Online Appendix to: Endogenous Gentrification and Housing Price Dynamics

Veronica Guerrieri*
University of Chicago and NBER

Daniel Hartley†
Federal Reserve Bank of Cleveland

Erik Hurst‡
University of Chicago and NBER

November 1, 2012

R1 Online Robustness Appendix

This is the online robustness appendix for our paper titled "Endogenous Gentrification and Housing Price Dynamics". This appendix contains additional tables, sample description, and proofs which we refer to in our paper. All data and code for this paper can be found at `http://faculty.chicagobooth.edu/erik.hurst/research/`. As noted in the main paper, we are contractually prevented from posting the Case-Shiller data. However, we posted all other data online. Additionally, we have posted all the relevant code used to make the tables and figures reported both in the main paper and in this robustness appendix.

This appendix is broken into two sections. In the first section, we provide proofs for Propositions 1 - 4 in the paper. In the second section, we provide several further empirical results and sample selection details that bolster our analysis presented in the paper. In the first two subsections, we present greater (and more systematic) detail on within-city house price movements for many cities during different time periods (in support of Facts 2 and 3 in the paper). In the third section, we provide details of how we selected census tracts for our broadest census tract samples. In the final section, we show the results of Table 3 controlling for the age distribution of the neighborhood housing stock.

*Contact: *Veronica.Guerrieri@chicagobooth.edu*
†Contact: *Daniel.Hartley@clev.frb.org*
‡Contact: *Erik.Hurst@chicagobooth.edu*

R2 Theory Appendix

R2.1 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 (7). Using expression (8), these two conditions can be rewritten as

$$K^R(A+H_t(i))^{\frac{\delta_R}{\beta}} \leq K^P(A+H_t(i))^{\frac{\delta_P}{\beta}} + \frac{r}{1+r}\left(C^R - C^P\right) \text{ for all } i \in \left[I_t, \bar{I}_t\right], \quad \text{(R1)}$$

$$K^P(A+H_t(i))^{\frac{\delta_P}{\beta}} \leq K^R(A+H_t(i))^{\frac{\delta_R}{\beta}} - \frac{r}{1+r}\left(C^R - C^P\right) \text{ for all } i \in [0, I_t]. \quad \text{(R2)}$$

Combining (10) with (14) and (15) 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 (R1) 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 (R2) 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. It follows that $\delta_R \geq \delta_P$ 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 $\delta^P \leq \delta^R$.

R2.2 Proof of Proposition 2

The initial price schedule is:

$$p_t(i) = \begin{cases} p_t^R(i) & \text{for } i \in [0, I_t] \\ p_t^P(i) & \text{for } i \in \left[I_t, \bar{I}_t\right] \end{cases}, \tag{R3}$$

where $p_t^R(i)$ and $p_t^P(i)$ are given by (18) and (19). 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 the economy is hit by an unexpected and permanent increase in population, so that the measure of both rich and poor households increase by a proportion of $\phi > 1$, *i.e.* $N_{t+1}^s = \phi N_t^s$ for $s = R, P$. We then have

$$\frac{p_{t+1}(i)}{p_t(i)} = \begin{cases} \left(\frac{A+\gamma+\min\{\gamma, I_{t+1}-i\}}{A+\gamma+\min\{\gamma, I_t-i\}}\right)^{\frac{\delta_R}{\beta}} & \text{for } i \in [0, I_t] \\ \left[\left(\frac{A+\gamma}{A}\right)^{\frac{\delta_P}{\beta}} + \frac{C^R-C^P}{C^P}\right] \frac{\left(1+\frac{\min\{\gamma, I_{t+1}-i\}}{A+\gamma}\right)^{\frac{\delta_R}{\beta}}}{\left(1+\frac{\max\{\gamma+I_t-i,0\}}{A}\right)^{\frac{\delta_P}{\beta}}} & \text{for } i \in [I_t, I_{t+1}] \\ \left(\frac{A+\max\{\gamma+I_{t+1}-i,0\}}{A+\max\{\gamma+I_t-i,0\}}\right)^{\frac{\delta_P}{\beta}} & \text{for } i \in [I_{t+1}, \bar{I}_t] \end{cases}. \tag{R4}$$

