FHFA STAFF WORKING PAPER SERIES

Local House Price Dynamics: New Indices and Stylized Facts

Alexander N. Bogin
William M. Doerner
William D. Larson

June 2016 (revised)
April 2016 (original)

Working Paper 16-01

FEDERAL HOUSING FINANCE AGENCY
Office of Policy Analysis & Research
400 7th Street SW
Washington, DC 20024, USA

Please address correspondences to William Larson (william.larson@fhfa.gov). Many thanks to Morris Davis, Jim Follain, Andy Leventis, Scott Smith, Tony Yezer, and seminar participants at the ASSA Annual Meetings, the AREUEA Mid-Year Meetings, the ARES National Conference, and GWU Microeconomics Workshop, who provided helpful questions, comments, and discussion.

Local House Price Dynamics: New Indices and Stylized Facts

Alexander N. Bogin, William M. Doerner, and William D. Larson

FHFA Staff Working Paper 16-01

June 2016 (revised)

April 2016 (original)

Abstract

We construct annual house price indices for nearly 18,000 ZIP codes in the United States over a 40 year period. Between 1990 and 2015, house price gradients within large cities steepen, documenting a reversal of decades of increasing relative desirability of suburban locations. Additionally, real house prices are more likely to be non-stationary near the city-center, consistent with a higher elasticity of housing supply near the edge of the city.

Keywords: aggregation bias · house prices · standard urban model · land leverage · transportation costs

JEL Classification: C43 · R14 · R30

Alexander N. Bogin
Federal Housing Finance Agency
Office of Policy Analysis & Research
400 7th Street SW
Washington, DC 20024, USA
alexander.bogin@fhfa.gov

William M. Doerner
Federal Housing Finance Agency
Office of Policy Analysis & Research
400 7th Street SW
Washington, DC 20024, USA
william.doerner@fhfa.gov

William D. Larson
Federal Housing Finance Agency
Office of Policy Analysis & Research
400 7th Street SW
Washington, DC 20024, USA
william.larson@fhfa.gov

1. Introduction

Much of the 20th century in the United States can be characterized as one of suburbanization. Rapid improvements in both automobile technology and road capacity decreased transportation costs, increasing the desirability of areas far from city centers which were often plagued by pollution, crime, and other hazards (Mieszkowski and Mills, 1993; Glaeser, 2011). Consequently, as relative demand for center-city housing fell, so too did prices. This trend persisted into the 1980s, but recent evidence suggests a reversal: after decades of hollowing out, center-cities are becoming increasingly popular. Median home values in center-cities have been increasing, according to Glaeser, Gottlieb, and Toibo (2012), Edlund, Machado, Sviatchi (2015), and others. Popular media is awash in articles on "millennials," (New York Times, 2014) and the rise of center-cities as destinations for the young and eager "creative class" (Florida, 2004; Couture and Handbury, 2016). This is coincident with Moretti's (2012) "Great Divergence" of jobs, income, and increased inequality in U.S. cities (see also, Diamond, 2016).

While this body of research has produced robust evidence of center-city revitalization, what is lacking is direct evidence of shifts in house price gradients within cities. The house price gradient within a city captures the tradeoff between demand for housing at different locations, and is therefore, a fundamental economic indicator of relative center-city demand. In this paper, we provide the first evidence documenting the steepening of house price gradients over a broad cross-section of cities, and simultaneously, over long time periods. Until now, within-city house price measurement has been a persistent blind spot for those interested in urban research, policy, and finance. This has limited our ability to perform panel analyses and quasi-natural experiments within and across cities.

