Robustness Appendix for 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

August 12, 2011

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

R1 Introduction

In this robustness appendix we do three things. First, we document three empirical facts about within-city housing price dynamics. Second we specify exactly how the census tract samples that we use in Table 3 columns 6 - 8, Table 5 column 3 - 5, Table 6, and Table 7. Third we present a version of Table 3 from the main paper that also includes controls for the initial age of residential structures in the neighborhood.

R2 Three Facts About Within-City House Price Dynamics

R2.1 Fact 1: Low Price Neighborhoods Appreciated More on Average Than High Price Neighborhoods During City-Wide Housing Booms

This section of the robustness appendix documents fully the results discussed in section 2.2 of the main text showing the three systematic facts about the differences in house price appreciation across neighborhoods during city wide housing price booms. To better fully describe the data, we now estimate the following simple relationship using the different housing price series:

$$\frac{\Delta P_{t,t+k}^{i,j}}{P_t^{i,j}} = \mu_j + \omega_1 \ln(HP_t^{i,j}) + \epsilon_{t,t+k}^{i,j} \tag{1}$$

where $\Delta P_{t,t+k}^{i,j}/P_t^{i,j}$ 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 $HP_t^{i,j}$ 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 panel A, we show results where t and $t+k$ are 2000 and 2006, respectively, using both the Case-Shiller (columns 1 and 2) and the Zillow data (columns 3 and 4). For all the specifications in panel A, our definition of neighborhood is a zip code within the MSA. We restrict our attention to zip codes where both the Case-Shiller and Zillow indices exist. In panel B, we show the analagous results for the 1990 to 2000 period using the Case-Shiller and the Census data. In columns 1-4 of panel B, our definition of a neighborhood is a Case-Shiller zip code. In columns 5-8, our definition of a neighborhood is a census tract. We can only examine the census tract patterns using the Census data. As noted above, when using Census measures to compute house price appreciation, we also include the 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 in panel B, we include all census tracts that span the Case-Shiller zip codes (columns 5 and 6) or all census tracts in MSAs that contain at least 100 tracts (columns 7 and 8).

The results for the full sample of data mimic the results of the selected cities/MSAs shown in Table R2. For example, using the Case-Shiller data, neighborhoods where the initial median house price is twice as large as another neighborhood appreciated at a 23 percentage point lower rate during the 2000 to 2006 period (panel A, column 1). The results were nearly identical using the Zillow data (panel A, column 2). Again, these specifications are identified from within-MSA movements in house prices. 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.

R2.2 Fact 2: The Difference in House Price Appreciation Between Low and High Price Neighborhoods Increases With the Size of the MSA-Wide House Price Boom

In this subsection, we 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:

$$\frac{\Delta P_{t,t+k}^{i,j}}{P_t^{i,j}} = \mu_j + \omega_1 \ln(HP_t^{i,j}) + \omega_2 \ln(HP_t^{i,j}) * \frac{\Delta P_{t,t+k}^{j}}{P_t^{j}} + \epsilon_{t,t+k}^{i,j} \quad (2)$$

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, $\Delta P_{t,t+k}^{j}/P_t^{j}$. 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. These results are shown in columns 2 (for the Case-Shiller data) and 4 (for the Zillow data). 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 1, 3, 5, and 7 are close to zero is because most MSAs did not experience MSA-wide house price increases during this period. These columns estimate equation (1) and do not include the interacted term.[1] 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 Tables 2 in the main paper and 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 and New York). 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.

R2.3 Fact 3: The Variation in House Price Appreciation Among Lower Price Neighborhoods is Higher

Returning to Figure 1 of the main paper, another feature of the data for the New York MSA is that house price appreciation among initially low priced neighborhoods exhibits substantially more volatility than the price appreciation rates among initially high priced neighborhoods. In particular, the standard deviation of housing prices in the lowest initial house price quartile was 29 percent while the standard deviation of house prices in the top initial house price quartile was only 5 percent. The differences were significant at the less than 1 percent level.

This 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. To illustrate this, we estimate:

$$\left|\widehat{\epsilon}^{i,j}_{t,t+k}\right| = \mu_j + \alpha_1 \ln(HP^{i,j}_t) + \alpha_2 \ln(HP^{i,j}_t) * \frac{\Delta P^j_{t,t+k}}{P^j_t} + \eta^{i,j}_{t,t+k} \tag{3}$$

where $\left|\widehat{\epsilon}^{i,j}_{t,t+k}\right|$ is the absolute value of the estimated error term from (2) 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 R2 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 equation (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

[1] Again, Appendix Table A3 in the main paper shows the MSAs that did experience large price appreciation during the 1990s.

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.

It is this variation among low priced neighborhoods that we will exploit to directly test the mechanism at the heart of the model we present in the next section. Why is it that some low price neighborhoods within a city (like Harlem in the 2000s) appreciate at substantial rates while other (like Midtown in the 2000s) equally low price neighborhoods do not experience similar house price growth?

R3 Census Tract Sample Specification

As we specify in the main paper, when using samples of census tracts we focus only on cities that had at least 30 tracts in the initial year. Furthermore we drop any census tracts that were in the top or bottom 1% of either house price growth or income growth to make sure that our results are not being driven by outliers. Also, since census tracts may change over time, in our 1980 - 1990 samples we limit our sample to census tracts that did not change at all or changed very little (the centroid must have moved by less than 100 meters and the area must have changed by by less than 40,469 square meters: the size of a Chicago city block).

