighborhood, the initial fraction of African-
American residents in the neighborhood, city fixed effects, and the own predicted income shock of the
neighborhood. The regression indicates that an exogenous shock to the earnings of the residents in one’s
zip code is associated with an increase in house prices in the neighborhood. For example, going from the
shock experienced by the 5th percentile of the zip codes to the 95th percentile of the zip codes (a 0.033
increase) will result in an increase in house prices of 32.2 percent (0.033 * 9.76).

In the second column, we include both the own income shock and the average income shock experienced
by one’s 10 closest neighbors. The results show that a shock to one’s neighbors income increases property
prices in that neighborhood - even controlling for the predicted income shock in that neighborhood. This
effect is also large. Moving from the 5th percentile of the average neighbor’s income shock (17.1 percent)
to the 95th percentile of the average neighbor’s income shock (19.2 percent) increases house prices in
one’s own neighborhood by 17.6 percent (0.021 * 8.39). A shock to income of neighboring zip codes has
a large effect on the house prices in one’s own zip code. This is very much consistent with the results of our model. Notice, in column 3, including the income shocks of those zip codes that are 11-20 and 21-30 zip codes away have no effect on house prices in the neighborhood. It is only the shock to income for those zip codes that are spatially close to you that affects your property prices. In the fourth column, we show the results are relatively similar if we include variables that control for commuting times, change in commuting times, and proximity to the natural amenities. Again, it is not the traditional urban stories of proximity to jobs or proximity to fixed natural amenities that are driving the results. Even with these controls, moving from the 5th percentile of the average neighbor’s income shock to the 95th percentile of the average neighbor’s income shock increases house prices in one’s neighborhood by 17.0 percent (0.021 * 8.08). All of these estimates are statistically significant at standard levels.

6.2 Industry Wage Shocks and Gentrification

Do house prices increase in the neighboring areas because those neighboring areas gentrified? In our model, gentrification is the mechanism which leads house prices in neighboring areas to increase. To explore the effects of our instrument on neighborhood gentrification, we estimate:

\[ Y_{i,j}^{t,t+k} = \delta_0 + \delta_1 \text{OwnIncShock}_{i,j}^{t,t+k} + \delta_4 \text{NeighborIncShock}_{i,j}^{t,t+k} + \Gamma X_{i,j}^{t} + \Psi Z_{i,j}^{t} + \mu_j + \epsilon_{i,j}^{t,t+k} \]

where all the variables are defined as above. The results of this regression are shown in Table 7. Again our three measures of gentrification are percent change in median income in the neighborhood (column 1), percentage point decline in the poverty rate in the neighborhood (column 2) and percentage change in the median tenure of residents in the neighborhood (column 3). A predicted shock to income of residents in the neighborhood leads to increased income and decreased tenure in the zip code, but has no statistically significant effect on the poverty rate. However, the results are much stronger when one’s neighbors receive a positive income shock. Receiving a positive income shock in one neighbor causes the neighboring zip codes to experience a large statistically significant increase in median income, a statistically significant decline in the poverty rate, and a statistically significant decline the tenure of residents. On net, an income shock in one area causes richer households to move into neighboring neighborhoods. Their moving in is also associated with housing prices increasing dramatically. This is exactly the mechanism highlighted in the paper.

7 Cross-City Variation in Price Appreciation

In the previous sections, we have shown that the within city patterns in house price movements are consistent with our model. In this section, we explore whether it is consistent with cross city differences in house price changes.
Recently, important work has been done by Saiz (2010) and Gyourko, Mayer, and Sinai (2006) exploring the importance of cross-city differences in housing price appreciations. Both papers emphasize differences in supply constraints across cities to explain differences in house price appreciations across cities. Saiz focuses on supply restrictions due to geographical or regulatory constraints. Cities differ in the extent to which land is easily developable for new construction and the extent to which regulation hinders to the ability to adjust the housing stock. For a given demand shock, cities where housing supply is more inelastic will experience larger house price changes, all else equal. Gyourko et al. highlight the importance of what they term “super star cities”. These cities are endowed with a fixed desirable amenity (like good weather). Because proximity to the amenity is in fixed supply, households will bid up the land prices around the amenity as households become richer. Again, such a model predicts that there should be big differences in the relationship between changes in household income and changes in house prices between cities where housing supply can adjust easily and where housing supply is more inelastic.

Our model differs from these other models in that changes in income and housing prices are predicted to be strongly correlated even in cities where housing supply is relatively elastic. The reason for this is that the influx of richer households into a city will endogenously increase the desirability of living in the neighborhoods where these households will move. As seen from the model section above, the gentrification of such neighborhoods will increase the average level of housing prices in the city as a whole.

Figure 6 shows the bivariate relationship between income growth and house price growth across MSAs between 1990 and 2000. The MSAs included in the figure had to meet two requirements. First, we only included MSAs where we had MSA level price indices from FHFA. Second, we only included MSAs for which Saiz computed a measure of the housing supply elasticity within the city. More elastic cities are more easily able to adjust supply and, as a result, should have prices that respond less to housing demand shocks. This restrictions left us with 152 MSAs.

In Panel A of Figure 6, we examine the bivariate relationship between house price growth and income growth for all MSAs in our sample. Our measure of income growth comes from using mean income for the MSA from the 1990 and 2000 U.S. Censuses. In Panel B of Figure 6, we show the same relationship but for those cities for which supply can adjust easily according to the Saiz measure. In particular, we only graph the relationship for those cities in the bottom half of the Saiz elasticity measure (out of all cities included in Panel A). As seen from comparing the two figures, there is an equally strong relationship between house prices and income growth within the cities with low supply constraints and for the sample of all cities.

To formally examine the extent to which changes in income are correlated with changes in house prices

---

36See Saiz (2010).
even among cities with relatively elastic supply, we estimate:

$$\frac{\Delta P_j^{t,t+k}}{P_j^t} = \delta_0 + \delta_1 \frac{\Delta Inc_j^{t,t+k}}{Inc_j^t} + \delta_2 SupplyConstraints_j^t + \delta_3 \frac{\Delta Inc_j^{t,t+k}}{Inc_j^t} \times SupplyConstraints_j^t$$

$$+ \delta_4 \frac{\Delta Commute_j^{t,t+k}}{Commute_j^t} + \epsilon_j^{t,t+k}$$

where $\Delta P_j^{t,t+k}/P_j^t$, $\Delta Inc_j^{t,t+k}/Inc_j^t$, and $\Delta Commute_j^{t,t+k}/Commute_j^t$ is the percentage growth in housing prices, income, and commuting time, respectively, for city $j$ between $t$ and $t+k$. $SupplyConstraints_j^t$ is a measure of the extent to which supply can adjust in the metro area. As discussed above, we use the measure of housing supply elasticity constructed in Saiz (2010) as our measure of supply constraints. We restrict our analysis to include only the 155 MSAs for which the measure of housing supply elasticity is available.

As seen from column 1 of Table 8, changes in income alone explain 49 percent of the variation in house prices growth across MSAs between 1990 and 2000s. Our change in income measure comes from the 1990 and 2000 U.S. Censuses. All of our observations are at the MSA level. This is consistent with the results seen in Figure 6. In column 2 of Table 8, we include only the supply elasticity measures. Higher housing supply elasticity should predict lower changes in house prices. During the 1990s, the Saiz supply elasticity measure alone explains none of the variation across MSAs. In column 3, we include both the change in income and the supply elasticity measure. The incremental adjusted R-squared does not change as we add the supply elasticity measure to the growth in income measure. In essence, changes in income explain a lot of the variation in house prices across MSAs in the 1990s.