Previously, researchers have relied on two main types of data sources to investigate within-city price movements. Both have serious issues that limit their usefulness in a variety of applications. The first category consists of value measures such the American Community Survey, the Decennial Census, or Zillow's home value index, which confound price and quantity changes unless strong assumptions are valid.[1] The second category consists of proprietary price index data such as Black Knight, CoreLogic, or Case-Shiller, which are produced using

[1] Value changes can be decomposed into price and quantity changes ($V = P \times Q$, so $\%\Delta V \approx \%\Delta P + \%\Delta Q$ for small changes), and therefore only acts as a price index in areas where the quantity of housing services provided by new units is identical to the existing stock ($\Delta Q = 0$).

limited transactions data prior to the late 1980s or early 1990s, forcing reliance on geographic pooling of transactions, smoothing of series over space or time, or limiting coverage.[2]

To address this gap in the availability of constant-quality house price measures, we construct a comprehensive set of annual HPIs over four decades for cities, counties, 3-digit ZIP codes (ZIP3s), and 5-digit ZIP codes (ZIP5s) using a repeat-sales methodology.[3] The tradeoff of this lower level of geographic aggregation is a higher level of time aggregation. The result is a panel of annual indices for every area for which data are available, each estimated using the same method and source data.[4] Our final database includes house price indices for 914 CBSAs (381 MSAs and 533 MicroSAs), 2,742 counties, 879 3-digit ZIP codes, and 17,936 5-digit ZIP codes.[5]

Two main stylized facts emerge with these new HPIs. Within large cities, appreciation rates and volatility between 1990 and 2015 are higher near the central business district (CBD) than in the suburbs.[6] According to augmented Dickey-Fuller tests, real house prices are non-stationary in the majority of ZIP codes over the sample. Within cities, non-stationarity occurs at higher rates near the CBD compared to the suburbs, and house prices in large cities are more often non-stationary than those in small cities. In small cities, the appreciation gradient is flat.

[2]For instance, the Case-Shiller ZIP code house price data used by Mian and Sufi (2009) and Guerrieri, Hartley, and Hurst (2013) "GHH" is proprietary and begin in the late 1980s, with coverage including 1,498 ZIP codes beginning in 1990 (see Column 3 of Table 1 in GHH). A major issue with proprietary data is highlighted in footnote 5 of GHH: "Unfortunately, we only have the data through 2008 and, as a result, we cannot systematically explore within-city house price patterns during the recent bust. We have been unsuccessful in our attempts to secure the post 2008 from [the company overseeing the Case-Shiller index] Fiserv."

[3]We use the term "cities" to refer to Core Based Statistical Areas (CBSAs) that may be further defined into Metropolitan Statistical Areas (MSAs) and Micropolitan Statistical Areas (MicroSAs). It should be emphasized that no data from other time periods or areas are incorporated into a particular area's index calculation. This is opposed to the common practice of augmenting index values constructed with sparse observation counts by temporal or spatial information, or with alternative methods. Rather we utilize only the repeat-sales method to produce a broad and consistent set of indices with an information set that is strictly limited to the time period and location of measure.

[4]A potential concern with such a comprehensive set of indices is that some periods, especially early in the sample and in sparsely populated areas, may have low observation counts which could lead to higher variance in estimates. Clapp, Giaccotto, and Tirtiroglu (1991) show, while short-term estimates of house prices may contain noise, these differences do not compound and instead offset after several years.

[5]These data are freely and publicly available at: `http://www.fhfa.gov/papers/wp1601.aspx`.

[6]A wide body of research seeks to empirically estimate house price and appreciation gradients, including Yinger (1979) and Coulson (1991). Recently, Edlund, Machado, and Sviatchi (2015), observe property values and changes decrease with distance from the CBD.

We then consider these facts in light of established urban theory. The Standard Urban Model (SUM) contains predictions about the movement of house prices over the recent period. While demand for housing increases at all locations in most cities due to increasing population and incomes, center-city desirability may be increasing in relative terms due to the broad trends of decreasing center-city crime, increasing transportation costs, and increasing amenties.[7] Our new house price indices show a steepening house price gradient in large cities, suggesting that demand for locations near the center-city is, in fact, increasing. We conduct further tests, and find the steepening house price gradient to be robust with respect to a variety of covariates, including city size, commuting time and modal choice, income amount and source, household characteristics, and labor force status. Finally, the volatility and stationarity results are consistent with a relatively high elasticity of housing supply in suburbs and in smaller cities (Glaeser, Gyourko, and Saiz, 2008).[8]

