working paper

10 08

Endogenous Gentrification and Housing-Price Dynamics

by Veronica Guerrieri, Daniel Hartley and Erik Hurst

FEDERAL RESERVE BANK OF CLEVELAND

Working Paper 10-08 July 2010

Endogenous Gentrification and Housing-Price Dynamics

by Veronica Guerrieri, Daniel Hartley and Erik Hurst

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 citywide 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 in which households fully segregate based on their income. In response to positive housing-demand shocks, the model predicts that the poor neighborhoods on the boundary with the rich ones are the most price elastic. 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.

JEL classication: R12, R21, I32
Keywords: gentrification, housing price dynamics, housing consumption externalities.

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, Tufts, and the Federal Reserve Banks of Atlanta, Boston, Chicago, Cleveland, Minneapolis, and St. Louis. They are particularly indebted to Daron Acemoglu, Gary Becker, Hoyt Bleakley, V. V. Chari, Morris Davis, Matt Kahn, Guido Lorenzoni, Enrico Moretti, Kevin Murphy, Matt Notowidigdo, John Quigley, Esteban Rossi-Hansberg, 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.

Veronica Guerrieri of the University of Chicago's Booth School of Business can be reached at Veronica.Guerrieri@chicagobooth.edu. Daniel Hartley of the Federal Reserve Bank of Cleveland can be reached at Daniel.Hartley@clev.frb.org. Erik Hurst of the University of Chicago's Booth School of Business can be reached at Erik.Hurst@chicagobooth.edu.

1 Introduction

It has been well documented that there are large differences in house price appreciation across US metropolitan areas.[1] 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 the 20 cities for which a Case-Shiller MSA index is publicly available, the standard deviation in real house price growth between 2000 and 2006 was 42 percent. Such variation is not a recent phenomenon. During the 1990s, the cross city standard deviation in house price growth was 21 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 was changing rapidly during the period when Harlem house prices were increasing at a relatively faster rate than other New York neighborhoods.[2]

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 in order to understand house price movements at the city (or MSA) level.[3] In addition, we show that 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.

[1] See, for example, Davis et al. (2007), Glaeser et al. (2008), Van Nieuwerburgh and Weill (2010), and Saiz (2010).

[2] See the article “No Longer Majority Black, Harlem Is In Transition” from the January 5th, 2010 New York Times.

[3] 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 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 one's 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 endogenously 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 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

[4]The importance of urban density in facilitating local consumption externalities has been recently emphasized in the work of Voith (1999), Voith and Gyourko (2002), Glaeser et al. (2001), Becker and Murphy (2003), Bayer et al. (2007), and Rossi-Hansberg et al. (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.

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 neighborhoods that are 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 et al. (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 after controlling for proximity to the jobs (commuting time of residents), 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.

We then test the causal effect of income shocks in a given neighborhood on property prices and gentrification in the surrounding neighborhoods. To do this, we follow the procedure used by Bartik (1991) and Blanchard and Katz (1992) in that we use the initial industry mix of residents in the neighborhood to

[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.

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 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.

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 (2010) 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 informs the set of policies they may choose to adopt when pursuing urban development plans.

Finally, our paper contributes to many literatures. First, we add to the small existing literature that has explored within city house price movements.[7] 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 et al. (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 et al. (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 test the model's predictions. During periods of declining housing demand in urban areas (like the suburbanization movement during the 1960s), the richer neighborhoods on the edge 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.[8]

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 et al. (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 shocks to housing demand like lower interest rates, changes in the skill premium, or even shocks to the productivity of 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.

[7]See, for example, Case and Mayer (1996) and Case and Marynchenko (2002) who examined house price movements during the 1980s and early 1990s across zip codes within Boston and across zip codes within Boston, Chicago and Los Angeles, respectively. Our work complements this literature by systematically examining within city movements in house prices during housing price booms and housing price busts across a large cross section of cities.

[8]There is a separate and important literature that looks at specific policy initiatives by local governments and explores how they affect the gentrification of surrounding neighborhoods. For example, Kahn et al. (2009) explore differences in land prices on the boundary of the Coastal Boundary Zone within California. Autor et al. (2010) explore differences in land prices that occur after the end of rent control in Cambridge, Massachusetts. Ellen et al. (2002) look at the effect of spill-overs from the subsidization of construction for owner-occupied housing within New York during the 1990s.

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 neighborhoods 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 across Chicago neighborhoods.

We chose 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 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. 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 within Chicago. 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.[9] 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.[10] We refer to this data as the Chicago Tribune Index. As we show in the online robustness appendix, if we use our methodology and compute the growth rate in housing prices for each Chicago zip code between 2000 and 2006, the growth rates found using our index correlates highly with the growth rates found using

[9]The hedonic regression used only single family home transactions since the Cook County Tax Assessor's office only lists detailed property characteristics for single family homes. The characteristics used in the hedonic regression were: log square footage and dummies for construction type (frame, masonry, combined), central air conditioning, number of bathrooms, attic (full/finished), basement (full/finished), garage (detached/2 car +), and age of the structure. We compute our price index for year 2000 using data from 2000 and 2001. Likewise, we compute our price index for 2006 using 2005 and 2006 data. We pool the data across the two years to ensure that we have enough transactions within each community area for each year. See the Data Appendix for specific details of how we compute the price index.

[10]See http://en.wikipedia.org/wiki/Community_areas_of_Chicago.

either the Case Shiller or Zillow indices over the same time period.[11] In subsequent sections, we use these two other indices when computing zip code level house price growth for other cities.

Panel A of Figure 1 shows the variation in year 2000 house prices (in year 2000 dollars) across Chicago community areas. 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 borders Lake Michigan (top right) are much higher than property prices in the south-eastern part of Chicago which also borders Lake Michigan. Panel B of Figure 1 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. Here we group neighborhoods into quintiles based on mean income. Panel B 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.[12] 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 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.[13] Our second measure of commuting costs is the distance to the "Chicago Loop" which is the name for the center business district of Chicago. As seen from column 1 of Table 1, the traditional measures of within city house price differences (proximity to jobs and fixed amenities) explain 63 percent of the variation in land prices across Chicago. Places with shorter commuting times and places closer to Lake Michigan have higher house prices.

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 with the controls for commuting costs and distance to Lake

[11]The online robustness appendix which has a variety of robustness excerises referred to in the paper can be found on the authors' webpage.

[12]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 perfectly 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.

[13]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. For a full discussion of how we converted the actual Census data into the mean time commuting within each community area, see the Data Appendix.

Michigan (column 3). The results in these columns show that differences in income, by itself, explains 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-squared of including income to the controls in column 1 is roughly 11 percentage points. The point estimate of the coefficient suggests that an elasticity of neighborhood average house prices to neighborhood average income of 0.72. These results underlie the relationships shown in Figures 1A and 1B. 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.[14] 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 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, what we show is 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. 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 directly isolate 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 that it illuminates the potential mechanisms that explain variation 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

[14]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 et al. (2009). Likewise, both Schelling (1969) and Card et al. (2008) also show how the desirability to have neighbors of a given race could lead to racial segregation within a city.

to Lake Michigan and similar commuting times relative to the south side of Chicago. 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 spatial 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 our mechanism.

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.[15] Agents are fully mobile and can choose to live in any location i. Denote by $n_t^s(i)$ the measure of households of type s who live in location i at time t and by $h_t^s(i)$ the size of the house they choose. In each location, there is a maximum space that can be occupied by houses which is normalized to 1,[16] that is,

$$n_t^R(i)\, h_t^R(i) + n_t^P(i)\, h_t^P(i) \leq 1 \text{ for all } i, t.$$

Moreover, market clearing requires

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

[15]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).

[16]Our notion of space is uni-dimensional: if there is need for more space to construct houses the neighborhoods have 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 is the extreme case with infinite adjustment cost on the vertical dimension and no adjustment costs on the horizontal dimension. Our conjecture is that our mechanism would go through if we allow some costly adjustment in the vertical margin and we believe this is an interesting direction for future work.

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 γ. Let $H_t(i)$ denote the total space occupied by houses of rich households in the neighborhood around location i,[17] that is,

$$H_t(i) = \int_{i-\gamma}^{i+\gamma} h_t^R(j)\, n_t^R(j)\, dj. \tag{2}$$

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 their consumption more if they live in locations with higher $H_t(i)$. The utility of a 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 α, β, and δ^s are non negative scalars.[18] 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. We want to stress that all of the implications of our model go through even if $\delta^R = \delta^P$.

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 houses: 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 $\tilde{s}$ into houses of type s, he has to pay $C^s - C^{\tilde{s}}$. The (per square foot) price of a house for households of type s in location i at time t is equal to $p_t^s(i)$. Hence there is going to be construction in any empty location i as long as $p_t^s(i) \geq C^s$. Moreover, if the firm wants to construct a house of type s in a location occupied by a house of type $\tilde{s}$, he has to pay the converting cost and the additional cost of convincing households of type $\tilde{s}$ to leave. Hence, there is going to be construction of houses of type s in any location occupied by agents of type $\tilde{s}$ if $p_t^s(i) \geq C^s - C^{\tilde{s}} + p_t^{\tilde{s}}(i)$.