The interesting result come in columns 4 and 5 where we include the interactions of the income growth with the measures of the supply elasticity. In column 4, we simply include the change in income measure with the continuous elasticity measure. In column 5, to ease the interpretation of the results, we replace the continuous elasticity measure with a dummy variable for whether the MSA was in the middle third of the housing supply elasticity measure and a dummy variable for whether the MSA was in the bottom third of the housing supply elasticity measure. The MSAs in the bottom third of the elasticity measure have the most inelastic housing supply, and as a result, should have the highest change in prices. While it is true, as theory predicts, is that MSAs with more inelastic housing supply have housing prices that respond more to income shocks than MSA with inelastic supply, the interesting result from Table 8 is that even within MSAs that have relatively elastic housing supply, house prices respond strongly to changes in MSA wide income. These results is exactly predicted by our model. Even in places where supply is relative elastic, populating the MSA with richer residents will endogenously increase the amenities within the MSA, making the MSA more desirable. According to the regression, even after controlling for differences in supply elasticities, the elasticity of housing price changes to changes in income is 0.80. This should be viewed as the elasticity of housing price changes to income changes for those cities where housing supply can easily adjust.
One other thing is of note in columns 4 and 5 of Table 8. In those regressions, we also include the growth rate in MSA wide commuting times, which add little to the explanatory power of the regression. Moreover, the coefficient on the income variable and the income-supply constraint interactions were unchanged by the inclusion of the change in commuting time controls. In separate regressions we included a cubic in the growth rate of MSA median commuting times and interactions of the growth rate in commuting times and the growth rate in income and the growth rate in commuting times and the supply elasticity. These regression results were quantitatively quite similar to those shown.

In columns 6-10 of Table 8, we show similar results for the 2000-2006 period. For this period, we use the change in MSA level income as computed from IRS zip code level information. During the recent period, differences in income growth across neighborhoods only explained 30 percent of the variation in house prices across cities. The Saiz elasticity measure alone, however, explained 33 percent of the variation in house price growth across MSAs. This seems to suggest that there a large demand shock for housing that was orthogonal to changes in income that also drove differences in housing prices during this period. This is consistent with many recent papers that suggested that low interest rates and the extension of credit to low income borrowers explained housing price dynamics in the recent period (see Mian and Sufi (2009)). But, as seen from columns 6-10 of Table 8 and from the within city results above, our mechanism was also active during the current cycle. In particular, even among MSAs were housing supply is fairly elastic, MSAs with high income growth during the 2000s also had sizeable increases in housing prices. The incremental R-squared of the income measures during the 2000s was much smaller than during the 1990s. This suggest that while our mechanism was active in the recent period, it was less important in explaining cross city variation in housing prices than it was during the 1990s. But, even among MSAs with a high housing supply elasticity, the elasticity of housing prices with respect to income changes was estimated to be 0.96.

8 Conclusions

In this paper, we explore the theoretical and empirical importance of neighborhood consumption externalities in explaining within and across city house price dynamics. The key assumption in the model is that all individuals prefer neighborhoods populated by richer households as opposed to neighborhoods populated by poorer households. The reason for this is that richer neighborhoods provide amenities that are desirable to individuals. While we do not take a stand on the exact amenities, we have in mind that richer neighborhoods have lower levels of crime, higher provisions of local public goods, better peer effects, and a more extensive provision of service industries (like restaurants and entertainment options).

Including this externality into a simple geographical model of a city generates a host of interesting

---

37 In a separate regression (not shown), we regressed the change in commuting time on the change in income and the change in income interacted with the supply elasticity measures. The results show that changes in MSA wide income does not explain much of the variation in changes in commuting times across cities during this period. As MSAs got richer, average commuting times did not change, suggesting that the reason that housing prices were going up in the elastic cities is not because the city was expanding outward which would have increased average commuting times.
theoretical predictions. We find that different neighborhoods within a city respond differentially to a housing demand shock. In particular, we find that poorer neighborhoods at the border between the poor and rich neighborhoods are the most price elastic to positive city wide housing demand shocks. Additionally, anything that increases the demand for housing by richer households will have persistent effects on housing prices within the city, even if the city is unbounded.

We then conduct a variety of empirical tests that provide support for the mechanism in the model. In particular, we test our mechanism against the importance of production externalities coupled with transportation costs and against a fixed amenity stories. Even after controlling for average commuting times and distance to fixed natural amenities, poor neighborhoods that are close to richer neighborhoods have much higher housing price appreciation during periods of positive housing demand shocks. In addition, these neighborhoods show marked signs of gentrification in response to the housing demand shocks. We also show that our mechanism can also explain some of the variation in cross-city housing price dynamics, which cannot be fully explained by differences in house supply elasticity.

Before concluding, we wish to emphasize a few additional points. First, we are not arguing that alternative mechanisms to explain house price movements are not important. Transportation costs, fixed natural amenities, and supply constraints are all important in explaining house price movements within and across cities. What we are saying, however, is that there is an additional mechanism – local consumption externalities – that is also quantitatively important in explaining housing price dynamics at different levels of aggregation. Prior to our work, this mechanism has not really been studied as an explanation for housing price dynamics.

Second, we feel it is economically important to distinguish between the different mechanisms. Although we have not highlighted the welfare implications of our model in this paper, we believe that they can differ substantially from the welfare implications of the alternative mechanisms mentioned above. We believe an interesting avenue for future research would be to explore the welfare implications of local productivity demand shocks (as in Moretti (2009) and Notowidigdo (2010)) in a world where amenities are endogenously provided.

Additionally, our paper focuses on one dimension of household preferences – the income of one’s neighbors. It is likely that households may have preferences over different features of their neighbors. For example, as discussed in Section 2, the existing literature has focused on individuals having preferences over the race of one’s neighbors. The work we show above, particularly in Section 6, shows that shocks to income (based on industry mix) do affect land prices within a city irrespective of the shocks to the race of one’s neighbors. However, we acknowledge that in most other contexts, it is hard to separate the effect of racial preferences from the effect of income preferences given that race and income are so highly correlated. This begs the question as to how much of the previous literature which documents a preference for the race of one’s neighbor is confounded with documenting a preference for the income of one’s neighbors. We also think that it would be fruitful, in future work, to try to tease out separately the importance of racial
preferences for one’s neighbors from the income preferences for one’s neighbors.

Lastly, throughout the paper, we have been silent on the political economy issues associated with gentrification. Local officials are often hesitant to allow large scale developers to buy up real estate in large sections of communities with the purpose of gentrification. Such political resistance can slow down the gentrification process in some neighborhoods. These factors may explain why it takes longer for some neighborhoods within a city to gentrify relative to others. Again, we leave a formal analysis of such political economy issues to future research.
References


Moretti, E. 2009. “Real Wage Inequality.”  


A: Hedonic Home Price Index.

B: Median Income.

Figure 1: Home Prices and Income by Zip Code in Chicago in 2000.