These results are also broadly consistent with the Land Leverage Hypothesis (LLH) of Bostic, Longhofer, and Redfearn (2007) and Davis and Heathcote (2007). This hypothesis uses an accounting identity to separate the value of a housing unit into structure and land components. Under the assumption that the elasticity of land supply is less than the elasticity of structure supply, the LLH predicts higher house price appreciation and volatility in high land leverage areas. With the additional assumption that land leverage decreases with distance from the CBD, as is shown to be the case in Washington, DC, by Davis, Oliner, Pinto, and Bokka (2016), the estimates here support the LLH.[9]

Overall, several results and predictions related to the Great Divergence are supported by our findings. As the price gradient rotates in large cities, wealth is redistributed from owners of land in the suburbs to owners in the center-city, potentially increasing inequality. Additionally, in the SUM, households and firms who consume less housing find the the center-city relatively more attractive, thus explaining the rise of millennials and the service industry in the center-city. Therefore, the house price results in this paper act as strong corroborating

[7] The foundations of the SUM are in Alonso (1964), Mills (1967), Muth (1969). Brueckner (1987) presents the textbook rendition of the model, along with derivations of key results.

[8] Under the supply elasticity hypothesis, a permanent demand increase will result in temporary house price increases in elastic areas. On the other hand, where supply is restricted through regulation, topography (see Saiz, 2010) or existing development, house price increases are more permanent. In the long run, the SUM predicts no differential house price effects due to differential supply elasticities.

[9] While estimates support the LLH over the last 30 years, prior to the mid 1980s, increasing suburbanization would suggest the opposite of the LLH, with house prices rising in low land-expenditure-share areas in the suburbs.

evidence for the findings of Couture and Handbury (2016) and Diamond (2016), who find that young, high human capital workers are attracted to center cities. While these demographic groups may be attracted to amenities, it is the high house prices and their relatively low housing demand that may allow these groups to outbid others for scarce center-city space.

The remainder of this paper is structured as follows. The next section outlines how our HPIs are constructed and compares them to existing indices. The third section gives a basic overview of Muth's Equation in the SUM, which frames the following discussion of stylized facts from the indices. The final section concludes with a summary, implications, and potential applications of these new house price indices.

2. Repeat-Sales Index Construction

It can be challenging to estimate a house price index. Data are often held by private listing services or public entities that limit access due to proprietary protections and privacy concerns. Even with accessible data, an accurate market price for a set of housing units can be difficult to compute because several years may pass between individual sales and characteristics may vary across units. Due to these facts, a large volume of transactions is necessary to construct a constant-quality index, often requiring aggregation over time or space.

Table 1 shows the major United States HPIs that are available online at no pecuniary cost, along with the time horizon, frequency, and level of geographic aggregation. Geographic aggregation at the city level is currently the standard, allowing for high-frequency (monthly or quarterly) house price measurement, but at the cost of smoothing over variation within the city. This leaves a serious gap in our knowledge of house price dynamics within submarkets of cities. Therefore, instead of aggregating over space in order to capture enough housing transactions to estimate an index, we aggregate over time by reducing the frequency to an annual panel. As the table shows, we produce long panels of annual HPIs at various levels of geography. While this may introduce temporal aggregation bias, especially concerning volatility (Calhoun, Chinloy, and Megbolugbe, 1995), it allows us to document house price movements across smaller areas in ways that have not been explored before and are otherwise impossible to analyze when using higher levels of geography, like city, county, or ZIP3 HPIs.

Using a rich, proprietary dataset of mortgage transactions going back to the 1970s, we are able to construct indices down to the 5-digit ZIP code-level across the nation using a repeat-sales methodology. This technique is attractive because of its limited data requirements for each observed transaction (i.e. only a property identifier, sales price, and date are needed for each), though it results in discarding transactions when multiple sales for the same unit are not observed in the dataset.[10] The repeat-sales methodology is explained below.