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.[19] The (per square foot) rent for a house of type s in location i at time t is denoted by $R_t^s(i)$. 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

[17] An alternative is to define the neighborhood externality $H_t(i)$ as the measure of rich households living in the neighborhood around location i (or even as their average income). However, this would make the model less tractable without affecting the substance of the mechanism. A more interesting extension would be to relax the assumption that a neighborhood has a fixed size and make the concept of a neighborhood more continuous. Again the main mechanism of the model would survive this change, but the price schedule would look smoother.

[18] Davis and Ortalo-Magné (2010) show that a Cobb-Douglas relationship between housing consumption and non-housing consumption fits the data well along a variety of dimensions.

[19] When the economy is hit by a positive demand shock, we will show that house prices appreciate by different amounts in different locations. If households own their houses this would introduce an additional source of heterogeneity in wealth which would complicate the analysis.

among intermediaries requires that for each location i the following arbitrage equations hold:

$$p_t^s(i) = R_t^s(i) + \left(\frac{1}{1+r}\right) p_{t+1}^s(i) \text{ for all } t, i, s. \tag{3}$$

3.2 Equilibrium

An equilibrium is a sequence of rent and price schedules $\{R_t^R(i), R_t^P(i), p_t^R(i), p_t^P(i)\}_{i \in R}$ and of allocations $\{n_t^R(i), n_t^P(i), h_t^R(i), h_t^P(i)\}_{i \in R}$ 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

$$\max_{c,h,i \in \mathcal{I}_t^s} c^\alpha h^\beta [A + H_t(i)]^{\delta^s},$$
$$s.t.\ c + hR_t^s(i) \le y^s,$$

where households take as given the function $H_t(i)$, the rent schedule $R_t^R(i)$, and the set $\mathcal{I}_t^s$ 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

$$h_t^s(i) = \frac{\beta}{\alpha+\beta} \frac{y^s}{R_t^s(i)} \text{ for all } t, s, i \in \mathcal{I}_t^s. \tag{4}$$

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 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 $i \in \mathcal{I}_t^s$.[20] Then it must be that

$$U_t^s(i) \equiv \alpha^\alpha \beta^\beta \left(\frac{y^s}{\alpha+\beta}\right)^{\alpha+\beta} \frac{[A + H_t(i)]^{\delta^s}}{R_t^s(i)^\beta} = \bar{U}_t^s \text{ for all } t, s, i \in \mathcal{I}_t^s. \tag{5}$$

This, in turn, requires that

$$R_t^s(i) = K^s [A + H_t(i)]^{\frac{\delta_s}{\beta}} \text{ for all } t, s, i \in \mathcal{I}_t^s, \tag{6}$$

for some constant K^s. This expression is intuitive, as rents must be higher in locations with a stronger externality. Moreover, rich households are more willing to pay higher rents for a given locational externality, all else equal.

Proposition 1. *If $\delta^R \ge \delta^P$, there exists an equilibrium with full segregation. If $C^R = C^P$, an equilibrium with full segregation exists if and only if $\delta^R \ge \delta^P$.*

[20] 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.

To prove this proposition we proceed by constructing 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. There may be other equilibria with full segregation with more centers of agglomeration of the rich households. It is interesting to notice that, as long as these centers are far enough from each other, the implications in terms of house prices are isomorphic to the equilibrium we focus on.

Let us now proceed to the construction of our equilibrium. As a normalization, we choose point 0 as the center of the city. It follows that $\mathcal{I}_t^R = [-I_t, I_t]$ and $\mathcal{I}_t^P = [-\bar{I}_t, -I_t,) \cup (I_t, \bar{I}_t]$, for some $\bar{I}_t > I_t > 0$. 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 $h_t^R(i)\, n_t^R(i)$ 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:

$$H_t(i) = \begin{cases} 2\gamma & \text{for } i \in [0, I_t - \gamma] \\ \max\{\gamma + I_t - i, 0\} & \text{for } i \in (I_t - \gamma, \bar{I}_t] \end{cases} . \tag{7}$$

That is, neighborhoods close to the city center are richer 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

$$K^R = R_t^R(I_t)(A+\gamma)^{-\frac{\delta_R}{\beta}} \text{ and } K^P = R_t^P\left(\bar{I}\right) A^{-\frac{\delta_P}{\beta}}, \tag{8}$$

so that we can rewrite the rent schedules as

$$R_t^R(i) = R_t^R(I_t)\left(1 + \frac{\min\{\gamma, I_t - i\}}{A+\gamma}\right)^{\frac{\delta_R}{\beta}} \quad \text{for } i \in [0, I_t], \tag{9}$$

$$R_t^P(i) = R_t^P\left(\bar{I}_t\right)\left(1 + \frac{\max\{\gamma + I_t - i, 0\}}{A}\right)^{\frac{\delta_P}{\beta}} \quad \text{for } i \in (I_t, \bar{I}_t]. \tag{10}$$

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 $p_t^P\left(\bar{I}_t\right) = C^P$ and $p_t^R(I_t) = p_t^P(I_t) + C^R - C^P$. 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

$$p_t^s(i) = \frac{1+r}{r} R_t^s(i) \text{ for all } t, i, s. \tag{11}$$

Combining this conditions we obtain

$$R_t^P\left(\bar{I}_t\right) = \frac{r}{1+r} C^P \text{ and } R_t^R(I_t) = R_t^P(I_t) + \frac{r}{1+r}\left(C^R - C^P\right), \tag{12}$$

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

$$R_t^P(I_t) = \frac{r}{1+r} C^P \left(\frac{A+\gamma}{A+\max\left\{\gamma + I_t - \bar{I}_t, 0\right\}} \right)^{\frac{\delta_P}{\beta}}. \tag{13}$$

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 $\mathcal{I}_t^R = [-I_t, I_t]$ 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 $\bar{I}_t$:

$$\begin{aligned} I_t &= \gamma + (A+2\gamma)^{-\frac{\delta_R}{\beta}} \left\{ \frac{\alpha}{\alpha+\delta} \frac{N^R y^R}{2K^R} - \frac{\beta}{\beta+\delta_R} \left[(A+2\gamma)^{\frac{\beta+\delta_R}{\beta}} - (A+\gamma)^{\frac{\beta+\delta_R}{\beta}} \right] \right\} &(14) \\ \bar{I}_t &= I_t + \gamma + A^{-\frac{\delta_P}{\beta}} \left\{ \frac{\alpha}{\alpha+\delta} \frac{N^P y^P}{2K^P} - \frac{\beta}{\beta+\delta_P} \left[(A+\gamma)^{\frac{\beta+\delta_P}{\beta}} - A^{\frac{\beta+\delta_P}{\beta}} \right] \right\}. &(15) \end{aligned}$$

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 the 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

$$\begin{aligned} U_t^R(i) &\leq \bar{U}_t^R \text{ for all } i \in \left[I_t, \bar{I}_t\right] \\ U_t^P(i) &\leq \bar{U}_t^P \text{ for all } i \in [0, I_t] \end{aligned}$$

where $U_t^s(i)$ 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.

Before proceeding to the analysis of the shocks, let us mention that Proposition 1 claims that an equilibrium with full segregation always exists when $\delta^R \geq \delta^P$, but not that no other equilibria exist. Actually, we can construct another equilibrium with partial segregation, where intervals with only poor people alternate with intervals where poor and rich people coexist. However, we can show that there is no equilibrium with full integration, that is, where poor and rich agents simultaneously live in every occupied location.

3.3 Demand shock

We are now interested in analyzing how house prices, both at an aggregate and at a disaggregate level, react to an unexpected increase in the demand for housing.[21] We will do so, by focusing on the equilibrium with full segregation that we constructed in the previous section.

In equilibrium, the aggregate price level is given by

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

where, from the analysis in the previous section,

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

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

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

For concreteness, we analyze the economy's reaction to a positive 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 want 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 into poor neighborhoods. This is what we refer to as endogenous gentrification. The house prices in gentrified neighborhoods are driven up due to our externality.

Let us define the function $g_t(.) : \left[C^P, \bar{p}\right] \mapsto [1, \infty)$, where $g_t(p)$ denotes the average gross growth rate

[21] The long run reaction of house prices would be symmetric in the case of a negative shock if we introduce some degree of depreciation that is big enough relative to the shock. However, after a negative shock the economy would not jump to the new steady state immediately, but there would be some transitional dynamics. Contact the authors if you are interested in the full analysis of a negative demand shock in the presence of depreciation.

between time t and $t+1$ in locations where the initial price is equal to p, that is,

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

The next proposition shows that after an unexpected permanent positive demand shock, the aggregate price level permanently increases and the price growth rate is higher in locations that had lower price levels initially, whenever prices are higher than the minimum level.[22]

Proposition 2. *Imagine that at time $t+1$ the economy is hit by an unexpected and permanent increase in income, that is, $y_{t+1}^s = \phi y^s$ 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 $\bar{I}_t$ 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 away from the rich neighborhoods. Most importantly, prices strictly increase in the rich neighborhoods that are not yet fully developed and, even more, in the poor neighborhoods that are physically close to the rich neighborhoods. 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 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 the closer to the richer neighborhoods.