Figure 2: Externality across space.
Figure 3: Reaction of house prices across space to a positive demand shock. We set $\alpha = .8$, $\beta = .8$, $\delta_R = 0.2$, $\delta_P = 0$, $A = 1$, $\phi = 1$, $\gamma = .1$, $N^R = N^P = .5$, $C^P = C^R = 25$, and $r = .03$.

Figure 4: Price growth rate across locations as a function of the initial price level and OLS regression. We set $\alpha = .8$, $\beta = .8$, $\delta_R = 0.2$, $\delta_P = 0$, $A = 1$, $\phi = 1$, $\gamma = .1$, $N^R = N^P = .5$, $C^P = C^R = 25$, and $r = .03$.

Figure 5: Price Growth vs. Initial Price 2000-2006 for Chicago City Community Areas using Hedonic Index.

R2 = 0.30, alpha = 3.56, $\beta = -0.25$ (0.05) N = 76
Table 1: The Relationship Between Home Price and Income in Chicago

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log Mean Income</td>
<td>1.15</td>
<td>0.68</td>
<td>0.62</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.12)</td>
<td>(0.15)</td>
<td>(0.13)</td>
<td></td>
</tr>
<tr>
<td>Log Mean Commuting Time</td>
<td>-3.22</td>
<td>-2.05</td>
<td>-1.26</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.29)</td>
<td>(0.36)</td>
<td>(0.42)</td>
<td></td>
</tr>
<tr>
<td>Log Distance to Loop</td>
<td>0.14</td>
<td>0.00</td>
<td>-0.06</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.09)</td>
<td>(0.07)</td>
<td>(0.07)</td>
<td></td>
</tr>
<tr>
<td>Log Distance to Lake</td>
<td>-0.04</td>
<td>-0.03</td>
<td>-0.08</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td></td>
</tr>
<tr>
<td>Fraction African American</td>
<td>-0.36</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.13)</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.59</td>
<td>0.48</td>
<td>0.68</td>
<td>0.71</td>
</tr>
</tbody>
</table>

Notes: This table shows the results of an OLS regression of log housing prices across Chicago neighborhoods in year 2000 on a series of neighborhood controls. For our measure of housing prices, we use the hedonically adjusted house price index created from the Chicago Tribune data. We discuss this procedure in greater detail within the text. Our measure of neighborhood is a Chicago community area. There are 77 community areas within Chicago. In column 1, we regress the log of the 2000 hedonically adjusted house price index on 1) log mean commuting time to work as reported by residents in the community area, 2) log distance from the community area’s center (in miles) to the “Chicago Loop” which is Chicago’s center business district, and 3) log distance from the community area’s center (in miles) to Lake Michigan. In column 2, we only include the log of mean household income of residents in the community area as a regressor. In column 3, we combine the regressors from columns 1 and 2. In column 4, we also add a control for the fraction of the residents in the community area who are African American. The commuting time, income and race data by community area comes from the 2000 U.S. Census. See the text for details.
Table 2: Housing Price Growth Variation

<table>
<thead>
<tr>
<th></th>
<th>FHFA b/w</th>
<th>C-S b/w</th>
<th>C-S MSA w/i</th>
<th>C-S city w/i</th>
<th>Zillow city w/i</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000-2006</td>
<td>0.36</td>
<td>0.53</td>
<td>0.17</td>
<td>0.18</td>
<td>0.25</td>
</tr>
<tr>
<td>1990-2000</td>
<td>0.16</td>
<td>0.28</td>
<td>0.15</td>
<td>0.16</td>
<td></td>
</tr>
</tbody>
</table>

Notes: This table shows the cross MSA standard deviation in house price growth for both the 2000-2006 period and the 1990-2000 period using the FHFA housing price index (column 1) and using the Case Shiller housing price index (column 2). Both the FHFA index and the Case Shiller index are discussed in the text. When computing the standard deviation in house price growth using the FHFA data we include 152 MSAs. The MSAs included are all MSAs with a population above 200,000 for which FHFA indices exist and where there is also a Saiz estimate of an MSA level housing supply elasticity. This latter restriction makes the set of MSAs included in the FHFA analysis the same as the set of MSAs analyzed in Section 7 of the paper. When computing the standard deviation in house price growth using the Case Shiller indices, we restrict the MSAs to 26 MSAs where we have data for the 1990-2000 period and the 28 MSAs where we have data fro the 2000-2006 period. In columns 3-5, we measure the between zip code within city/MSA standard deviation in house price growth. In columns 3 and 4, we restrict our analysis to those zip codes within the cities/MSAs for which we have Case Shiller data. Again, the measure of house price appreciation comes from using the zip code level Case Shiller price indices. As discussed in the paper, we only restrict our analysis to those zip codes for which a Case Shiller zip code is computed using actual (non-imputed) data. Column 3 shows the results from analyzing the cross zip code variation within the MSA as whole. Column 4 shows the results from analyzing the cross zip code variation only within the primary city within the MSA. In column 5, we show the within city variation using house price appreciations computed using the Zillow price index. For comparability, we restrict the cities in column 5 to be the same as the cities in column 4. As discussed in the text, however, we have a Zillow index for all zip codes within the city.

Table 3: Housing Price Growth by Initial Price Quartile

<table>
<thead>
<tr>
<th></th>
<th>Quartile 4</th>
<th>Quartile 3</th>
<th>Quartile 2</th>
<th>Quartile 1</th>
<th>p-val of Quartile 4 = Quartile 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000-2006: Housing Booms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chicago, City Level, Chicago Tribune</td>
<td>0.37</td>
<td>0.54</td>
<td>0.71</td>
<td>0.68</td>
<td>0.00</td>
</tr>
<tr>
<td>Chicago, City Level, Zillow</td>
<td>0.36</td>
<td>0.52</td>
<td>0.57</td>
<td>0.63</td>
<td>0.00</td>
</tr>
<tr>
<td>Chicago, City Level, Case Shiller</td>
<td>0.53</td>
<td>0.66</td>
<td>0.72</td>
<td>0.88</td>
<td>0.00</td>
</tr>
<tr>
<td>Chicago, MSA Level, Case Shiller</td>
<td>0.47</td>
<td>0.50</td>
<td>0.49</td>
<td>0.66</td>
<td>0.00</td>
</tr>
<tr>
<td>New York City, MSA Level, Case Shiller</td>
<td>0.64</td>
<td>0.75</td>
<td>0.86</td>
<td>1.09</td>
<td>0.00</td>
</tr>
<tr>
<td>Boston, MSA Level, Case Shiller</td>
<td>0.43</td>
<td>0.50</td>
<td>0.55</td>
<td>0.61</td>
<td>0.00</td>
</tr>
<tr>
<td>Los Angeles, MSA Level, Case Shiller</td>
<td>1.22</td>
<td>1.42</td>
<td>1.60</td>
<td>1.74</td>
<td>0.00</td>
</tr>
<tr>
<td>San Francisco, MSA Level, Case Shiller</td>
<td>0.36</td>
<td>0.38</td>
<td>0.45</td>
<td>0.59</td>
<td>0.00</td>
</tr>
<tr>
<td>Washington D.C., MSA Level, Case Shiller</td>
<td>1.29</td>
<td>1.42</td>
<td>1.54</td>
<td>1.59</td>
<td>0.00</td>
</tr>
<tr>
<td>1990 - 1997: Housing Booms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Denver, MSA Level, Case Shiller</td>
<td>0.51</td>
<td>0.50</td>
<td>0.50</td>
<td>0.88</td>
<td>0.00</td>
</tr>
<tr>
<td>Portland, MSA Level, Case Shiller</td>
<td>0.41</td>
<td>0.52</td>
<td>0.49</td>
<td>0.69</td>
<td>0.00</td>
</tr>
<tr>
<td>1984-1989: Housing Booms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New York City, City Level, Furman</td>
<td>0.33</td>
<td>0.57</td>
<td>0.69</td>
<td>1.06</td>
<td>0.00</td>
</tr>
<tr>
<td>Boston, MSA Level, Case Shiller</td>
<td>0.65</td>
<td>0.70</td>
<td>0.75</td>
<td>0.84</td>
<td>0.00</td>
</tr>
<tr>
<td>1990 - 1997: Housing Busts</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>San Francisco, MSA Level, Case Shiller</td>
<td>-0.06</td>
<td>-0.08</td>
<td>-0.09</td>
<td>-0.13</td>
<td>0.00</td>
</tr>
<tr>
<td>Boston, MSA Level, Case Shiller</td>
<td>-0.05</td>
<td>-0.08</td>
<td>-0.13</td>
<td>-0.11</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Notes: This table shows the mean house price appreciation rates for neighborhoods grouped by quartile of initial housing prices, during different time periods, and using different housing price indices to measure the price appreciation. Quartile 4 has the highest initial price zip codes while quartile 1 has the lowest initial price zip codes within the city. Each row labels a city or metro area for a given time period using a given house price index to compute the appreciation rates. See the text for details about the different house price surveys. The last column reports the p-value of a t-test of whether the highest quartile mean appreciation is equal to the lowest quartile mean appreciation.
### Table 4: The Relationship Between House Price Growth Among Low Price Neighborhoods and Distance to High Price Neighborhoods