Suppose the (natural log) value y of house i at time t can be written as follows,

$$y_{it} = x'_{it}\beta + D'_t\delta + \epsilon_{it} \tag{1}$$

where x is a vector of attributes, β is a vector of relative implicit prices, δ is a vector of price levels over time, and D is a vector of traditional dummy variables set equal to one in period t and 0 otherwise. The empirical estimation of β could include a wealth of information about the house's structural characteristics, location, and surrounding neighborhood attributes. Fortunately, the conventional weighted repeat-sales (WRS) methodology uses a differencing technique that eliminates the need to estimate β by "pairing" the same house across periods, or

$$y_{it} - y_{i\tau} = (x_{it} - x_{i\tau})'\beta + (D_t - D_\tau)'\delta + (\epsilon_{it} - \epsilon_{i\tau}). \tag{2}$$

Under the assumption that the house's characteristics, both observable and unobservable, are constant across both transaction periods, $x_{it} - x_{i\tau} = 0$, the model becomes

$$y_{it\tau} = D'_{t\tau}\delta + \epsilon_{it\tau} \tag{3}$$

where $y_{it\tau} \equiv y_{it} - y_{i\tau}$, $D_{t\tau} \equiv D_t - D_\tau$, and $\epsilon_{it\tau} \equiv \epsilon_{it} - \epsilon_{i\tau}$. There are some important shortcomings of the repeat-sales assumptions, which are noted in the footnote below. Generally, these include issues related to depreciation, renovations, distressed sales, and other factors which may influence value differentials across sales that are unrelated to movements

[10] Repeat-sales house price indices have a long tradition, dating back to Bailey, Muth, and Nourse (1963). The index approach gained wide notoriety in the late 1980s with Case and Shiller's (1987, 1989) seminal work, and has since been the subject of intense research, including Calhoun (1996) that describes the Federal Housing Finance Agency (FHFA)-specific methodology. Recently, newer methods have been developed to exploit the broader information sets (e.g., public and private appraisals, real estate listings, official recordings, geographic maps). Unfortunately, even though data digitization has provided more information about property- and neighborhood-level characteristics, it seldom captures sales information prior to the mid-1990s. While newer techniques may have empirical advantages over the repeat-sales methodology, they tend to be data intensive and are constrained to shorter time periods as well as limited geographic coverage.

in the price of housing services.[11] Despite these potential problems, the constant-quality assumption is necessary for the index to be estimated.

A statistical issue with Equation 3 is that the error variance may be predictable based on the time between transactions, $t - \tau$. This knowledge gives justification for the use of a FGLS procedure, where the estimated first stage residuals are in turn modeled as a function of the time between transactions and the time squared:[12]

$$\hat{\epsilon}^2_{it\tau} = \alpha_1 + \alpha_2(t-\tau) + \alpha_3(t-\tau)^2 + e_{it\tau}. \tag{4}$$

The fitted residuals from this auxiliary equation, $\tilde{\epsilon}^2_{it\tau} = \hat{\alpha}_1 + \hat{\alpha}_2(t-\tau) + \hat{\alpha}_3(t-\tau)^2$ are applied as weights in a second-stage regression written as

$$\frac{y_{it\tau}}{\sqrt{\tilde{\epsilon}^2_{it\tau}}} = \left(\frac{D_{t\tau}}{\sqrt{\tilde{\epsilon}^2_{it\tau}}}\right)' \delta + \frac{\epsilon_{it\tau}}{\sqrt{\tilde{\epsilon}^2_{it\tau}}}. \tag{5}$$

The house price index is then computed by exponentiating the estimated $\hat{\delta}_t$ as $I_t = \exp \hat{\delta}_t$, which we normalize so that the base year is 1990, or $I_{1990} = 100$.

In this paper, housing unit transaction data are from the FHFA's "all-transactions" sample, which includes conventional mortgages of single-family purchases and refinances that are acquired or guaranteed by Fannie Mae or Freddie Mac.[13] This dataset contains over 97

[11] There are some well-known issues with the assumption that characteristics do not change. For instance, a house can depreciate over time by as much as 2% per year (Harding, Rosenthal, Sirmans, 2007). Additionally, renovations may improve the quality or size of a housing unit, with recent evidence suggesting the renovation or "flip" bias can range from 1.5% to as much as 20% (McMillen and Thorsnes, 2006; Depken, Hollans, Swidler, 2009; Billings, 2015). In addition to changes in structure attributes, the conditions of the sale may also affect the price. For example, house prices can be depressed when homeowners terminate mortgage payments and are either forcefully evicted through the foreclosure process or voluntarily dispose of their houses in a lender-permitted "distressed" sale (Leventis, 2009). Foreclosures have been shown to depress the value of surrounding houses (Immergluck and Smith, 2006; Harding, Rosenblatt, Yao, 2009) and a growing concentration of them can lead to price discounts in excess of 25% (Campbell, Giglio, and Pathak, 2011; Depken, Hollans, Swidler, 2015). The large discount is partially attributable to the long foreclosure process after delinquency as well as possible deterioration of property condition. Since there is a shorter timeline between when a borrower stops payments and a distressed sale happens, the time-varying discounts are smaller, with a range between 5 and 20% (Depken, Hollans, Swidler, 2015; Doerner and Leventis, 2015). This negative impact eventually vanishes as the house sells again and better information is available about its physical condition.