Proposition 3. *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*

$$\frac{d\left(p_{t+1}(i)/p_t(i)\right)}{di} \leq 0 \textit{ for } p_t(i) = C^P.$$

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

[22] See the online appendix for all the proofs that are not in the text.

should appreciate the most during a city wide housing demand shock. This proposition underlies the variation in appreciation rates among the poorer neighborhoods. 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.

Proposition 4. *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 the growth rate in the aggregate price level is larger the larger is the increase in ϕ and, if the shock is large enough, $d^2 g_t(p)/dpd\phi \geq 0$ 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 a 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 preceding 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.[23] 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

[23]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)

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. In most of the analysis below, we will use Case-Shiller house price indices within cities, rather than within metropolitan areas. By focusing our results on housing price movements within the primary city within the MSA, we can hold such factors as the level of taxes and public goods fixed by including MSA fixed effects. In the online Robustness Appendix, we show that all our results are robust to using all zip codes within the MSA as opposed to just the zip codes in the primary city.

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 26 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 10 zip codes in either 1990 and 2000 or 2000 and 2006.[24]

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. 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.[25] 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. Restricting our analysis to only the primary Case-Shiller city (within each metro area) and to only the zip codes with price indices based on actual transaction data, we have data for 508 zip codes during the 1990-2000 period and for 497 zip codes during the 2000-2006 period.

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 imperfectly, take steps to minimize the effect of potential remodeling and renovations when computing the index.[26]

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

[25] New York City is an outlier. Only 9 zip codes within New York City had price indices computed using actual (non-imputed) data during the 2000-2006 period. To overcome this problem, we used a supplementary data set to explore the patterns in New York City. We discuss this data below.

[26] 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

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 perfectly capture 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. Our primary robustness analysis uses zip code level data on house prices from the Zillow Home Value Indices.[27] 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 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. The Zillow data is available at the monthly level for most zip codes starting in the late 1990s. Finally, some zip codes do not have enough transactions during the month to create a reliable house price index. The Zillow data that we have access to indicates which zip codes within the city that Zillow feels do not have enough transactions to create a reliable price index. We exclude such zip codes from our analysis.

The Zillow data, at least partially, overcomes some of the deficiencies of the Case-Shiller indices. First, the Zillow data allows for the attributes of the structure to change over time. Second, a reliable Zillow index is available for more zip codes than the Case-Shiller indices. The reason for this is that the Case-Shiller data is based off of repeat sales transactions while the Zillow data uses all sales regardless of whether or not they could match the sale with a previous transaction. The Zillow data will allow us to see if our results using the Case-Shiller data change in any substantial way when we include a broader set of zip codes. Neither the restricted set of zip codes nor the failure to control for changing structure attributes change any of our key empirical results in anyway. As we show in the online Robustness Appendix, none of our key results change when we use the Zillow data instead of the Case Shiller data. We choose to use the Case Shiller data for our main results shown in the paper and relegate the results using the Zillow data to the Robustness Appendix because the Zillow data is only available for the 2000-2006 period.

Also, below and in the robustness appendix, we show that our results are robust to using two other price indices which are available for the universe of zip codes within the cities of Chicago and New York. Again, these results suggest that the selection issues associated with the Case Shiller zip codes being only

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. 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.

[27] See http://www.zillow.com/wikipages/Real-Estate-Market-Reports-FAQ/ for details. We thank Amir Sufi for providing us with the zip code level indices including the information he received from Zillow on which zip codes had too few observations to make a reliable price index. We posted all such data on our web pages. See the online robustness appendix for details.

available for some of the zip codes within some cities is not biasing our results in any meaningful way. For Chicago, we use the Chicago tribune index described above. For New York City (NYC), we use the Furman Center repeat sales index which covers all of NYC. The Furman data use NYC community districts as its level of aggregation as opposed to zip codes and, as a result, have enough observations to make reliable indices for all areas within NYC.[28]

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.[29]

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 nominal house prices into real house prices.

5 Within City House Price Variation and Gentrification

In this section, we test the predictions of our model in three steps. First, we show descriptive features about 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 to systematically document 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 (row 1) and 1990-2000 (row 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 large variation in price appreciation across MSAs during both the 1990s and the 2000s. This is consistent with the well documented facts discussed in Davis et al. (2007), Glaeser et al. (2008),

[28] See http://furmancenter.org/. There are 59 community districts in NYC which represent clusters of several zip codes. The Furman data for NYC extend back to 1974.

[29] Additional information about the construction of these housing price indices can be found at http://www.fhfa.gov/.

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. Also, according to Proposition 3, the variation in housing price appreciation should be higher among poorer neighborhoods closer to richer neighborhoods. 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.[30]

Table 3 shows the relationship between the initial housing price and the subsequent housing price growth across neighborhoods within a city for a large selection of cities and metro areas during different time periods. 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 a few things from the results in Table 3.[31] First, within both the Chicago city and MSA, initially low price neighborhoods (quartile 4) appreciated at much higher rates than initially high price neighborhoods (quartile 1) during the 2000-2006 period. For example, within the city of Chicago, low price neighborhoods appreciated at close to 90 percent while high price neighborhoods appreciated at about 50 percent. Second, the results for Chicago are found in all other cities. For example, within the

[30] Let us mention that our model has also similar implications about the variation in rental prices within a city. Getting within city 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 metro area. He replicated the main empirical findings in our paper using the rental price data. Using the rental data, he found that the main patterns in our paper still hold. A discussion of these results can be found in Davis (2010).

[31] 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. Additionally, we show the same results using the Zillow data.

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 them, 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 depreciated 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.

Although not shown in Table 3, we also find that there is more variation in price changes among poorer neighborhoods than richer neighborhoods during a given time period. For example, within the New York metro area during the 2000 - 2006 period, the standard deviation of house price appreciation 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.

Finally, as we show in the online Robustness Appendix, the gap in appreciation rates between initially high priced and initially low priced neighborhoods is larger when the city wide property price changes were higher, as implied by Proposition 4. 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. When instead 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.

5.2 Housing Price Dynamics and Proximity to Rich Neighborhoods

One of the main predictions of our model is that variation 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:

$$\frac{\Delta P_{t,t+k}^{i,j}}{P_t^{i,j}} = \alpha + \beta \ln(Dist_t^{i,j}) + \Gamma X_t^{i,j} + \Psi Z_t^{i,j} + \mu_j + \epsilon_{t,t+k}^{i,j}$$

where $\Delta P_{t,t+k}^{i,j}/P_t^{i,j}$ 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 β, the coefficient on $\ln(Dist_t^{i,j})$ which measures the log of the distance (in miles) to the nearest zip code in the 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 $X_t^{i,j}$ 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 African American.[32] All of the X controls are 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 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, μ_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

[32] As shown by Rosenthal (2008), the age of the housing stock could be an important determinant of predicting which neighborhoods will subsequently gentrify. In the online robustness appendix we show that our results hold even when controlling for structure age.

whether there are differences in property price appreciation across neighborhoods based upon 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 a constant.[33]

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 since 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 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

[33]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 2006.

including controls for distance to jobs or distance to natural amenities like oceans, lakes, or rivers.

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 between 5.3 and 5.8 percent. In other words, zip codes that are within 1 mile of the high priced neighborhoods increased by between 10.6 and 11.6 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 about 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 neighborhoods and housing price appreciation.

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:

$$\frac{\Delta P_{t,t+k}^{i,j}}{P_t^{i,j}} = \alpha + \beta Y_{t,t+k}^{i,j} + \Gamma X_t^{i,j} + \mu_j + \epsilon_{t,t+k}^{i,j}$$

where $Y_{t,t+k}^{i,j}$ is some measure of gentrification in neighborhood i of city/metro area j during the period t to $t+k$ and where $\Delta P_{t,t+k}^{i,j}/P_t^{i,j}$, $X_t^{i,j}$, and μ_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 from 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.[34] Given that the mean increase in house price growth across the zip codes during this period was 28 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.

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 use variation in national earnings by industry between 1990 and 2000. Using that, we 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 (1991) and Blanchard and Katz (1992)).