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Case-Shiller Zips (276 obs.)</td>
<td>Case-Shiller Zips (269 obs.)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log Dist. to Nearest</td>
<td>-0.058</td>
<td>-0.055</td>
<td>-0.039</td>
<td>-0.032</td>
<td>-0.054</td>
</tr>
<tr>
<td>High Price Neighborhood</td>
<td>(0.022)</td>
<td>(0.023)</td>
<td>(0.017)</td>
<td>(0.017)</td>
<td>(0.021)</td>
</tr>
<tr>
<td>Log Dist. to Nearest</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.061</td>
</tr>
<tr>
<td>High Price Neighborhood</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.029)</td>
</tr>
<tr>
<td>Bust_{t,t+k}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log Mean</td>
<td>-0.102</td>
<td>-0.140</td>
<td>-0.119</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Commuting Time_{t}</td>
<td>(0.134)</td>
<td>(0.107)</td>
<td>(0.107)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percent Change in</td>
<td></td>
<td></td>
<td></td>
<td>-0.377</td>
<td>-0.357</td>
</tr>
<tr>
<td>Commuting Time_{t,t+k}</td>
<td></td>
<td></td>
<td></td>
<td>(0.125)</td>
<td>(0.125)</td>
</tr>
<tr>
<td>Distance to River,</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Lakes, Oceans</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>City Fixed Effects</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Initial Income and House Price Controls</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Fraction African-American Controls</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Structure Age Controls</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Mean Dist. to High Price</td>
<td>3.9 miles</td>
<td>3.9 miles</td>
<td>3.9 miles</td>
<td>3.9 miles</td>
<td>3.9 miles</td>
</tr>
<tr>
<td>Mean Commuting Time</td>
<td>28.9 minutes</td>
<td>27.4 minutes</td>
<td>27.4 minutes</td>
<td>27.4 minutes</td>
<td>27.4 minutes</td>
</tr>
<tr>
<td>Mean Gr. Commute Time</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10.0%</td>
</tr>
</tbody>
</table>

Notes: This table shows the results from a regression of neighborhood house price growth between period t and t+k on the log distance from that neighborhood to a high housing price neighborhood, city fixed effects, and a vector of neighborhood controls. Our measure of a neighborhood is a zip code. The sample is restricted to only those zip codes for which we have a Case-Shiller housing price index. Furthermore, we restrict the sample to only include those zip codes within the city that had housing prices in period t that were in the bottom half of zip codes with respect to their period t housing price. In columns 1 and 2, we explore house price appreciation between 2000 and 2006. In columns 3-5, we explore house price appreciation between 1990 and 2000. There were 276 and 269 observations respectively in these sets of regressions. In all regressions, we control for city fixed effects and a vector of initial time t neighborhood controls. These controls include: the median house price within the zip code, the median income of residents in the zip code, the former two variables interacted with the city wide house price increase, the fraction of residents that were Black in the zip code, and dummies for the average age of the housing stock within the zip code. In columns 2, 4, and 5, we also control for the log of the mean commuting time of residents within the zip code and log distance to oceans, rivers, and lakes (if they exist within the city limits). The variable of interest is the log distance from the center of the zip code (in miles) to the center of the nearest zip code that is in the top quartile of median housing prices within the city. We measure time t housing prices using the time t U.S. Census. Likewise, all time t measures of income, race, commuting time, and age of the housing stock comes from the corresponding time t U.S. Census. Given that some cities experienced a housing price bust between 1990 and 2000, we interacted the log distance to high price variable with whether the city experienced a housing price bust in column 5. In columns 4 and 5, we also control for the change in commuting time within the neighborhood between t and t+k. We can only do this for the 1990-2000 period given that we do not have a commuting time measure at the zip code level in 2006.
Table 5: Relationship Between House Price Growth and Measures of Gentrification

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case-Shiller Zips</td>
<td>582 obs.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percent Change in Neighborhood Median Income</td>
<td>0.32</td>
<td>(0.08)</td>
<td></td>
</tr>
<tr>
<td>Percentage Point Change in Neighborhood Poverty Rate</td>
<td>-1.08</td>
<td>(0.21)</td>
<td></td>
</tr>
<tr>
<td>Percent Change in Mean Tenure in Home</td>
<td>-0.17</td>
<td>(0.07)</td>
<td></td>
</tr>
<tr>
<td>City Fixed Effects</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Initial Income and House Price Controls</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Fraction African-American Controls</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Table shows the coefficient on measures of gentrification in separate regressions of housing price growth on each measure of gentrification and metro area fixed effects. The variation in the table is across neighborhoods within a center city. The measure of house price growth in these regressions is the percentage increase in house prices at the zip code level using the Case-Shiller index between 1990 and 2000. Our measures of gentrification are from zip code data from the Censuses. Initial income and house price controls include initial period median household income, initial period median home value, and each of the former interacted with MSA home price growth over the period.