[12] Note that the FHFA methodology omits the constant term in Equation 4 due to the possibility of negative predicted variance.

[13] This sample does not include Federal Housing Administration or non-conventional loans.

million transactions with nearly 54 million transaction pairs in the United States from 1975 to 2015.

We construct annual HPIs for each area with at least 100 repeat sales (RS) in the data, subject to two filters, the purpose of which are to eliminate transaction pairs from the sample that are likely to violate the constant-quality assumption.[14] The first removes any pair of transactions with annual average appreciation rates greater than +/-40%. This filter is common in repeat-sales index construction and serves to remove homes from the sample that have undergone substantial quality changes. The second filter removes any pair of transactions where the two sales are within the same 12 month period. This filter is meant to remove "flips," or homes that are purchased, rehabilitated, then sold in a short length of time.[15] For each area with over 100 RS, the index begins once 25 half-pairs (HP) are observed in a single year in the admissible RS sample.[16]

Table 2 shows our local HPIs have longer horizons and greater granularity than is currently available from any other publicly available data source. About half of the MicroSA, county, and 5-digit ZIP code level indices begin in 1990 or earlier. The vast majority of MSA and 3-digit ZIP codes begin in 1980 or earlier. Geographic coverage at the turn of each decade is given in Figure 1, showing the increasingly broad cross-section of data available over time. As an illustration of several of the indices constructed, Figure 2 shows selected real HPIs (net of the all-goods CPI) for various ZIP codes and counties in the Washington D.C. metropolitan statistical area. There is wide variation in index values across different measures of geography, especially in later periods. Noise is often high in early periods due to low observation counts, but there is enough signal to extract some useful information using econometric methods.

[14] There is also non-random selection into our sample which we do not believe to be problematic. The loan values in our dataset must be below the FHFA's conforming loan limit. This right-censoring can substantially reduce the number of transactions we observe in areas where the price level is high. While this would obviously distort a median value index as well as price index estimation error variance, we do not find compelling evidence that it biases our estimated appreciation rates.

[15] The appreciation rate filter is implemented as the annual average log-difference of 0.3, which is approximately the FHFA's filter of 40% per year. The holding period filter of one year is larger than the FHFA's filter of 3 months, but is necessary due to the annual frequency of the present indices. Beyond these two filters, the FHFA also implements a small number of supplemental data screens. To maximize the data available for estimating the 5-digit ZIP code indexes, the supplemental screens have not been used in this paper.

[16] We create the term "half-pairs" as a measure of the information content in a particular year. This is calculated as $HP_t = \sum(abs(D_{it}) + abs(D_{i\tau}))$.

To compare similar submarkets across cities, we define several regions based on distance to the CBD for 5-digit ZIP codes.[17] These include: the center city, which consist of ZIP codes within 5 miles of the CBD; the middle city, which consists of ZIP codes between 5 and 15 miles; the suburbs, which consists of ZIP codes between 15 and 25 miles; and the exurbs, which consist of ZIP codes beyond 25 miles from the CBD. ZIP codes in CBSAs are more likely to have high transaction counts, and thus begin in earlier periods. Within CBSAs, the closer a ZIP code is to the CBD, the earlier the start date. These submarket classifications are important because they enable measurement and tests of sub-market variation in house price dynamics, including gradients of the rate of appreciation, appreciation volatility, and stationarity.