Specifically, we estimate the following regression:

$$\begin{aligned}\frac{\Delta P_{t,t+k}^{i,j}}{P_t^{i,j}} &= \delta_0 + \delta_1 OwnIncShock_{t,t+k}^{i,j} + \delta_4 NeighborIncShock_{t,t+k}^{i,j} + \\ &\quad \Gamma X_t^{i,j} + \Psi Z_t^{i,j} + \mu_j + \epsilon_{t,t+k}^{i,j}\end{aligned}$$

[34] The mean increase in income between 1990 and 2000 was 6 percent with a standard deviation of 14 percent. As a result, a 20 percent increase is slightly bigger than a one standard deviation increase in income growth across the zip codes. The mean change in the poverty rate for our sample over this time period was 1 percentage point with a standard deviation of 4 percentage points. The mean change in the tenure rate of residents across the zip codes was 2 percent with a standard deviation of 13 percent.

where $\Delta P^{i,j}_{t,t+k}/P^{i,j}_t$, $Z^{i,j}_t$, $X^{i,j}_t$ and μ_j are defined as above. The variable $OwnIncShock^{i,j}_{t,t+k}$ 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 $NeighborIncShock^{i,j}_{t,t+k}$ 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 growth 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 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 483 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 16.8 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 do 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 4.55 with a standard error of 0.96 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 only the predicted income measure is 18.2.

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 34.5 percent (0.033 * 10.46).

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 * 8.36). 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 17.9 percent (0.021 * 8.52). 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:

$$Y_{t,t+k}^{i,j} = \delta_0 + \delta_1 OwnIncShock_{t,t+k}^{i,j} + \delta_4 NeighborIncShock_{t,t+k}^{i,j} + \Gamma X_t^{i,j} + \Psi Z_t^{i,j} + \mu_j + \epsilon_{t,t+k}^{i,j}$$

where all the variables are defined as above. The results of these regressions are shown in Table 7. Again our three measures of gentrification (one for each regression) 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 in 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 the predictions of our model are consistent with cross city differences in house price changes.

Recently, important work has been done by Gyourko et al. (2006) and Saiz (2010) exploring the causes of cross city differences in housing price appreciation. Both papers emphasize differences in supply constraints across cities to explain differences in house price appreciation 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 the ability to adjust the housing stock. For a given demand shock, cities where housing supply is more inelastic experience larger house price changes, all else equal. Gyrouko 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 poor neighborhoods will endogenously increase the desirability of living there. 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 A1 in the Appendix 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.[35] More elastic cities are more easily able to adjust supply and, as a result, should have prices that respond less to housing demand shocks. These restrictions left us with 152 MSAs.

In Panel A of Appendix Figure A1, 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 median income for the MSA from the 1990 and 2000 U.S. Censuses. In Panel B of the same figure, we show the same relationship but for those cities for which supply can adjust easily according to the Saiz measure (those cities in the bottom half of the distribution 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 even among cities with relatively elastic supply, we estimate:

$$\begin{aligned}\frac{\Delta P^j_{t,t+k}}{P^j_t} &= \delta_0 + \delta_1 \frac{\Delta Inc^j_{t,t+k}}{Inc^j_t} + \delta_2 SupplyConstraints^j_t + \delta_3 \frac{\Delta Inc^j_{t,t+k}}{Inc^j_t} * SupplyConstraints^j_t \\ &\quad + \delta_4 \frac{\Delta Commute^j_{t,t+k}}{Commute^j_t} + \epsilon^j_{t,t+k}\end{aligned}$$

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 MSAs for which the measure of housing supply elasticity is available.

As seen from column 1 of Table 8, changes in MSA income alone explain 49 percent of the variation in house price growth across MSAs between 1990 and 2000s. This is consistent with the results seen in Appendix Figure A1. The interesting results come in columns 2 and 3 where we include the interactions of the income growth with the measures of the supply elasticity. In column 2, we simply include the change in income measure with the continuous elasticity measure. In this specification, we also include the supply elasticity measure as a separate control. In column 3, 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

[35] See Saiz (2010).

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. In this specification, we also include separate dummies for whether the MSA was in the bottom third or the middle third of housing supply elasticity measures.

While it is true, that MSAs with more inelastic housing supply have housing prices that respond more to income shocks than MSAs with elastic housing 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 are exactly what is predicted by our model. Even in places where supply is relatively 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 2 and 3 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.[36] 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 4-6 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 alone only explained 30 percent of the variation in house prices across cities (column 4). As seen from columns 4 and 5 of Table 8, 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. Specifically, the elasticity of housing prices with respect to income changes was estimated to be 0.96. However, we also wish to note that 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. This seems to suggest that a large demand shock for housing that was orthogonal to changes in income also drove housing prices during the 2000-2006 period. This is consistent with many recent papers that suggest that low interest rates and the extension of credit to low income borrowers were important in explaining

[36] 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.

housing price dynamics in the recent period (see Mian and Sufi (2009)).

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).

Our model implies that poor neighborhoods that are close to richer neighborhoods have much higher housing price appreciation during periods of positive housing demand shocks. This is because these are the neighborhoods that gentrify in response to housing demand shocks. We conduct a variety of empirical tests that provide support for the mechanism at the heart of our model. In particular, we test our mechanism against the importance of production externalities coupled with transportation costs and against fixed amenity stories. We show that our results survive even after controlling for average commuting times and distance to fixed natural amenities. 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 housing 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 (2010) 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. We show 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

Alonso, William, *Location and Land Use*, Harvard University Press, 1964.

Autor, David H., Christopher J. Palmer, and Parag A. Pathak, "Housing Market Spillovers: Evidence from the End of Rent Control in Cambridge Massachusetts," 2010.

Bartik, T.J., "Who Benefits from State and Local Economic Development Policies?," *W.E. Upjohn Institute for Employment Research: Kalamazoo, Mich.*, 1991.

Bayer, Patrick, Hanming Fang, and Robert McMillan, "Separate when equal? Racial Inequality and Residential Segregation," 2009.

_ , Robert McMillan, and Fernando Ferreira, "A Unified Framework for Measuring Preferences for Schools and Neighborhoods," *The Journal of Political Economy*, 2007, *115* (4), 588–638.

Becker, Gary and Kevin Murphy, *Social Economics: Market Behavior in a Social Environment*, Harvard University Press, 2003.

Blanchard, O.J. and L.F. Katz, "Regional Evolutions," *Brookings Papers on Economic Activity*, 1992, pp. 1–75.

Brueckner, Jan K., "The Determinants of Residential Succession," *Journal of Urban Economics*, 1977, *4* (1), 45–59.

_ and Stuart S. Rosenthal, "Gentrification and Neighborhood Housing Cycles: Will America's Future Downtowns be Rich?," 2008.

Busso, Matias and Patrick Kline, "Do Local Economic Development Programs Work? Evidence from the Federal Empowerment Zone Program," Unpublished 2007.

Card, David, Alexandre Mas, and Jesse Rothstein, "Tipping and The Dynamics of Segregation," 2008, pp. 177–218.

Case, Karl E. and Christopher J. Mayer, "Housing Price Dynamics within a Metropolitan Area," *Regional Science and Urban Economics*, 1996, *26* (3-4), 387–407.

_ and Maryna Marynchenko, "Home Price Appreciation in Low-and Moderate-Income Markets," *Low-income Homeownership: Examining the Unexamined Goal*, 2002, pp. 239–256.

Courant, P.N., *Urban Residential Structure and Racial Prejudice*, Institute of Public Policy Studies, University of Michigan, 1974.

_ and J. Yinger, "On models of Racial Prejudice and Urban Residential Structure," *Journal of Urban Economics*, 1977, *4* (3), 272–291.

Davis, Morris A., "Discussion of Endogenous Gentrification and Housing Price Dynamics," `http://morris.marginalq.com/2010-02.md_discuss_efg.pdf` 2010.

_ and Francois Ortalo-Magné, "Household Expenditures, Wages, Rents," *Review of Economic Dynamics*, 2010.

_ , _ , and Peter Rupert, "What's Really Going on in Housing Markets?," Federal Reserve Bank of Cleveland, Economic Commentary 2007.

Ellen, I.G., M.H. Schill, S. Susin, and A.E. Schwartz, "Building Homes, Reviving Neighborhoods: Spillovers from Subsidized Construction of Owner-Occupied Housing in New York City," *Low Income Homeownership: Examining the Unexamined Goal*, 2002, pp. 447–77.

Glaeser, Edward L. and Joseph Gyourko, "The Impact of Building Restrictions on Housing Affordability," *Economic Policy Review*, 2003, *9* (2).

_ , Jed Kolko, and Albert Saiz, "Consumer city," *Journal of Economic Geography*, 2001, *1* (1), 27.

_ , Joseph Gyourko, and Charles Nathanson, "Housing Dynamics," Updated Version of NBER Working Paper 12787 2008.

_ , _ , and Raven E. Saks, "Why Have Housing Prices Gone Up?," *American Economic Review*, 2005, *95* (2), 329–333.

Gyourko, Joseph, Christopher J. Mayer, and Todd M. Sinai, "Superstar Cities," NBER Working Paper 12355 2006.

Jr., Robert E. Lucas and Esteban Rossi-Hansberg, "On the Internal Structure of Cities," *Econometrica*, 2002, pp. 1445–1476.