Table 6: House Price and Commuting Time Response to Industry Earnings Shocks of Own and Adjacent Neighborhoods

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependent Variable:</td>
<td>Neighborhood House Price Growth</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Predicted Income Shock</td>
<td>9.76</td>
<td>6.68</td>
<td>6.91</td>
<td>6.14</td>
</tr>
<tr>
<td>(1.43)</td>
<td>(1.68)</td>
<td>(1.72)</td>
<td>(1.52)</td>
<td></td>
</tr>
<tr>
<td>Avg. Neighborhood Shock: 1-10</td>
<td>8.39</td>
<td>8.68</td>
<td>8.08</td>
<td></td>
</tr>
<tr>
<td>(1.76)</td>
<td>(1.82)</td>
<td>(1.72)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avg. Neighborhood Shock: 11-20</td>
<td>-1.11</td>
<td>0.17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1.84)</td>
<td>(1.72)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avg. Neighborhood Shock: 21-30</td>
<td>-2.28</td>
<td>-1.51</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1.75)</td>
<td>(1.63)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log Commuting Time</td>
<td>-0.51</td>
<td>-0.54</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.07)</td>
<td>(0.12)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentage Change in Commuting Time</td>
<td>-0.51</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>City Fixed Effects</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Initial House Price Controls</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Fraction African-American Controls</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Distance to Rivers, Lakes, Oceans</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Columns (1)-(4) of this tables reports the coefficients of different variables from a regression of house price change between 1990 and 2000 (measured by growth in the Case-Shiller index) in neighborhood i of city j on the predicted income change in neighborhood i of city j during the same period as well as the average predicted income shock of neighborhood of various degrees of proximity to neighborhood i of city j. Our measure of neighborhood in this regression is zip code. In the regression, we partition proximity to neighbors based on the 10 zip codes that are spatially closest (measured in distance to midpoint) to the reference zip code. Avg. Neighborhood Shock: 11-20 is a simple average of the income shock over the zip codes that were the 11th to the 20th closest to the reference zip code. Neighborhood income shocks are predicted using the industry mix of residents in the neighborhood. See text for details. The regression also includes controls for the initial house price in the zip code (from the 1990 census), log commuting time of residents in the zip code (from the 1990 census), the change in commuting time of residents in the zip code (between the 1990 and 2000 census), fraction African American, distance to rivers, lakes, and oceans, and city fixed effects. The regression is restricted to 582 zip codes in the Case-Shiller data residing within the main sample cities. In the last column, the dependent variable is the growth in commuting time of the residents in neighborhood i of city j between 1990 and 2000.
Table 7: Relationship Between Measures of Gentrification and Industry Earnings Shocks of Own and Adjacent Neighborhoods

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>(1) Percentage Change in Median Income</th>
<th>(2) Percentage Point Change in Poverty Rate</th>
<th>(3) Percentage Change in Tenure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicted Income Shock</td>
<td>3.29</td>
<td>-0.28</td>
<td>-3.25</td>
</tr>
<tr>
<td></td>
<td>(1.16)</td>
<td>(0.35)</td>
<td>(1.18)</td>
</tr>
<tr>
<td></td>
<td>(1.64)</td>
<td>(0.46)</td>
<td>(1.55)</td>
</tr>
<tr>
<td>Avg. Neighborhood Shock: 11-20</td>
<td>-1.90</td>
<td>0.43</td>
<td>-1.08</td>
</tr>
<tr>
<td></td>
<td>(1.33)</td>
<td>(0.38)</td>
<td>(1.68)</td>
</tr>
<tr>
<td>Avg. Neighborhood Shock: 21-30</td>
<td>-0.75</td>
<td>0.46</td>
<td>0.63</td>
</tr>
<tr>
<td></td>
<td>(1.33)</td>
<td>(0.37)</td>
<td>(1.70)</td>
</tr>
<tr>
<td>Percentage Change in Commuting Time</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>City Fixed Effects</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Initial House Price Controls</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Fraction African-American Controls</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Distance to Rivers, Lakes, Oceans</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

This table reports the coefficients of different variables from a regression of measures of neighborhood gentrification between 1990 and 2000 on measures of predicted neighborhood income shock between 1990 and 2000 and the average predicted income shock of adjacent neighborhoods between 1990 and 2000. See the note to Table 6 for details. The regression also includes controls for growth in commute time, initial housing prices, fraction African American, distance to rivers, lakes, and oceans, and city fixed effects. Our three measures of gentrification are as defined in the note to Table 5. As in Table 6, the sample for this regression are the Case-Shiller zip codes (restricting the sample to only those in the main city of the metro areas sampled by Case-Shiller).
Table 8: Income Growth, Supply Constraints, and Cross City Differences in House Price Appreciation

<table>
<thead>
<tr>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
<th>(9)</th>
<th>(10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Growth in Inc</td>
<td>1.82</td>
<td>1.82</td>
<td>2.84</td>
<td>0.80</td>
<td>3.58</td>
<td>2.74</td>
<td>4.12</td>
<td>0.96</td>
<td></td>
</tr>
<tr>
<td>(0.17)</td>
<td>(0.18)</td>
<td>(0.37)</td>
<td>(0.37)</td>
<td>(0.38)</td>
<td>(0.32)</td>
<td>(0.62)</td>
<td>(0.48)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Saiz Housing Supply Elasticity</td>
<td>-0.53</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.76</td>
</tr>
<tr>
<td>Growth in Inc</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.16)</td>
</tr>
<tr>
<td>(0.16)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.26)</td>
</tr>
<tr>
<td>1(Low Saiz Supply Elasticity)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.42</td>
</tr>
<tr>
<td>Growth in Inc</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.45)</td>
</tr>
<tr>
<td>(0.45)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.71)</td>
</tr>
<tr>
<td>1(Med Saiz Supply Elasticity)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.27</td>
</tr>
<tr>
<td>Growth in Inc</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.45)</td>
</tr>
<tr>
<td>(0.45)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.96)</td>
</tr>
<tr>
<td>Saiz Housing Supply Elasticity</td>
<td>0.01</td>
<td>-0.00</td>
<td>0.05</td>
<td>-0.20</td>
<td>-0.16</td>
<td>-0.11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.02)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1(Low Saiz Supply Elasticity)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.12</td>
</tr>
<tr>
<td>Growth in Inc</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.05)</td>
</tr>
<tr>
<td>(0.05)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.07)</td>
</tr>
<tr>
<td>1(Med Saiz Supply Elasticity)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.15</td>
</tr>
<tr>
<td>Growth in Med Commute</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.05)</td>
</tr>
<tr>
<td>(0.05)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.06)</td>
</tr>
<tr>
<td>Growth in Med Commute</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.07)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.07)</td>
</tr>
<tr>
<td>R²</td>
<td>0.49</td>
<td>0.00</td>
<td>0.49</td>
<td>0.53</td>
<td>0.56</td>
<td>0.30</td>
<td>0.33</td>
<td>0.50</td>
<td>0.51</td>
</tr>
<tr>
<td>0.50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This table reports the coefficient of a regression of growth in house prices at the metropolitan area level from 1990 to 2000 and from 2000 to 2006 against the growth in median income at the metropolitan area level from 1990 to 2000 and from 2000 to 2006, Saiz’s measure of supply elasticity, and the growth in commuting time for residents in the metro area between 1990 and 2000. The growth in house prices at the metro level came from the FHFA repeat sales price index. Median income for 1990 and 2000 and median commuting times come from the Census. See text for details of Saiz’s measure of housing supply elasticity. 1(Low Saiz Supply Elasticity) and 1(Med Saiz Supply Elasticity) are dummy variables indicating whether the MSA is in the lowest or middle tercile of housing supply elasticity.
D1 Appendix (Included on the Journal’s Website)