3. Theory and Stylized Facts

These new constant-quality house price indices allow us to explore two aspects of submarket appreciation that have previously been difficult to investigate because suitable indices have not been widely available to researchers.[18] First, between 1990 and 2015, real house price appreciation is consistently higher near the centers of large cities. Echoing these results are augmented Dickey-Fuller tests, which indicate that areas near the center of the city are less likely to have stationary real house prices than areas in the suburbs, suggesting greater levels of mean reversion in house price appreciation the further an area is from the CBD. Second, appreciation volatility shows similar spatial patterns as appreciation, with higher values closer to the CBD, and to a greater degree in large cities, controlling for the number of repeat-sales used in the construction of the index. These two facts show that house price appreciation rates and volatility vary both within and across cities in systematic ways. Further investigation shows these results to be highly correlated with known elements of the Standard Urban Model.

3.1. Muth's Equation and the House Price Gradient

We now introduce the SUM as it relates to transportation costs in order to frame the discussion of facts revealed by our house price indices. The standard rendition of the SUM assumes a monocentric city with exogenous employment in the CBD (Alonso, 1964; Mills, 1967; Muth, 1969). Households commute to this center at a cost t per unit of distance k,

[17] The location of the CBD is identified as the ZIP code within a CBSA with the largest standardized fraction of housing units in 20+ unit structures minus the standardized land area. Visual inspection shows this gives a reasonable approximation of distance to the CBD. Distance is calculated from centroid to centroid.

[18] While these concepts have been examined in previous work, samples are frequently limited to a small number of large cities and value measures are often used as proxies for prices.

creating a downward-sloped bid-rent curve. Equation 6, below, is termed "Muth's Equation," and shows that house prices fall with distance to the CBD at a rate of $-t/H(k;\alpha)$, where H is housing consumption at location k under a vector of exogenous demand shifters α.

$$\frac{\partial P^H(k)}{\partial k} = \frac{-t}{H(k;\alpha)} \tag{6}$$

This equation shows that the house price gradient is driven by transportation costs and other exogenous factors within α that would influence the desirability of housing as it relates to distance to the CBD. These exogenous factors may consist of variables such as income (combined with an income elasticity of demand for housing that is less than 1), exogenous increases or decreases in center-city crime, or exogenous demographic shifts related to household size and formation.[19]

Empirically, however, the absolute distance to the CBD may be an imperfect measure. For instance, non-radial transportation networks and city polycentricity may result in non-monotonic transportation costs as a function of absolute distance (McMillen and Smith, 2003). Commuting costs are also likely correlated with other variables that are measurable, such as structure density and the housing consumption gradient (Brueckner, 1987). Because these gradients are each governed by the same relationships in the SUM, these variables may provide additional information.

Beyond the canonical rendition of the SUM, a large body of literature has found other gradients are logically consistent with the model. For example, the modal choice of commuting (Voith, 1991) is related to k, with more households choosing to commute via automobile the further from the CBD because car ownership is land-intensive. Other variables have known gradients that can be derived from the consumer's utility function, including incomes, marital status, the number of children, and employment and labor force status. For non-workers, income is endowed and commuting transportation costs are eliminated, causing such households to reside further from the CBD where transportation costs for would-be commuters are higher (Blackley and Follain, 1987). Similarly, households with children demand more housing, suggesting optimal residence in the suburbs and in low housing-cost cities (Black et al., 2002). Finally, an income elasticity of demand for housing less than one suggests higher

[19] Edlund, Machado, and Sviatchi (2015) identify and discuss several potential determinants of rotations in the house value gradient between 1980 and 2010, including those listed above.

income households will live in the suburbs, where housing expenditure shares are lower.[20] These alternative correlates of commuting costs can be used to further evaluate changes to the desirability of proximity to the center-city during the time period.

3.2. Appreciation Within Cities

The ZIP code house price indices suggest price appreciation tends to be highest near the CBD between 1990 and 2015.[21] Figure 3 shows average appreciation rates for nine large cities over a 25 year period. Dark blue cells indicate higher appreciation rates and are generally clustered in center-city ZIP codes while lighter cells indicate lower appreciation rates and are more common on the peripheries. This pattern exists in a variety of other cities, as well. Figure 4 measures this appreciation gradient across all of the cities in our sample from 1990 through 2015. It illustrates a negative correlation between ZIP code-level house price appreciation and distance to the CBD. In CBSAs with over 500,000 housing units, house price appreciation declines markedly with distance (most apparent in the first 15 miles from the CBD); in smaller cities the appreciation gradient is flat. For large cities, average real appreciation is about 2% per year in areas near the CBD compared to an appreciation rate of about 1% in areas that are 10 miles from the CBD.[22] At 25 miles from the CBD, average real appreciation in both large and small cities is about 0.3% per year. In general, ZIP code-level house price measures suggest that house price gradients have steepened in large cities over the last several decades.