Also, from equations (16) and (17), 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 (R4), 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

$$\frac{d\left(\frac{p_{t+1}(i)}{p_t(i)}\right)}{di} = \frac{\delta_R}{\beta\left(A+\gamma\right)}\left(\frac{A+2\gamma}{A+\gamma}\right)^{\frac{\delta_R}{\beta}}\left(\frac{A+\gamma+I_t-i}{A+\gamma}\right)^{-\frac{\delta_R}{\beta}-1} > 0 \tag{R5}$$

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

$$\frac{d\left(\frac{p_{t+1}(i)}{p_t(i)}\right)}{di} = \frac{\delta_R}{\beta}\left(\frac{A+\gamma+I_{t+1}-i}{A+\gamma+I_t-i}\right)^{\frac{\delta_R}{\beta}}\left[\frac{1}{A+\gamma+I_t-i} - \frac{1}{A+\gamma+I_{t+1}-i}\right] > 0 \tag{R6}$$

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

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

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

$$\frac{d\left(\frac{p_{t+1}(i)}{p_t(i)}\right)}{di} = \frac{\tilde{C}}{A}\left(\frac{A+2\gamma}{A+\gamma}\right)^{\frac{\delta_R}{\beta}}\left(\frac{A+\gamma+I_t-i}{A}\right)^{-\frac{\delta_P}{\beta}-1} > 0 \tag{R7}$$

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

$$\frac{d\left(\frac{p_{t+1}(i)}{p_t(i)}\right)}{di} = \frac{\tilde{C}}{\beta}\frac{\left(\frac{A+\gamma+I_{t+1}-i}{A+\gamma}\right)^{\frac{\delta_R}{\beta}}}{\left(\frac{A+\gamma+I_t-i}{A}\right)^{\frac{\delta_P}{\beta}}}\left[\frac{\delta_P}{A+\gamma+I_t-i} - \frac{\delta_R}{A+\gamma+I_{t+1}-i}\right] \tag{R8}$$

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

$$\frac{d\left(\frac{p_{t+1}(i)}{p_t(i)}\right)}{di} = \frac{\delta_P}{\beta}\left(\frac{A+\gamma+I_{t+1}-i}{A+\gamma+I_t-i}\right)^{\frac{\delta_P}{\alpha}}\left[\frac{1}{A+\gamma+I_t-i} - \frac{1}{A+\gamma+I_{t+1}-i}\right] > 0. \tag{R9}$$

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

R2.3 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 population, that is, $N^s_{t+1} = \phi N^s_t$ with $\phi > 1$ for $s = P, R$. From expressions (16) and (17) it follows that $I_{t+1} > I_t$ and $\bar{I}_{t+1} > \bar{I}_t$. Then, from expression (R4), 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.

R2.4 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 $\left[I_t, \bar{I}_t\right]$. From expression (R3), 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 (16) where N^s_t is substituted by ϕN^s_t 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 $i \in \left[0, I^B_{t+1} - \gamma\right]$ are also in $\left[0, I^B_{t+1} - \gamma\right]$, but some $i \in \left[I^B_{t+1} - \gamma, I^B_{t+1} + \gamma\right]$ may be in $\left[0, I^A_{t+1} - \gamma\right]$ or some $i \in \left[I^B_{t+1}, I^B_{t+1} + \gamma\right]$ may be in $\left[I^A_{t+1} - \gamma, I^A_{t+1}\right]$. Given that, from inspection of expression (R3), $p_{t+1}(i)$ 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