D1.1 Proof of Proposition 1

Most of the proof of Proposition 1 is in the text. As we argue in the text, we are left to check only that

\[ U^R(i) \leq \bar{U}^R \text{ for all } i \in [I_t, \bar{I}_t], \]
\[ U^P(i) \leq \bar{U}^P \text{ for all } i \in [0, I_t], \]

where \( U^s(i) \) is defined in expression (5). Using expression (6), these two conditions can be rewritten as

\[ K_R (A + H_t(i))^{\frac{\delta_R}{\delta_P}} \leq K_P (A + H_t(i))^{\frac{\delta_R}{\delta_P}} + \frac{r}{1 + r} (C_R - C_P) \text{ for all } i \in [I_t, \bar{I}_t], \]
\[ K_P (A + H_t(i))^{\frac{\delta_P}{\delta_R}} \leq K_R (A + H_t(i))^{\frac{\delta_P}{\delta_R}} - \frac{r}{1 + r} (C_R - C_P) \text{ for all } i \in [0, I_t]. \]

Combining (8) with (12) and (13) we obtain

\[ K^R = \frac{r}{1 + r} \left[ C^P \left( \frac{A}{A + \gamma} \right)^{-\frac{\delta_P}{\delta_R}} + (C_R - C_P) \right] (A + \gamma)^{-\frac{\delta_P}{\delta_R}}, \]
\[ K^P = \frac{r}{1 + r} C^P A^{-\frac{\delta_P}{\delta_R}}, \]

Using these expressions, condition (18) can be rewritten as

\[ \left( \frac{A + H_t(i)}{A + \gamma} \right)^{\frac{\delta_P - \delta_R}{\delta_P}} \leq 1 + \left( \frac{C_R - C_P}{C^P} \right) \left( \frac{A}{A + H_t(i)} \right)^{\frac{\delta_P}{\delta_R}}. \]

for all \( i \in [I_t, \bar{I}_t] \). This implies that \( H_t(i) < \gamma \) and hence the RHS is not smaller than 1 and that, if \( \delta_R \geq \delta_P \), the LHS is not bigger than 1. Hence, \( \delta_R \geq \delta_P \) is a sufficient condition for this condition to be satisfied. Notice that if \( C^R = C^P \), this is also a necessary condition.

Next, condition (19) can be rewritten as

\[ \left( \frac{A + H_t(i)}{A + \gamma} \right)^{\frac{\delta_R - \delta_P}{\delta_R}} \leq 1 + \left( \frac{C_R - C_P}{C^P} \right) \left( \frac{A}{A + \gamma} \right)^{\frac{\delta_P}{\delta_R}} \left[ 1 - \left( \frac{A + \gamma}{A + H_t(i)} \right)^{\frac{\delta_P}{\delta_R}} \right] \]

for all \( i \in [0, I_t] \). In these locations, by construction, \( H_t(i) > \gamma \), which implies that the RHS is not smaller than 1 and that, if \( \delta_R \geq \delta_P \) the LHS is not bigger than 1. Hence, \( \delta_R \geq \delta_P \) is also a sufficient condition for this equation to hold. Again, it is also a necessary condition if \( C^R = C^P \). Hence, this completes the proof that a fully segregated equilibrium exists if \( \delta^P \leq \delta^R \).

D1.2 Proof of Proposition 2

The initial price schedule is:

\[ p_t(i) = \begin{cases} 
C^P \left( 1 + \frac{\gamma}{A} \right)^{\frac{\delta_P}{\delta_R}} + C^P - C^P \left( 1 + \min \left( \frac{\gamma + i - 1}{A + \gamma} \right) \right)^{\frac{\delta_P}{\delta_R}} \text{ for } i \in [0, I_t] \\
C^P \left( 1 + \max \left( \frac{\gamma + i - 0}{A - \gamma} \right) \right)^{\frac{\delta_P}{\delta_R}} \text{ for } i \in [I_t, \bar{I}_t] 
\end{cases} \]
First, notice that if \( i \geq I_t + \gamma \), then \( p_t (i) = C_P \), and if \( i < I_t + \gamma \), then \( p_t (i) > C_P \). Also, if \( i < I_t - \gamma \), then \( p_t (i) = \bar{p} \), where

\[
\bar{p} = \left[ C_P \left( 1 + \frac{\gamma}{A} \right)^{\frac{\delta_R}{\beta}} + C_P - C_P \right] \left( 1 + \frac{\gamma}{A + \gamma} \right)^{\frac{\delta_R}{\beta}}.
\]

Now, imagine that a permanent unexpected income shock hits the economy, so that the income of both rich and poor increase by a proportion of \( \phi \). Denote with a primus the equilibrium objects after the shock. We have

\[
p_{t+1} (i) = \begin{cases} 
\left( \frac{A + \gamma + \min \{ \gamma, \gamma_i - i \} }{ A + \gamma + \min \{ \gamma, \min \} } \right)^{\frac{\delta_R}{\beta}} & \text{for } i \in [0, I_t] \\
\left( \frac{A + \gamma}{ A + \gamma + \min \{ \gamma, \gamma_i - i \} } \right)^{\frac{\delta_R}{\beta}} + \left( \frac{C_P - C_P}{ A + \gamma + \min \{ \gamma, \gamma_i - i \} } \right)^{\frac{\delta_R}{\beta}} & \text{for } i \in [I_t, I_{t+1}] \\
\left( \frac{A + \max \{ \gamma, I_{t+1} - i \} + i }{ A + \max \{ \gamma, I_{t+1} - i \} } \right)^{\frac{\delta_R}{\beta}} & \text{for } i \in [I_{t+1}, \bar{I}_t] 
\end{cases}
\]  

(21)

Also, from equations (14) and (15), we obtain \( I_{t+1} > I_t \) and \( \bar{I}_{t+1} > \bar{I}_t \). Then, if \( i < I_t - \gamma \), it must be that \( p_{t+1} (i) / p_t (i) = 1 \), which implies that

\[
E_{t+1} \left[ \frac{p_{t+1} (i)}{p_t (i)} \right] | p_t (i) = \bar{p} = 1.
\]

Moreover, \( I_{t+1} > I_t \), together with expression (21), immediately implies that \( p_{t+1} (i) / p_t (i) \geq 1 \) for \( i > I_t - \gamma \), and hence

\[
E_{t+1} \left[ \frac{p_{t+1} (i)}{p_t (i)} \right] | p_t (i) < \bar{p} > 1,
\]

which proves the first statement or the proposition.

We now want to prove the second statement of the proposition, that is, that the price ratio \( p_{t+1} (i) / p_t (i) \) is non-increasing in \( p_t (i) \). First, notice that \( p_t (i) \) is non-increasing in \( i \), so proving that \( p_{t+1} (i) / p_t (i) \) is non-increasing in \( p_t (i) \) is equivalent to prove that \( p_{t+1} (i) / p_t (i) \) is non-decreasing in \( i \). The ratio \( p_{t+1} (i) / p_t (i) \) is continuous and differentiable except at a finite number of points. Hence, in order to prove that it is non-decreasing in \( i \), it is enough to show that \( d [p_{t+1} (i) / p_t (i)] / di \) is non-negative, for all \( i \) where this derivative exists. Let us show that.