R4 Initial Structure Age

In order to make sure that our results are not simply being driven by a cycle of housing stock depreciation and rehabilitation, Table R3 reproduces Table 3 from the main paper, but adds controls for the initial age of the housing stock in the neighborhood. These controls include: the fraction of housing units built in the past 5 years, 5 - 10 years ago, 10 - 20 years ago, 20 - 40 years ago, and 40 - 50 years ago.

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

Panel A: 2000 - 2006 Appreciation Rates

	Case-Shiller		Zillow	
	(1)	(2)	(3)	(4)
Initial Level of Neighborhood House Price	-0.23	0.01	-0.28	0.04
	(0.05)	(0.04)	(0.09)	(0.05)
Initial Level of Neighborhood House Price * MSA Wide House Price Appreciation	-	-0.35	-	-0.48
		(0.04)		(0.09)
Number of Observations	1,617	1,617	1,617	1,617

Panel B: 1990 - 2000 Appreciation Rates

	Case-Shiller		Census Median		Census Median		Census Median	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Initial Level of Neighborhood House Price	-0.02	0.01	0.02	0.03	-0.08	-0.06	-0.06	-0.03
	(0.04)	(0.03)	(0.03)	(0.02)	(0.06)	(0.06)	(0.02)	(0.02)
Initial Level of Neighborhood House Price * MSA Wide House Price Appreciation	-	-0.70	-	-0.28	-	-0.31	-	-0.27
		(0.23)		(0.17)		(0.23)		(0.14)
	1,506	1,506	1,506	1,506	9,670	9,670	39,709	39,709
	Zip Code		Zip Code		Census Tracts that span Case-Shiller zip codes		Census Tracts in all MSAs that have at least 100 tracts	

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 panel A (the 2000's), we use the Case-Shiller house price indices for the appreciation rate (columns 1 and 2) and the Zillow house price indices for the appreciation rate (columns 3 and 4). In panel B (the 1990's), we use the Case-Shiller indices for the appreciation rate (columns (1) and (2)) and the Census data for the appreciation rate (columns 3 - 6). 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. 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 text for additional details.

Table R2: Regression of Absolute Value of Zip Code House Price Appreciation Residuals on Initial Average House Price in Zip Code

	Case-Shiller 2000 - 2006		Case-Shiller 1990 - 2000	
	(1)	(2)	(3)	(4)
Initial Level of Neighborhood House Price	-0.01	0.02	-0.00	0.00
	(0.02)	(0.02)	(0.01)	(0.01)
Initial Level of Neighborhood House Price $*$ MSA wide House Price Appreciation	-	-0.04		-0.14
		(0.02)		(0.07)
Number of Observations	1,617	1,617	1,506	1,506

Note: Table reports the regression of the absolute value of residuals from (2) on the initial level of house prices in the neighborhood (columns 1 and 3) and the initial level of house prices interacted with MSA wide house price appreciation (columns 2 and 4). All regressions use Case-Shiller data. The first two columns use house price growth between 2000 and 2006. The last two columns use house price growth between 1990 and 2000. 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 R3: Regression of Neighborhood House Price Appreciation on Distance to Nearest High-Price Neighborhood and Other Controls (Including Initial Age of Structure), Across Different Samples With Different House Price Measures

Time Period	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log Distance to Nearest High-Price Neighborhood	-0.055 (0.020)	-0.052 (0.019)	-0.033 (0.029)	-0.020 (0.030)	-0.047 (0.038)	-0.237 (0.042)	-0.147 (0.033)	-0.145 (0.027)
Log Distance to Nearest High-Price Neighborhood * City Wide Bust Indicator	-	-	-	-	0.079 (0.050)	0.066 (0.049)	0.082 (0.028)	0.074 (0.031)
House Price Measure/ Neighborhood Aggregation	C-S Zip Code	C-S Zip Code	Zillow Zip Code	C-S Zip Code	C-S Zip Code	Census Census Tract	Census Census Tract	Census Census Tract
Time Period	00-06	00-06	00-06	90-00	90-00	90-00	90-00	80-90
Vector of Z Controls Included	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of Observations	208	208	208	223	223	3,015	7,981	4,251
Mean Log Distance to Nearest High-Price Neighborhood	1.23	1.23	1.23	1.22	1.22	0.397	0.497	0.320
Std. Dev. Log Distance to Nearest High-Price Neighborhood	0.538	0.538	0.538	0.488	0.488	0.784	0.719	0.717

Note: Table shows regression of neighborhood house price appreciation between period t and t+k on log distance to nearest high price neighborhood within the neighborhood's city, city fixed effects, and a vector of neighborhood controls. In addition to the controls described in the main paper, controls for the fraction of housing units built in the past 5 years, 5 - 10 years ago, 10 - 20 years ago, 20 - 40 years ago, and 40 - 50 years ago are also included. High price neighborhoods are those neighborhoods that are within the top quartile of average neighborhood house prices in year t. We restrict our analysis in this table to those neighborhoods within the city which were in the bottom half of the house price distribution in period t. See text for additional sample descriptions and discussion of the controls included. 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.