$$\frac{d^2\left(p_{t+1}(i)/p_t(i)\right)}{dp_t(i)\, d\phi} \geq 0$$

for all $p_t(i) > C^P$ where the derivative is well-defined. Equations (R5)-(R9) in the proof of Proposition (2) define $d\left[p_{t+1}(i)/p_t(i)\right]/di$ for all i where this derivative is well-defined and $p_t(i) > C^P$. If the increase in ϕ is big enough, $d\left[p_{t+1}(i)/p_t(i)\right]/di > 0$ for all $p_t(i) > C^P$. Moreover, by inspection, it is easy to see that $d\left[p_{t+1}(i)/p_t(i)\right]/di$ 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 $\min\left\{I^B_{t+1} - \gamma, I^A_t\right\} < i < I^A_t$ but $I^B_t - \gamma < i < \min\left\{I^A_t - \gamma, I^B_t\right\}$, or that $\max\left\{I^H, I^L_B - \gamma\right\} < i < I^H + \gamma$ and $I^H < i < \min\left\{I^H + \gamma, I^L_A - \gamma\right\}$, or that $I^L_B < i < I^H + \gamma$ but $\max\left\{I^H, I^L_A - \gamma\right\} < i < I^H + \gamma$. It is easy to see that expression (R5) is not smaller than expression (R6) and that expression (R7) is not smaller than expression (R8). Finally expression (R8) is bigger than expression (R9) iff

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

which is true if the shock is large enough so that $I_{t+1} - I_t$ is big enough, as we assumed. This proves that $d^2 \left[p_{t+1}(i) / p_t(i) \right] / did\phi$ 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.

R3 Empirical Appendix

R3.1 Further Evidence for Fact 2: Initially Low Price Neighborhoods Within a City Appreciate More than High Price Neighborhoods During City-Wide Housing Booms

As mentioned in the paper, we estimate the following simple relationship using different housing price series:

$$\frac{\Delta P^{i,j}_{t,t+k}}{P^{i,j}_t} = \mu_j + \omega_1 \ln(HP^{i,j}_t) + \epsilon^{i,j}_{t,t+k} \tag{R10}$$

where $\Delta P^{i,j}_{t,t+k}/P^{i,j}_t$ is the growth in housing prices between period t and $t+k$ within neighborhood i in city or MSA j using the various house price series and $HP^{i,j}_t$ is the median house price in neighborhood i in city or MSA j in year t as measured by the U.S. Census. Given that we also include city or MSA fixed effects, μ_j, all of our identification comes from variation across neighborhoods within a city/MSA. The variable of interest from this regression is ω_1 which estimates the relationship between initial median house prices in the neighborhood and subsequent neighborhood housing price growth. We run this regression using different neighborhood house price series and for different time periods. For all specifications, we weight the data using the number of owner occupied housing units in the neighborhood during period t (from the Census).

The results from these regressions are shown in Table R1. In columns (1) - (3), we show results where t and $t+k$ are 2000 and 2006, respectively, using the Case Shiller data. For all specifications except those in the last two columns, our definition of neighborhood is a zip code within the MSA. The specification in column (2) is he same as column (1) except that we instrument for the initial level of house prices (2000) using the lagged level of house prices (1990). In columns (4) - (7), we show the analogous results for the 1990 to 2000 period using the Case Shiller and the Census data. In columns (8) and (9), our definition of a neighborhood is a census tract. We can only examine the census tract patterns using the Census data. As noted in the paper, when using Census measures to compute house price appreciation, we also include controls to proxy for the changing neighborhood housing stock characteristics. These controls include: the change in the fraction of homes in the tract that are single-family-detached, the change in the fraction that have zero or one bedrooms, the change in the fraction that have two bedrooms, the change in the fraction that have three bedrooms, the change in the fraction built in the past 5 years, the change in the fraction built between 5 and 20 years ago, the change in the fraction built between 20 and 40 years ago, and the change in the fraction built between 40 and 50 years ago. When examining the census tract patterns we include all census tracts in MSAs (columns 8 and 9).