For \( i \in [0, I_t - \gamma] \), \( p_{t+1} (i) / p_t (i) = 1 \) and hence \( p_{t+1} (i) / p_t (i) \) is constant in \( i \). For \( i \in [I_t - \gamma, I_t] \), we have that

1. if \( I_t - \gamma < i < I_{t+1} - \gamma \), then

\[
d \left( \frac{p_{t+1} (i)}{p_t (i)} \right) = \frac{\delta_R}{\beta} \left( A + 2 \gamma \right)^{\frac{\delta_R}{\beta}} \left( A + \gamma + I_t - i \right)^{-\frac{\delta_R}{\beta} - 1} > 0 \quad \text{(22)}
\]

2. if \( I_t - \gamma < i < \min \{ I_{t+1} - \gamma, I_t \} \), then

\[
d \left( \frac{p_{t+1} (i)}{p_t (i)} \right) = \frac{\delta_R}{\beta} \left( A + \gamma + I_t - i \right)^{\frac{\delta_R}{\beta}} \left[ \frac{1}{A + \gamma + I_t - i} - \frac{1}{A + \gamma + I_{t+1} - i} \right] > 0 \quad \text{(23)}
\]

given that \( I_t < I_{t+1} \).
For $i \in [I_t, I_t + \gamma]$ we have that

1. if $I_t < i < \min \{I_t + \gamma, I_{t+1} - \gamma\}$

$$
\frac{d}{di} \left( \frac{p_{t+1}(i)}{p_t(i)} \right) = \frac{\ddot{C}}{\beta} \left( \frac{A + \gamma + I_t - i}{A + \gamma + I_{t+1} - i} \right)^{\frac{\delta_R}{\beta}} \left[ \frac{1}{A + \gamma + I_t - i} - \frac{1}{A + \gamma + I_{t+1} - i} \right] > 0
$$

(24)

where

$$
\ddot{C} = \left[ \frac{(A + \gamma) \delta_R}{\beta} \right] + \frac{C_R - C_P}{C_P}
$$

2. if $\max \{I_t, I_{t+1} - \gamma\} < i < \min \{I_{t+1}, I_t - \gamma\}$

$$
\frac{d}{di} \left( \frac{p_{t+1}(i)}{p_t(i)} \right) = \frac{\ddot{C}}{\beta} \left( \frac{A + \gamma + I_{t+1} - i}{A + \gamma + I_t - i} \right)^{\frac{\delta_R}{\beta}} \left[ \frac{\delta_P}{\delta_R} - \frac{\delta_R}{\delta_P} < \frac{A + \gamma + I_{t+1} - i}{A + \gamma + I_t - i},
$$

(25)

which is true if the shock is big enough and $I_{t+1} - I_t$ is big enough;

3. if $\max \{I_t, I_{t+1}\} < i < I_t + \gamma$

$$
\frac{d}{di} \left( \frac{p_{t+1}(i)}{p_t(i)} \right) = \frac{\delta_P}{\beta} \left( \frac{A + \gamma + I_t - i}{A + \gamma + I_{t+1} - i} \right)^{\frac{\delta_R}{\beta}} \left[ \frac{1}{A + \gamma + I_t - i} - \frac{1}{A + \gamma + I_{t+1} - i} \right] > 0.
$$

(26)

This proves that, if the shock is big enough, the second statement of the proposition holds.

**Proof of Proposition 3**

The proof of this Proposition is straightforward. Imagine that at time $t + 1$ the economy is hit by an unexpected and permanent increase in income, that is, $y^s_{t+1} = \phi y^s_t$ with $\phi > 1$ for $s = P, R$. From expressions (14) and (15) it follows that $I_{t+1} > I_t$ and $\bar{I}_{t+1} > \bar{I}_t$. Then, from expression (21), we immediately obtain that for all $i \leq I_{t+1} - \gamma$, that is, for all $i$ such that $p_t(i) = C^P$, $d (p_{t+1}(i) / p_t(i)) / di < 0$, as we wanted to show.

**D1.3 Proof of Proposition 4**

First, notice that at time $t$, each location $i$ may lie in four possible intervals that implies different pricing behavior: $[0, I_t - \gamma]$, $[I_t - \gamma, I_t]$, $[I_t, I_t + \gamma]$, and $[I_t, \bar{I}_t]$. From expression (20), it is immediate that prices at time $t + 1$ in each location $i$ are weakly increasing in $I_{t+1}$, whenever $i$ is in the same type of interval at $t$ and $t + 1$. From expression (14) where $y^s$ is substituted by $\phi y^s$ for $s = R, P$, $I_{t+1}$ is non-decreasing in $\phi$ and hence prices are weakly increasing in $\phi$ for all $i$ which remain in the same type of interval. Let us consider any $\phi^A > \phi^B > 1$, with $I^A_{t+1} > I^B_{t+1}$. Then all $i \in [0, I^A_{t+1} - \gamma]$ are also in $[0, I^B_{t+1} - \gamma]$, but some $i \in [I^B_{t+1} - \gamma, I^B_{t+1} + \gamma]$ may be in $[0, I^A_{t+1} - \gamma]$ or some $i \in [I^B_{t+1} - \gamma, I^B_{t+1} + \gamma]$ may be in $[I^A_{t+1} - \gamma, I^A_{t+1}]$. 48
Given that, from inspection of expression (20), \( p_{t+1}(i) \) is non-increasing in \( i \), this implies that aggregate prices \( P_{t+1} \) must be non-decreasing in \( \phi \). Hence, if at time \( t+1 \) the economy is hit by an unexpected and permanent increase in \( \phi \), then \( P_{t+1} \) is going to be higher, the larger is the increase in \( \phi \). Given that \( P_t \) is given, this immediately proves the first statement of the proposition that the percentage increase in aggregate price is higher the larger is the increase in \( \phi \).

Second, we want to prove the second statement of the proposition, that

\[
\frac{d^2 (p_{t+1}(i)/p_t(i))}{dp_t(i) d\phi} \geq 0
\]

for all \( p_t(i) > C^P \) where the derivative is well-defined. Equations (22)-(26) in the proof of Proposition (2) define \( d[p_{t+1}(i)/p_t(i)]/di \) for all \( i \) where this derivative is well-defined and \( p_t(i) > C^P \). If the increase in \( \phi \) is big enough, \( d[p_{t+1}(i)/p_t(i)]/di > 0 \) for all \( p_t(i) > C^P \). Moreover, by inspection, it is easy to see that \( d[p_{t+1}(i)/p_t(i)]/di \) is increasing in \( I_{t+1} \), and hence increasing in \( \phi \), whenever \( i \) is in the same type of interval after a small or a large shock, say \( \phi^A \) or \( \phi^B \). Moreover, given that \( I^A_{t+1} > I^B_{t+1} \), \( i \) may lie in different types of interval in the two cases. In particular, it could be that \( \min \{ I^B_{t+1} - \gamma, I^A_t \} < i < I^A_t \) but \( I^B_t - \gamma < i < \min \{ I^A_t - \gamma, I^B_t \} \), or that \( \max \{ I^H_t, I^L_B - \gamma \} < i < I^H + \gamma \) and \( I^H < i < \min \{ I^H + \gamma, I^L_t - \gamma \} \), or that \( I^L_B < i < I^H + \gamma \) but \( \max \{ I^H, I^L_A - \gamma \} < i < I^H + \gamma \). It is easy to see that expression (22) is not smaller than expression (23) and that expression (24) is not smaller than expression (25). Finally expression (25) is bigger than expression (26) iff

\[
\left( \frac{A + \gamma + I_{t+1} - i}{A + \gamma} \right)^{\frac{\delta_R - \delta_P}{\delta_P}} \left[ 1 - \frac{(\delta_R - \delta_P)(A + \gamma + I_t - i)}{\delta_P (I_{t+1} - I_t)} \right] > 1,
\]

which is true if the shock is large enough so that \( I_{t+1} - I_t \) is big enough, as we assumed. This proves that \( d^2 [p_{t+1}(i)/p_t(i)]/didi\phi \) is positive for all \( i \) such that the derivative exists and \( p_t(i) > C^P \). Given that \( p_t(i) \) is non-increasing in \( i \), this completes the proof of the second claim of the proposition.