The stationarity of real house prices exhibits similar spatial patterns. Augmented Dickey-Fuller (ADF, 1981) regressions are estimated for each ZIP code's full time series in order to test for the presence of a unit root.[23] Figure 5 visually depicts how stationarity changes

[20]This elasticity was previously established to be less than one (Deaton and Muellbauer, 1980, for example), though recent research suggests that it may in fact be equal to one (Davis and Ortalo-Magne, 2011).

[21]We discuss ZIP5 indices for expository purposes. The sample period of 1990 to 2015 is chosen for two reasons. By 1990, a large number of ZIP codes have reliable house price index estimates. Additionally, in later sections, we examine effects of various initial ZIP code characteristics on appreciation. Because a decennial census occurred in 1990, this year becomes a natural choice for the start of the analysis. Figures and estimates are robust to the inclusion of all years, though the panel becomes unbalanced.

[22]City size is measured in 1990, so categories are based on the number of housing units at this date. However, between 1990 and 2015, cities may grow at different rates. As a robustness exercise, we reproduced the same figure, splitting the sample into growing and declining cities. Appreciation has similar growth rates near the CBD, but the mean growth rate is higher in growing cities in the suburbs. Thus, the appreciation gradient is slightly flatter in growing cities.

[23]A deterministic constant term is modeled but not a trend, because while areas may have different real levels of house prices due to some economic advantage, it is difficult to justify a model with real house prices trending upward *ad infinitum*.

across a city. The null hypothesis of a unit root is rejected more often (1.5 to 2.5 times) in smaller cities than in large cities. Additionally, unit roots are increasingly rejected the further a ZIP code is located from the CBD in both small and large cities. For both small and large cities, real house prices are more likely to be stationary near the edge of the city and less likely to be stationary the closer the ZIP code is to the CBD.

We further explore appreciation rate differentials by stochastically specifying house price appreciation in ZIP code z in city c as a function of the distance to the CBD, k, where ZIP code-level house price appreciation is approximated by the log difference, Δp. Distance to the CBD is not a perfect measure of commuting costs in a real-world city, so we also include a vector of covariates X which may include the modal choice of commuting, income, household demographics, and other factors that may be correlated with transportation costs. The following estimating equation includes CBSA fixed effects, allowing us to isolate within-city variation,

$$\Delta p_{c,z} = \alpha_c + \gamma k_z + X_z' \beta + v_{c,z} \tag{7}$$

Equation 7 is estimated for a variety of potential covariates, calculated using data from the 1990 Decennial Census.[24]

Table 3 shows how different initial covariates are related to cumulative real house price appreciation between 1990 and 2015. This cross-section is the 2,742 ZIP codes within 25 miles of the CBD ZIP code that exist in a single large (over 500,000 housing units) CBSA. Covariates are measured at the ZIP code-level and are loosely grouped into five main categories: transportation, structure, labor force, earnings, and housing demand. Each covariate relates to a SUM concept with a known gradient slope as a function of commuting costs.

In general, results indicate that commuting costs and other correlates of proximity to center-city locations drive house price changes in large cities over the sample period. The distance to the CBD has a significant, negative relationship with appreciation, with a doubling of CBD

[24] Because ZIP codes change over time, every transaction pair has a ZIP code, but not every ZIP code has 1990 Census information. Of the 30,790 ZIP codes in the house price data, there are 2,526 ZIP codes that do not exist in the 1990 Census data but do exist in the 2015 ACS. Additionally, there are 2,067 ZIP codes that exist in the 1990 Census that do not exist in the housing transactions database. Because ZIP codes tend to be created in fast-growing areas, fast-growing areas tend to have high land values, and high land values are correlated with low transportation costs, this selection problem may introduce a downward bias in the transportation cost effects.