The results for the full sample of data mimic the results of the selected cities/MSAs shown in Table 2 of the paper. For example, using the Case Shiller data, neighborhoods where the initial median house price is twice as large as another neighborhood appreciated at a 24 percentage point lower rate during the 2000 to 2006 period (column 1). The results were nearly identical when we used lagged house price levels to instrument for initial house price levels in order to ensure that our results were not driven by measurement error in year 2000 house price levels (column 2). During the 2000s, when most US cities experienced a house price boom, a systematic feature of the data is that house prices in initially poorer neighborhoods systematically appreciated at higher rates than house prices in initially richer neighborhoods.

We also show that the difference in house price growth between low and high priced neighborhoods grows with the size of the MSA-wide housing price boom. To do this, we estimate:

$$\frac{\Delta P^{i,j}_{t,t+k}}{P^{i,j}_t} = \mu_j + \omega_1 \ln(HP^{i,j}_t) + \omega_2 \ln(HP^{i,j}_t) * \frac{\Delta P^{j}_{t,t+k}}{P^{j}_t} + \epsilon^{i,j}_{t,t+k} \tag{R11}$$

where all similar variables are defined as above. From this regression, we are interested in the coefficient on the interaction between initial house prices in the neighborhood and the MSA wide house price appreciation, $\Delta P^{j}_{t,t+k}/P^{j}_t$. The coefficient ω_2 assesses whether the relationship between initial neighborhood median house price and subsequent neighborhood house price growth differs between MSAs that experience large MSA-wide housing price booms relative to MSAs that experience smaller MSA-wide housing price booms, or even a bust. To measure the MSA-wide housing price booms, we use the FHFA MSA-level house price appreciation.

Returning to Table R1, we show that all of the house price differential between poor and rich neighborhoods occurs in MSAs that experienced a positive city-wide housing boom. This result is shown in columns (3). When housing prices were fairly constant in the MSA, there was no difference in the appreciation rates of low price neighborhoods relative to high-price neighborhoods on average. However, the larger the house price increase within the city between 2000 and 2006, the more the low price neighborhoods appreciated relative to higher price neighborhoods.

Similar patterns are also found during the 1990s using the Case Shiller data at the zip code level, the Census data at the zip code level, and the Census data at the census tract level. The reason that during the 1990s the coefficients in columns (4), (6), and (8) are close to zero is because most MSAs did not experience MSA-wide house price increases during this period. These columns shows estimates of Equation (R10) and do not include the interacted term. However, for those MSAs that did experience an MSA housing price gain during the 1990s (like Denver and Portland), initially low priced neighborhoods appreciated at a substantially higher rate than initially higher priced neighborhoods during this time period.

Taken together, the results in Table 2 (in the paper) and Table R1 convincingly show that during city-wide housing price booms, neighborhoods with lower initial housing prices appreciated at much higher rates than neighborhoods with higher initial housing prices. These results are not isolated to just the current housing price boom. The patterns are similar within cities that experienced sizeable housing price booms in the 1990s (Denver and Portland) and in the 1980s (Boston). So, not only is there large variation in house prices within a city/MSA, the variation exhibits some consistent and robust patterns during city-wide housing price booms.

R3.2 Further Evidence for Fact3: The Variation in Appreciation Rates is Also Higher for Initially Low Price Neighborhoods During City-Wide Housing Booms

The difference in volatility between initially low priced neighborhoods and initially high priced neighborhoods is a robust feature of the data across the many cities in our sample. Table R2 shows the result of a test of whether the standard deviation of the high price (Quartile 4) neighborhoods is equal to the standard deviation of the low price neighborhoods (Quartile 1). The p-value indicates that we can reject at the less than 1% level that the standard deviations are equal.