Kahn, Matthew E., Ryan Vaughn, and Jonathan Zasloff, "The Housing Market Effects of Discrete Land Use Regulations: Evidence from the California Coastal Boundary Zone," 2009.

Kain, J.F. and J.M. Quigley, *Housing Markets and Racial Discrimination: A Microeconomic Analysis*, National Bureau of Economic Research: distributed by Columbia University Press, 1975.

Kolko, Jed, "The Determinants of Gentrification," Unpublished 2007.

Mian, Atif and Amir Sufi, "The Consequences of Mortgage Credit Expansion: Evidence from the U.S. Mortgage Default Crisis," *Quarterly Journal of Economics*, 2009, *124* (4).

Mills, Edwin S., "An Aggregative Model of Resource Allocation in a Metropolitan Area," *The American Economic Review*, 1967, pp. 197–210.

Moretti, E., "Real Wage Inequality," NBER Working Paper No. 14370 2010.

Muth, Richard F., *Cities and Housing*, Univ. of Chicago Press, 1969.

Nieuwerburgh, S. Van and P.O. Weill, "Why Has House Price Dispersion Gone Up?," *The Review of Economic Studies*, 2010.

Notowidigdo, M.J., "The Incidence of Local Labor Demand Shocks," 2010.

Roback, Jennifer, "Wages, Rents, and the Quality of Life," *The Journal of Political Economy*, 1982, *90* (6), 1257.

Rose-Ackerman, S., "Racism and Urban Structure," *Journal of Urban Economics*, 1975, *2* (1), 85–103.

Rosen, Sherwin, "Wage-based Indexes of Urban Quality of Life," *Current Issues in Urban Economics*, 1979, *3.*

Rosenthal, Stuart S., "Old homes, Externalities, and Poor Neighborhoods. A Model of Urban Decline and Renewal," *Journal of Urban Economics*, 2008, *63* (3), 816–840.

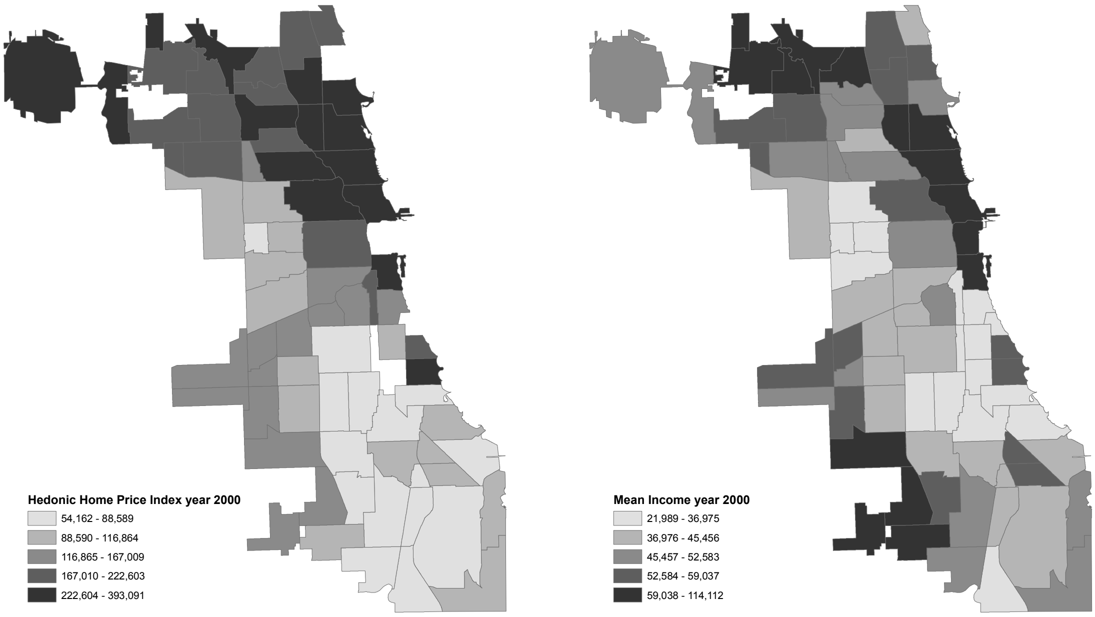
Rossi-Hansberg, Esteban, Pierre-Daniel Sartre, and Raymond Owens III, "Housing Externalities," *Journal of Political Economy*, 2010, *118* (3), 485–535.

Saiz, A., "The Geographic Determinants of Housing Supply," *Quarterly Journal of Economics*, 2010.

Schelling, T.C., "Models of Segregation," *The American Economic Review*, 1969, *59* (2), 488–493.

Voith, R. and J. Gyourko, "Capitalization of Federal Taxes, the Relative Price of Housing, and Urban Form: Density and Sorting Effects," *Regional Science and Urban Economics*, 2002, *32* (6), 673–690.

Voith, Richard, "Does the Tax Treatment of Housing Create an Incentive for Exclusionary Zoning and Increased Decentralization?," Federal Reserve Bank of Philadelphia Working Paper No. 99-22 1999.



A: Hedonic Home Price Index. B: Mean Income.

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

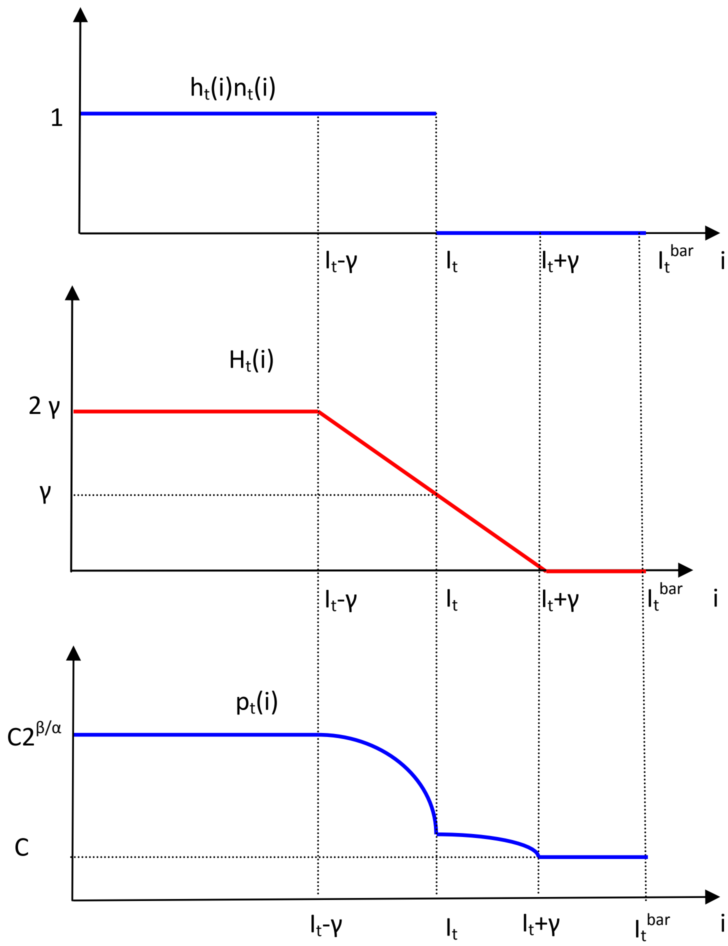


Figure 2: Externality across space.

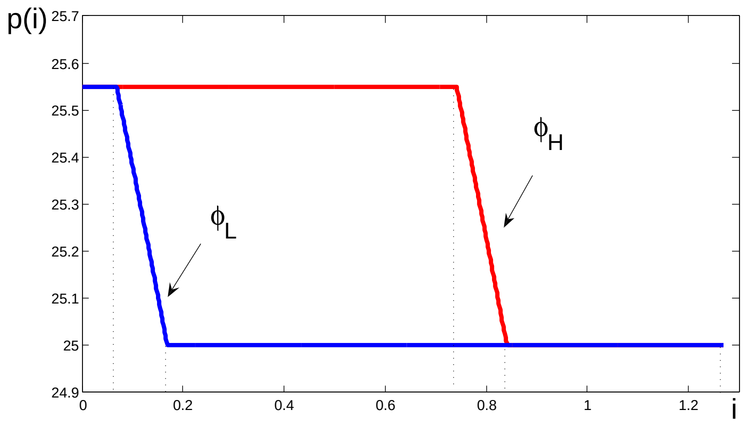


Figure 3: Reaction of house prices across space to a positive demand shock. We set $\alpha = .8$, $\beta = .8$, $\delta_R = 0.2$, $\delta_P = 0$, $A = 1$, $\phi = 1$, $\gamma = .1$, $N^R = N^P = .5$, $C^P = C^R = 25$, and $r = .03$.