### D.1.4 Proof of Proposition 5

Consider two cities, \( A \) and \( B \), with both \( y^A_R > y^B_R \) and \( y^A_B > y^B_B \). First, we want to prove the claim that if at time \( t+1 \) they are both hit by an unexpected and permanent increase in income of rich and poor of the same proportion \( \phi \), the percentage increase in the aggregate price level \( P_t \) is larger in city \( A \). The two cities are exactly the same except for the size of the city and of the rich neighborhoods, which, from expressions (14) and (15), are so that \( I^A_t > I^B_t, I^A_{t+1} > I^B_{t+1}, I^A_t > I^B_t \) and \( I^A_{t+1} > I^B_{t+1} \). Hence city \( A \) has a larger center and a larger size overall. After the income increase, prices do not change for all \( i < I^A_j - \gamma \) and for all \( i > I^A_{t+1} + \gamma \) if \( I^A_{t+1} + \gamma < I^L_t \), for both \( j = A \) and \( j = B \). Moreover, given that \( I^A_{t+1} > I^B_{t+1} \), the growth rate \( p^A_{t+1}(i)/p^A_t(i) \) is weakly higher than \( p^B_{t+1}(i + I^B_t - I^A_t)/p^B_t(i + I^B_t - I^A_t) \) for all \( i \in [I^A_t - \gamma, I^A_{t+1}] \). This implies that the gross growth rate in aggregate prices is also higher in city \( A \) than in city \( B \) if the shock is big enough that the higher expansion in city \( A \) dominates the zero growth rate in the rich neighborhoods in the center.
To prove the second claim notice that the price in city $A$ at time $t$ for $i \in [I^A_t - \gamma, I^A_t + \gamma]$ are exactly the same as in city $B$ for $i \in [I^B_t - \gamma, I^B_t + \gamma]$. However, $I^A_t > I^B_t$, so that the interval $[0, I^A_t]$ is larger than $[0, I^B_t]$. When the income increase hits both cities, expressions (14) and (15) give that both $I$ and $\bar{I}$ increase more in city $A$. Hence, the expression for $d(p_{t+1}(i)/p_t(i))/di$ in city $A$ for all $i \in [I^A_t - \gamma, I^A_t + \gamma]$ is equivalent to $d(p_{t+1}(i)/p_t(i))/di$ in city $B$ for all $i \in [I^B_t - \gamma, I^B_t + \gamma]$ if city $B$ was facing a larger income increase. From the proof of Proposition 4, this immediately implies that $d(p^A_{t+1}(i)/p^A_t(i))/di > d(p^B_{t+1}(i+I^A_t-I^B_t)/p^B_t(i+I^A_t-I^B_t))/di$, that is, $g^A_t(p) > g^B_t(p)$ for all $p \in (\bar{p}, C^P)$. Finally, the gross growth rate of prices in all locations where the initial price was $\bar{p}$ is everywhere equal to 1 so that is not sensitive to $i$ and $g^A_t(\bar{p}) = g^B_t(\bar{p})$. This completes the proof that $g^A_t(p) \geq g^B_t(p)$ for all $p > C^P$ whenever this derivative is well defined.

D2 Data Appendix (Included on the Journal’s Website)

D2.1 Chicago Deeds

Property records for Chicago, IL were downloaded from a section of the Chicago Tribune’s website which gives access to Record Information Services’ Property Transfers Database. During the spring of 2008, we downloaded all records that were available on the website for property transfers pertaining to properties located in the City of Chicago through May 30, 2008. We dropped any records that had prices recorded as either zero or ten dollars. We also dropped any records which had not already been geocoded and which we were also unable to geocode. Finally, we dropped transactions in the top and bottom percentile of the price distribution in each year. (The top and bottom percentile of the price distribution for all years were transactions over $1,400,000 or under $17,600, respectively.) Next, we downloaded building characteristic data for each parcel identification number present in the Record Information Services’ Property Transfers Database. We obtained this building characteristic data from the Cook County Tax Assessor’s website. While the Cook County Tax Assessor database contains a rich set of building characteristics for single-family and multi-family homes, only the age of the structure is available for condominiums. The variables include indicators for whether the property is a condominium or a multi-family building, and indicators for the age of the building. We ran a regression of log price on these indicator variables and a vector of community area * year indicator variables. To form the hedonic index, we evaluated the regression equation at the mean building characteristic values. Thus, differences in the index value are driven by differences in the estimates of the community area * year effects, and the index can be interpreted as representing the value of the mean structure type in different neighborhoods at different points in time.

39 All prices are left in nominal terms until hedonic index is computed. All prices are then converted to year 2000 dollars.
40 Web Address: http://cookcountyassessor.com/.
D2.2 Furman Center Repeat Sales Index for NYC

We use a repeat sales index for New York City community districts that was produced by NYU’s Furman Center for Real Estate and Urban Policy. These data consist of a repeat sales index for each community district in New York City reported at an annual frequency and running from 1974 through 2008.41

D2.3 Census Data

We use tabulations from the United States Census from 1980, 1990, and 2000. Some data come from Summary File 1 (SF1) which contains 100 percent counts and information for all people and housing units, while other data come from Summary File 3 (SF3) which contains a 1-in-6 sample and is weighted to represent the entire population. In general, we obtained data from the 2000 Census from the American Factfinder Website.42 However, zip code and census tract tabulations for the 1980 and 1990 Censuses were obtained from the Inter-University Consortium for Political and Social Research.43 Finally, variables indicating census tracts that had not changed boundaries between 1980 and 1990 were obtained from the Neighborhood Change Database produced by Geolytics.44 Median home value data for NYC community districts and are available from NYC’s GIS department at http://gis.nyc.gov.

D2.4 Chicago Community Area Mean Commuting Time

To compute mean commute time for each community area in Chicago, we aggregate from the 2000 Census data by census tract. Specifically, the 2000 Census reports the number of workers 16 years and over that do not work at home and whose commuting times are in each of the following bins: less than 5, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-59, 60-89, and 90 minutes or over. Each Chicago community area is made up of a group of census tracts. We sum the number of workers in each bin over all the census tracts for each community area and then divide by the total number of workers in the community area to obtain the fraction of workers in each bin for each community area. Finally, we calculate mean commuting time for each community area by multiplying these fractions by the midpoint of the bin and summing across the bins. For the 90 minute and over bin we multiplied by 90.

D2.5 MSA-Level Housing Price Growth

Data on MSA-level housing price growth come from the, publicly available, Federal Housing Finance Agency MSA repeat sales indices (formerly called the OFHEO indices).45

---

41See http://www.furmancenter.org for more information.
42Web Address: http://factfinder.census.gov/.
43Web Address: http://www.icpsr.umich.edu/icpsrweb/ICPSR/.