To further illustrate this feature of the data, we estimate:

$$\left|\hat{\epsilon}^{i,j}_{t,t+k}\right| = \mu_j + \alpha_1 \ln(HP^{i,j}_t) + \alpha_2 \ln(HP^{i,j}_t) * \frac{\Delta P^j_{t,t+k}}{P^j_t} + \eta^{i,j}_{t,t+k} \tag{R12}$$

where $\left|\hat{\epsilon}^{i,j}_{t,t+k}\right|$ is the abolute value of the estimated error term from (R10) and $HP^{i,j}_t$, $\Delta P^j_{t,t+k}/P^j_t$, and μ_j are defined as above. This regression is designed to uncover whether there is more variability in housing price growth, for a given level of initial house prices, for neighborhoods with initially low housing prices across the cities/MSAs in our sample. Table R3 shows the results of this regression using the Case Shiller data for the 2000-2006 period (columns 1 and 2) and for the 1990-2000 period (columns 3 and 4). Columns 1 and 3 estimate (3) without the interaction term in either the first or second stages. For columns 2 and 4, we included the interaction term only in the second stage. All of these regressions focus on the MSA level data. As seen from the regression, during MSA level housing price booms, the volatility of house price growth is higher among neighborhoods with initially low levels of house prices compared to neighborhoods with initially high levels of house prices. This higher volatility among low price neighborhoods relative to high price neighborhoods dramatically increases when the MSA as whole experiences a larger housing price boom.

R3.3 Criteria for Broadest Census Tract Samples

To measure housing price appreciation at the census tract level, we use the Neighborhood Change Database (NCDB) variables for median owner-occupied home price for 2000 and 1990. All variables in the NCDB have been adjusted so that comparisons can be made across time using consistently defined year 2000 census tract boundaries. Our broadest tract-level sample for 1990-2000 uses all cities for which we could define appreciation rates of median house prices for at least 30 census tracts. We use the Census Bureau tract-level tabulation file to measure median owner-occupied house prices in 1980. Our sample of tracts for the 1980-1990 period contains the additional restriction that the tract boundary must have either not changed at all or changed only slightly from 1980 to 2000. We allow tracts to remain in the sample if their centroid has moved less than 100 meters and their area has changed by less than one eighth mile by one eighth mile (the size of a Chicago city block).

R3.4 Adding Age of Structure Controls to Table 3 of Main Text

Table R4 re-estimates columns 2, 5, 6, and 7 of Table 3 (from the paper) adding controls for the distribution of the age of residential buildings in the zip code or census tract. Variables indicating the fraction of buildings built in the past 0 to 5 years, 6 to 20 years, 21 to 40 years are included. The fraction built more than 41 years ago is excluded. Adding these controls does not markedly change the magnitude of the coefficient on log distance to nearest rich neighborhood. These results suggest that our main results in the paper are robust to the stories of gentrification put forth by Brueckner and Rosenthal (2009).

Table R1: Regression of House Price Appreciation on Initial Average House Price

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Initial Level of House Price	-0.24	-0.26	0.00	-0.02	0.02	0.01	0.03	-0.10	-0.07
	(0.05)	(0.05)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)
Initial Level of House Price ∗			-0.35		-0.70		-0.28		-0.28
MSA-wide House Price Appreciation			(0.04)		(0.23)		(0.20)		(0.11)
House Price Measure	C-S	C-S	C-S	C-S	C-S	Census	Census	Census	Census
Neighborhood Aggregation	Zip Code	Zip Code	Zip Code	Zip Code	Zip Code	Zip Code	Zip Code	Census Tract	Census Tract
Time Period	00-06	00-06	00-06	90-00	90-00	90-00	90-00	90-00	90-00
Observations	1,600	1,600	1,600	1,498	1,498	1,498	1,498	49,660	49,660