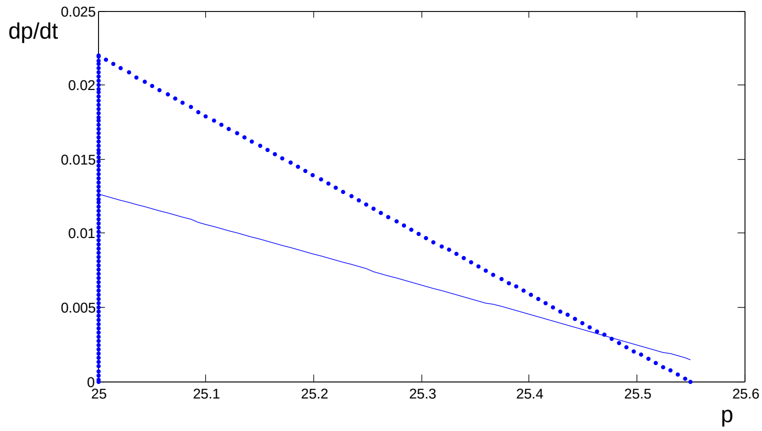


Figure 4: Price growth rate across locations as a function of the initial price level and OLS regression. We set $\alpha = .8$, $\beta = .8$, $\delta_R = 0.2$, $\delta_P = 0$, $A = 1$, $\phi = 1$, $\gamma = .1$, $N^R = N^P = .5$, $C^P = C^R = 25$, and $r = .03$.

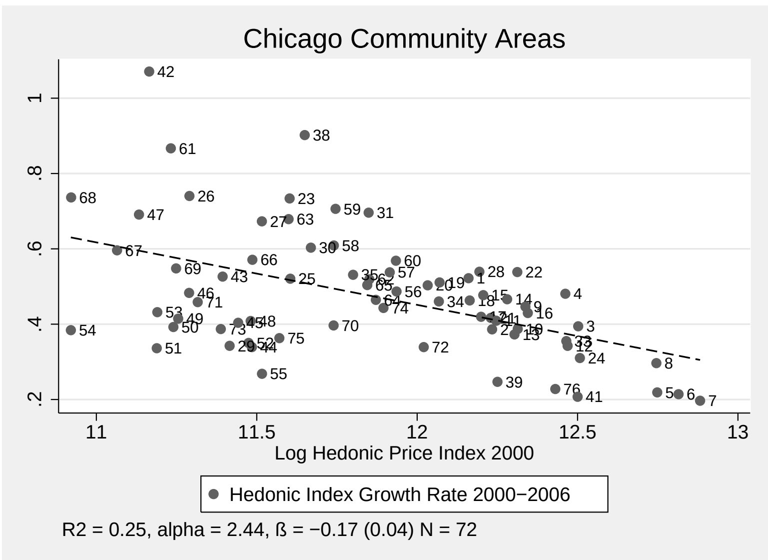


Figure 5: Price Growth vs. Initial Price 2000-2006 for Chicago City Community Areas using Hedonic Index.

Table 1: The Relationship Between Home Price and Income in Chicago (n = 72)

	(1)	(2)	(3)	(4)
Log Mean Income		1.21	0.72	0.65
		(0.11)	(0.12)	(0.11)
Log Mean Commuting Time	-3.45		-2.35	-1.47
	(0.27)		(0.30)	(0.33)
Log Distance to Loop	0.16		0.03	-0.03
	(0.09)		(0.06)	(0.06)
Log Distance to Lake	-0.03		-0.02	-0.07
	(0.06)		(0.05)	(0.06)
Fraction African American				-0.38
				(0.12)
R^2	0.63	0.50	0.74	0.78

Notes: This table reports the results from OLS regressions of log house prices across Chicago community areas 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. The commuting time, income and race data come from the 2000 U.S. Census. Robust standard errors shown in parentheses. See text for additional details.

Table 2: Housing Price Growth Variation

	(1) FHFA b/w	(2) C-S b/w	(3) C-S MSA w/i	(4) C-S city w/i	(5) Zillow city w/i
2000 - 2006	0.36	0.42	0.18	0.18	0.28
1990 - 2000	0.16	0.21	0.16	0.17	

Notes: Columns 1 and 2 of this table show 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 row 1, n = 153; column 1 row 2, n = 152) and using the Case Shiller housing price index (column 2 row 1, n = 20; column 2 row 1, n = 17). In columns 3-5, this table shows the cross zip code/within city or MSA standard deviations in house price growth during the same periods. Columns 3 and 4 use the Case Shiller house price indices within Case Shiller MSAs (column 3 row 1, n = 1,693; column 3 row 2, n = 1,529) and within Case Shiller main cities (column 4 row 1, n = 497; column 4 row 2, n = 508). Column 5 uses the Zillow house price indices within Case Shiller main cities (n = 596).

Table 3: Housing Price Growth by Initial Price Quartile

	(1) Quartile 4	(2) Quartile 3	(3) Quartile 2	(4) Quartile 1	(5) p-val of Quartile 4 = Quartile 1
2000 - 2006: Housing Booms					
Chicago, City Level, Case-Shiller	0.53	0.66	0.72	0.88	0.00
Chicago, MSA Level, Case-Shiller	0.47	0.50	0.49	0.69	0.00
New York City, MSA Level, Case-Shiller	0.64	0.76	0.86	1.11	0.00
Boston, MSA Level, Case-Shiller	0.40	0.47	0.54	0.62	0.00
Los Angeles, MSA Level, Case-Shiller	1.21	1.40	1.58	1.76	0.00
San Francisco, MSA Level, Case-Shiller	0.35	0.41	0.49	0.61	0.00
Washington D.C., MSA Level, Case-Shiller	1.29	1.37	1.49	1.61	0.00
1990 - 1997: Housing Booms					
Denver, MSA Level, Case-Shiller	0.51	0.50	0.52	0.89	0.00
Portland, MSA Level, Case-Shiller	0.41	0.52	0.49	0.69	0.00
1984 - 1989: Housing Booms					
New York City, City Level, Furman	0.33	0.57	0.69	1.06	0.00
Boston, MSA Level, Case-Shiller	0.65	0.69	0.75	0.84	0.00
1990 - 1997: Housing Busts					
San Francisco, MSA Level, Case-Shiller	-0.08	-0.07	-0.11	-0.14	0.00
Boston, MSA Level, Case-Shiller	-0.05	-0.08	-0.13	-0.11	0.00

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.

Table 4: The Relationship Between House Price Growth Among Low Price Neighborhoods and Distance to High Price Neighborhoods

	(1)	(2)	(3)	(4)	(5)
	2000-2006 (n = 225)		1990-2000 (n = 211)		
Log Dist. to Nearest High-Price Neighborhood_t	-0.057	-0.053	-0.062	-0.044	-0.058
	(0.019)	(0.019)	(0.023)	(0.022)	(0.029)
Log Dist. to Nearest High-Price $\text{Neighborhood}_t * \text{Bust}^j_{t,t+k}$					0.041
					(0.036)
Log Mean Commuting Time_t		-0.114		-0.139	-0.141
		(0.147)		(0.120)	(0.121)
Percent Change in Commuting $\text{Time}_{t,t+k}$				-0.551	-0.549
				(0.178)	(0.177)
Distance to Rivers Lakes, Oceans	No	Yes	No	Yes	Yes
Additional Controls Included	Yes	Yes	Yes	Yes	Yes

Notes: This table shows the results from a regression of zip code house price growth between period t and t+k (using the Case Shiller house price indices) on the log distance from that zip code to a high housing price zip code, city fixed effects, and a vector of zip code controls. The sample for these regressions includes all Case-Shiller zip codes within the main city of the Case Shiller MSAs. 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. The vector of controls include in the regression are: the median house price within the zip code in period t, the median income of residents in the zip code in period t, the former two variables interacted with the city wide house price increase between t and t+k, and the fraction of residents that were African American in the zip code in period t. The mean distance to high price neighborhoods in both 1990 and 2000 was 3.8 miles and 3.9 miles, respectively. The mean commuting time in 1990 and 2000, respectively, was 27.6 minutes and 28.5 minutes. The mean growth in commuting times between 1990 and 2000 was roughly 10 percent. Robust standard errors shown in parentheses.

Table 5: Relationship Between House Price Growth and Measures of Gentrification

	(1)	(2)	(3)
Percent Change in Neighborhood Median Income	0.32 (0.09)		
Percentage Point Change in Neighborhood Poverty Rate		-1.12 (0.22)	
Percent Change in Mean Tenure in Home			-0.15 (0.07)
Additional Controls Included	Yes	Yes	Yes

Notes: This table shows the coefficient on measures of gentrification in three separate regressions of Case Shiller zip code level housing price growth between 1990 and 2000 on the separate measures of zip code gentrification between 1990 and 2000, city fixed effects, and a vector of additional zip code level controls. Each column represents a separate regression. The sample for these regressions includes all Case-Shiller zip codes within the main city of the Case Shiller MSAs. Our vector of additional zip code level controls include: the log of median income within the zip code in 1990, the log of the median house price within the zip code in 1990, the fraction of residents that were African American in 1990, the interaction of the FHFA MSA-level home price index growth rate with the log of median income within the zip code in 1990, the interaction of the FHFA MSA-level home price index growth rate with the log of the median house price within the zip code in 1990, and city level fixed effects. For columns 1 and 2 (n = 504). For column 3 (n = 454). Lower sample size for column 3 is due to missing values in 2000 Census data for year moved into home variable. See the text for additional details. Robust standard errors shown in parentheses.

Table 6: House Price and Commuting Time Response to Industry Earnings Shocks of Own and Adjacent Neighborhoods (n = 483)

	(1)	(2)	(3)	(4)
Dependent Variable:	Neighborhood House Price Growth			
Predicted Income Shock	10.46 (1.53)	7.21 (1.86)	7.43 (1.91)	6.75 (1.70)
Avg. Neighborhood Shock: 1-10		8.36 (1.99)	8.70 (1.99)	8.52 (1.88)
Avg. Neighborhood Shock: 11-20			-1.33 (1.98)	-0.48 (1.85)
Avg. Neighborhood Shock: 21-30			-2.41 (1.83)	-1.21 (1.70)
Log Commuting Time				-0.46 (0.08)
Percentage Change in Commuting Time				-0.58 (0.13)
Additional Controls Included	Yes	Yes	Yes	Yes
Distance to Rivers, Lakes, Oceans	No	No	No	Yes

Notes: This table reports the coefficients of different variables from a regression of house price changes between 1990 and 2000 (measured by growth in the Case-Shiller index) in zip code i of city j on the predicted income change in zip code i of city j during the same period as well as the average predicted income shock of zip code of various degrees of proximity to neighborhood i of city j. The regression also includes controls for the log of the zip code level house price in 1990, the log commuting time of residents in the zip code in 1990, the change in commuting time of residents in the zip code between the 1990 and 2000, the fraction African American residents in the zip code in 1990, the log distance to rivers, lakes, and oceans (if applicable), and city fixed effects. The sample for these regressions includes all Case-Shiller zip codes within the main city of the Case Shiller MSAs. See text for additional details.

Table 7: Relationship Between Measures of Gentrification and Industry Earnings Shocks of Own and Adjacent Neighborhoods

	(1)	(2)	(3)
Dependent Variable	Percentage Change in in Median Income	Percentage Point Change in Poverty Rate	Percentage Change in Tenure
Predicted Income Shock	3.61	-0.41	-2.66
	(1.30)	(0.39)	(1.28)
Avg. Neighborhood Shock: 1-10	3.54	-1.07	-4.65
	(1.79)	(0.50)	(1.60)
Avg. Neighborhood Shock: 11-20	-1.18	0.41	-1.54
	(1.40)	(0.41)	(1.76)
Avg. Neighborhood Shock: 21-30	-0.71	0.36	0.55
	(1.39)	(0.40)	(1.72)
Additional Controls Included	Yes	Yes	Yes

Notes: This table reports the coefficients from different variables from three different regressions of measures of zip code gentrification between 1990 and 2000 on measures of predicted neighborhood income shocks between 1990 and 2000 and the average predicted income shock of adjacent zip codes between 1990 and 2000. Each column reports the results of a regression with a different measure of zip code gentrification. The regression also includes controls for growth in zip code commuting time between 1990 and 2000, initial house prices in the zip code, the initial fraction of African Americans in the zip code, distance to rivers, lakes and oceans (if applicable), and city fixed effects. The sample for these regressions includes all Case-Shiller zip codes within the main city of the Case Shiller MSAs. For columns 1 and 2 (n = 482). For column 3 (n = 432). Lower sample size for column 3 is due to missing values in 2000 Census data for year moved into home variable. See text for additional details.

Table 8: Income Growth, Supply Constraints, and Cross City Differences in House Price Appreciation

	(1)	(2)	(3)	(4)	(5)	(6)
	1990 - 2000 (n = 152)			2000 - 2006 (n = 155)		
Growth in Inc	1.82	2.84	0.80	3.58	4.12	0.96
	(0.17)	(0.37)	(0.37)	(0.38)	(0.62)	(0.48)
Saiz Housing Supply Elasticity * Growth in Inc		-0.53			-0.76	
		(0.16)			(0.26)	
1(Low Saiz Supply Elasticity) * Growth in Inc			1.41			2.13
			(0.45)			(0.71)
1(Med Saiz Supply Elasticity) * Growth in Inc			1.27			2.60
			(0.45)			(0.96)
Saiz Housing Supply Elasticity		0.04			-0.11	
		(0.02)			(0.02)	
1(Low Saiz Supply Elasticity)			-0.12			0.25
			(0.05)			(0.07)
1(Med Saiz Supply Elasticity)			-0.15			0.11
			(0.05)			(0.06)
Growth in Med Commute		-0.00	-0.00			
		(0.00)	(0.00)			
R^2	0.49	0.53	0.56	0.30	0.51	0.50

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. Robust standard errors shown in parentheses.

A1 Proof Appendix (Included on the Journal's Website)

A1.1 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

$$\begin{aligned} U^R(i) &\leq \bar{U}^R \text{ for all } i \in \left[I_t, \bar{I}_t\right], \\ U^P(i) &\leq \bar{U}^P \text{ for all } i \in [0, I_t], \end{aligned}$$

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

$$K_R(A + H_t(i))^{\frac{\delta_R}{\beta}} \leq K_P(A + H_t(i))^{\frac{\delta_P}{\beta}} + \frac{r}{1+r}\left(C^R - C^P\right) \text{ for all } i \in \left[I_t, \bar{I}_t\right], \tag{18}$$

$$K_P(A + H_t(i))^{\frac{\delta_P}{\beta}} \leq K_R(A + H_t(i))^{\frac{\delta_R}{\beta}} - \frac{r}{1+r}\left(C^R - C^P\right) \text{ for all } i \in [0, I_t]. \tag{19}$$

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

$$\begin{aligned} K^R &= \frac{r}{1+r}\left[C^P\left(\frac{A}{A+\gamma}\right)^{-\frac{\delta_P}{\beta}} + \left(C^R - C^P\right)\right](A+\gamma)^{-\frac{\delta_R}{\beta}} \\ K^P &= \frac{r}{1+r}C^P A^{-\frac{\delta_P}{\beta}}, \end{aligned}$$

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

$$\left(\frac{A + H_t(i)}{A+\gamma}\right)^{\frac{\delta_R - \delta_P}{\beta}} \leq \frac{1 + \left(\frac{C^R - C^P}{C^P}\right)\left(\frac{A}{A+H_t(i)}\right)^{\frac{\delta_P}{\beta}}}{1 + \left(\frac{C^R - C^P}{C^P}\right)\left(\frac{A}{A+\gamma}\right)^{\frac{\delta_P}{\beta}}}.$$

for all $i \in \left[I_t, \bar{I}_t\right]$. 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

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

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 $\delta^P \leq \delta^R$.

A1.2 Proof of Proposition 2

The initial price schedule is:

$$p_t(i) = \begin{cases} \left[C^P\left(1 + \frac{\gamma}{A}\right)^{\frac{\delta_P}{\beta}} + C^R - C^P\right]\left(1 + \frac{\min\{\gamma, I_t - i\}}{A+\gamma}\right)^{\frac{\delta_R}{\beta}} & \text{for } i \in [0, I_t] \\ C^P\left(1 + \frac{\max\{\gamma + I_t - i, 0\}}{A}\right)^{\frac{\delta_P}{\beta}} & \text{for } i \in \left[I_t, \bar{I}_t\right] \end{cases}. \tag{20}$$

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

$$\bar{p} \equiv \left[C^P\left(1 + \frac{\gamma}{A}\right)^{\frac{\delta_P}{\beta}} + C^R - C^P\right]\left(1 + \frac{\gamma}{A+\gamma}\right)^{\frac{\delta_R}{\beta}}.$$

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 ϕ. Denote with a primus the equilibrium objects after the shock. We have

$$\frac{p_{t+1}(i)}{p_t(i)} = \begin{cases} \left(\frac{A+\gamma+\min\{\gamma,I_{t+1}-i\}}{A+\gamma+\min\{\gamma,I_t-i\}}\right)^{\frac{\delta_R}{\beta}} & \text{for } i \in [0, I_t] \\ \left[\left(\frac{A+\gamma}{A}\right)^{\frac{\delta_P}{\beta}} + \frac{C^R-C^P}{C^P}\right] \frac{\left(1+\frac{\min\{\gamma,I_{t+1}-i\}}{A+\gamma}\right)^{\frac{\delta_R}{\beta}}}{\left(1+\frac{\max\{\gamma+I_t-i,0\}}{A}\right)^{\frac{\delta_P}{\beta}}} & \text{for } i \in [I_t, I_{t+1}] \\ \left(\frac{A+\max\{\gamma+I_{t+1}-i,0\}}{A+\max\{\gamma+I_t-i,0\}}\right)^{\frac{\delta_P}{\beta}} & \text{for } i \in [I_{t+1}, \bar{I}_t] \end{cases}. \tag{21}$$

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

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

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

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

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

$$\frac{d\left(\frac{p_{t+1}(i)}{p_t(i)}\right)}{di} = \frac{\delta_R}{\beta(A+\gamma)}\left(\frac{A+2\gamma}{A+\gamma}\right)^{\frac{\delta_R}{\beta}}\left(\frac{A+\gamma+I_t-i}{A+\gamma}\right)^{-\frac{\delta_R}{\beta}-1} > 0 \tag{22}$$

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

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

given that $I_t < I_{t+1}$.