Note: Regression of neighborhood level house price appreciation on the initial house price in the neighborhood and the initial house price in the neighborhood interacted with the MSA wide house price appreciation. All regressions also include MSA fixed effects. In columns (1) - (5), we use the Case Shiller house price indices for the appreciation rate. Columns (1) - (3) are for the 2000 - 2006 time period. Columns (1) and (2) are the same except that in column (2) we instrument initial level of house prices (2000) with the lagged level of house prices (1990). Columns (4) and (5) are for the 1990-2000 period and are analogous to columns (1) and (3). In all specifications, the initial house price in the neighborhood is computed using the Census data. Robust standard errors, clustered by MSA, are shown in parentheses. In columns (6) and (7) we measure neighborhood house price appreciation using the growth of the census median zip code price and use the same sample as columns (4) and (5). Columns (8) and (9) switch from zip codes to census tracts and broaden the sample to all census tracts within MSAs. All regressions are weighted by the number of owner occupied housing units in the neighborhood in the initial year. The census tract regressions use a growth in median home price measure which is trimmed at the top and bottom 1%. See paper for additional details.

Table R2: Standard Deviation of Housing Price Growth by Initial Price Quartile, Case-Shiller Data

	(1) Quartile 4	(2) Quartile 1	(3) p-val of Quartile 4 = Quartile 1
2000 - 2006			
Case-Shiller MSAs	0.46	0.61	0.00

Notes: This table shows the standard deviation of Case-Shiller house price appreciation rates for neighborhoods grouped by quartile of initial housing prices within each MSA. Quartile 4 has the highest initial price zip codes within the MSA while quartile 1 has the lowest initial price zip codes within the MSA. The sample includes all Case-Shiller MSAs.

Table R3: Regression of Absolute Value of Zip Code House Price Appreciation Residuals on Initial Average House Price in Zip Code

	(1)	(2)	(3)	(4)
Initial Level of House Price	-0.01	0.02	-0.00	0.00
	(0.01)	(0.02)	(0.01)	(0.01)
Initial Level of House Price ∗		-0.05		-0.14
MSA-wide House Price Appreciation		(0.02)		(0.07)
Time Period	00-06	00-06	90-00	90-00
Observations	1,600	1,600	1,498	1,498

Note: All regressions include MSA fixed effects. Robust standard errors, clustered by MSA, are shown in parentheses. All regressions are weighted by the number of owner occupied housing units in the neighborhood in the initial year.

Table R4: Regression of Neighborhood House Price Appreciation on Distance to Nearest High-Price Neighborhood, Age of Housing Stock, and Other Controls, Across Different Samples With Different House Price Measures

Time Period	(R2)	(R5)	(R6)	(R7)
Log Distance to Nearest High-Price	-0.046	-0.231	-0.142	-0.145
Neighborhood	(0.018)	(0.041)	(0.033)	(0.027)
Log Distance to Nearest High-Price	-	0.066	0.078	0.080
Neighborhood * City Wide Bust Indicator		(0.048)	(0.028)	(0.031)
House Price Measure/	C-S	Census	Census	Census
Neighborhood Aggregation	Zip	Census	Census	Census
	Code	Tract	Tract	Tract
Time Period	00-06	90-00	90-00	80-90
Vector of Z Controls Included	Yes	Yes	Yes	Yes
Number of Observations	236	3,099	7,955	4,253
Mean Log Distance to Nearest High-Price Neighborhood	1.23	0.401	0.499	0.322
Std. Dev. Log Distance to Nearest High-Price Neighborhood	0.524	0.778	0.719	0.716

Note: Table is analogous to Table 3 in the paper except that controls for the age of the neighborhood housing stock are added. These controls include the fraction of housing units built in the past 5 years, the fraction built in the past 6 - 20 year, the fraction built in the past 21 - 40 years, and the fraction built in the past 41 - 50 years. The fraction built 50 or more years ago is omitted. Columns are numbered to correspond to column numbers of Table 3. For further details see notes to Table 3.