For $i \in [I_t, I_t + \gamma]$ we have that

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

$$\frac{d\left(\frac{p_{t+1}(i)}{p_t(i)}\right)}{di} = \frac{\tilde{C}}{A}\left(\frac{A+2\gamma}{A+\gamma}\right)^{\frac{\delta_R}{\beta}}\left(\frac{A+\gamma+I_t-i}{A}\right)^{-\frac{\delta_P}{\beta}-1} > 0 \tag{24}$$

where

$$\tilde{C} \equiv \left[\left(\frac{A+\gamma}{A}\right)^{\frac{\delta_P}{\beta}} + \frac{C^R-C^P}{C^P}\right]$$

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

$$\frac{d\left(\frac{p_{t+1}(i)}{p_t(i)}\right)}{di} = \frac{\tilde{C}}{\beta}\frac{\left(\frac{A+\gamma+I_{t+1}-i}{A+\gamma}\right)^{\frac{\delta_R}{\beta}}}{\left(\frac{A+\gamma+I_t-i}{A}\right)^{\frac{\delta_P}{\beta}}}\left[\frac{\delta_P}{A+\gamma+I_t-i} - \frac{\delta_R}{A+\gamma+I_{t+1}-i}\right] \tag{25}$$

hence

$$\frac{d\left(\frac{p_{t+1}(i)}{p_t(i)}\right)}{di} > 0 \text{ iff } \frac{\delta_R}{\delta_P} < \frac{A+\gamma+I_{t+1}-i}{A+\gamma+I_t-i},$$

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

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

$$\frac{d\left(\frac{p_{t+1}(i)}{p_t(i)}\right)}{di} = \frac{\delta_P}{\beta}\left(\frac{A+\gamma+I_{t+1}-i}{A+\gamma+I_t-i}\right)^{\frac{\delta_P}{\alpha}}\left[\frac{1}{A+\gamma+I_t-i} - \frac{1}{A+\gamma+I_{t+1}-i}\right] > 0. \tag{26}$$

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 $i \leq I_t + \gamma$, 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.

A1.3 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 ϕy^s for $s = R, P$, I_{t+1} is non-decreasing in ϕ and hence prices are weakly increasing in ϕ 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 $i \in [0, I^B_{t+1}-\gamma]$ 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 $i \in [I^B_{t+1}, I^B_{t+1}+\gamma]$ may be in $[I^A_{t+1}-\gamma, I^A_{t+1}]$. 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 ϕ. Hence, if at time $t+1$ the economy is hit by an unexpected and permanent increase in ϕ, then P_{t+1} is going to be higher, the larger is the increase in ϕ. 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 ϕ.

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 ϕ 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 ϕ, whenever i is in the same type of interval after a small or a large shock, say ϕ^A or ϕ^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 $\max\{I^H, I^L_B-\gamma\} < i < I^H+\gamma$ and $I^H < i < \min\{I^H+\gamma, I^L_A-\gamma\}$, 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

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

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 $d^2\left[p_{t+1}(i)/p_t(i)\right]/did\phi$ 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.

A1.4 Proof of Proposition 5

Consider two cities, A and B, with both $y_R^A>y_R^B$ and $y_P^A>y_P^B$. 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 ϕ, 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 $I_t^A>I_t^B$, $I_{t+1}^A>I_{t+1}^B$, $\bar{I}_t^A>\bar{I}_t^B$ and $\bar{I}_{t+1}^A>\bar{I}_{t+1}^B$. Hence city A has a larger center and a larger size overall. After the income increase, prices do not change for all $i<I_t^j-\gamma$ and for all $i>I_{t+1}^j+\gamma$ if $I_{t+1}^j+\gamma<\bar{I}_t^j$, for both $j=A$ and $j=B$. Moreover, given that $I_{t+1}^A>I_{t+1}^B$, the growth rate $p_{t+1}^A(i)/p_t^A(i)$ is weakly higher than $p_{t+1}^B\left(i+I_t^B-I_t^A\right)/p_t^B\left(i+I_t^B-I_t^A\right)$ for all $i\in\left[I_t^A-\gamma,\bar{I}_t^A\right]$. 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 $i\in\left[I_t^A-\gamma,I_t^A+\gamma\right]$ are exactly the same as in city B for $i\in\left[I_t^B-\gamma,I_t^B+\gamma\right]$. However, $I_t^A>I_t^B$, so that the interval $\left[0,I_t^A\right]$ is larger than $\left[0,I_t^B\right]$. 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 $i\in\left[I_t^A-\gamma,I_t^A+\gamma\right]$ is equivalent to $d(p_{t+1}(i)/p_t(i))/di$ in city B for all $i\in\left[I_t^B-\gamma,I_t^B+\gamma\right]$ if city B was facing a larger income increase. From the proof of Proposition 4, this immediately implies that $d(p_{t+1}^A(i)/p_t^A(i))/di>d(p_{t+1}^B\left(i+I_t^A-I_t^B\right)/p_t^B\left(i+I_t^A-I_t^B\right))/di$, that is, $g_t^{A\prime}(p)>g_t^{B\prime}(p)$ for all $p\in\left(\bar{p},C^P\right)$. 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 $g_t^{A\prime}(\bar{p})=g_t^{B\prime}(\bar{p})$. This completes the proof that $g_t^{A\prime}(p)\geq g_t^{B\prime}(p)$ for all $p>C^P$ whenever this derivative is well defined.

A2 Data Appendix (Included on the Journal's Website)

A2.1 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.[37] 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.[38]) 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.[39] 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

[37] Web Address: `http://chicagotribune.public-record.com/realestate`.

[38] All prices are left in nominal terms until hedonic index is computed. All prices are then converted to year 2000 dollars.

[39] Web Address: `http://cookcountyassessor.com/`.

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.

A2.2 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.[40]

A2.3 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.[41] 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.[42] Finally, variables indicating census tracts that had not changed boundaries between 1980 and 1990 were obtained from the Neighborhood Change Database produced by Geolytics.[43] Median home value data for NYC community districts and are available from NYC's GIS department at `http://gis.nyc.gov`.

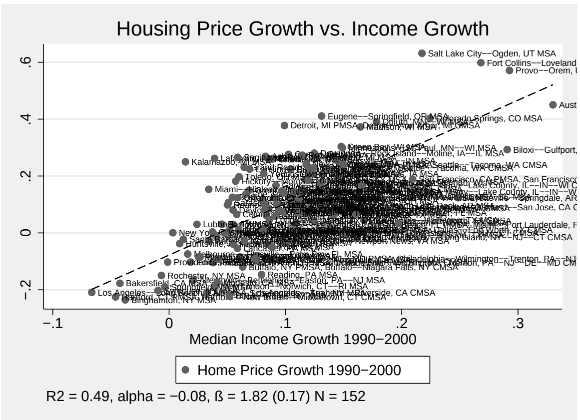
A2.4 Chicago Community Area Mean Commuting Time

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.

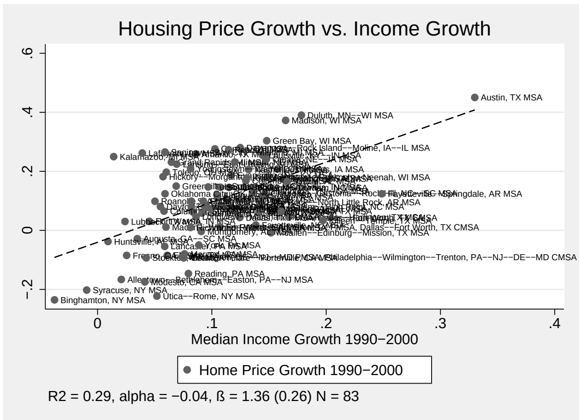
A2.5 MSA-Level Housing Price Growth

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).[44]

[40] See `http://www.furmancenter.org` for more information.
[41] Web Address: `http://factfinder.census.gov/`.
[42] Web Address: `http://www.icpsr.umich.edu/icpsrweb/ICPSR/`.
[43] Web Address: `http://www.geolytics.com/USCensus,Neighborhood-Change-Database-1970-2000,Products.asp`.
[44] Web Address: `http://www.fhfa.gov/Default.aspx?Page=216`.



A: All MSAs over 200K in Population



B: Top Half of Saiz Elasticity Measure for MSAs over 200K in Population

Figure A1: Housing Price Growth vs. Income Growth 1990 - 2000.