distance decreasing real house price appreciation by 15%.[25] The addition of transportation variables, as shown in Column 2, increases the explanatory power of the model, suggesting travel time and commuting method are important appreciation determinants over the sample period. Column 3 considers structure attributes, with denser structures and smaller units suggesting closer proximity to the CBD and an association with positive appreciation. In Column 4, the labor force fraction is predicted by the SUM to be positively associated with the proximity to the CBD because of the need for workers to commute. However, the estimated sign is negative, suggesting that the income and perhaps other effects are inducing those in the labor force to move to the suburbs. Retirees are predicted to live further from the CBD, which appears to be the case from the negative partial correlation with appreciation. Income effects in Column 5 are all as predicted, with higher income associated with lower appreciation, lending support to the notion that the income elasticity of demand for housing is less than one. Column 6 presents two variables that are associated with housing demand—children and being married. Both of these attributes suggest greater demand for space, a resulting suburban location, and reduced house price appreciation. A model which includes all these covariates suggests transportation, structure, labor force, and housing demand attributes are each driving some of the residual variation.

3.3. Volatility Within Cities

Appreciation volatility shows similar spatial characteristics to appreciation rates. Figure 6 shows the median absolute deviation of annual house price changes for each ZIP code as it relates to distance from the CBD.[26] Unconditionally, house price volatility in large cities decreases slightly with distance from the CBD. As illustrated, volatility in small cities appears to dip and then increase with distance from the CBD, though this may be confounded by increasing estimation error.[27]

Table 4 presents estimates of volatility as a function of the same covariates in Equation 7. These models relate the median absolute deviation in year-on-year log differences in house prices to SUM variables.[28] The volatility results echo the appreciation results, increasing

[25] These results do not hold for small cities, as Figure 4 and Appendix Table A-3 suggest.

[26] Median absolute deviation is defined as the median deviation from the median appreciation rate over the maximum sample period for a particular ZIP code, or $MAD_{c,z} = med_{c,z}(|Y_{c,z,t} - med_{c,z}(Y_{c,z})|)$. This measure is preferred to variance or standard deviation because it is not influenced by extreme values, which can occur in house price series in periods when transaction data are sparse.

[27] This figure could be susceptible to the variation in the number of transactions used to construct the index. A smaller number of transactions increases HPI estimation error, and thus measures of volatility.

[28] One new variable, the log of the number of repeat sales, is introduced for additional explanatory power.

with proximity to the CBD, controlling for the number of transactions in the source data. In general, the signs of parameter estimates are qualitatively similar to the appreciation results.

4. Discussion and Conclusions

This paper introduces a new panel of annual house price indices from 1975 through 2015. Prior to the introduction of this dataset, the lowest level of aggregation of publicly available, constant-quality indices was at the city level. This highly disaggregated panel will we hope unlock new and interesting research avenues that have remained closed due to lack of data availability.

We produce stylized facts related to house price appreciation gradients over a broad cross-section of cites over a long period of time. Overall, estimates suggest proximity to the center-city is a major factor explaining house price movements in the United States over the sample period, with house price gradients steepening in large cities between 1990 and 2015.

The Standard Urban Model (SUM) of Alonso (1964), Mills (1967) and Muth (1969) highlights the tradeoff between housing location and housing consumption, and how competition for scarce land leads to house price gradients within cities. Based on this model, there are many possible explanations for steepening gradients, including increases in traffic congestion, more extensive center-city amenities, lower center-city crime, or changing preferences, to name a few (see Glaeser, 2011 or Edlund, Machado, Sviatchi, 2015, for discussion). Within the SUM, migration of young, high income households—who have low relative preference for housing—to center-cities is expected to occur endogenously with the rotation in the price gradient. Previous findings of increases in center-city concentrations of households with relatively low housing demand are therefore corroborated by the price gradient results, including Black et al. (2002), Diamond (2016), and Couture and Handbury (2016). Additionally, because proximity to the CBD is correlated with land leverage, these results lend support to the Land Leverage Hypothesis of Bostic, Longhofer, and Redfearn (2007) and Davis and Heathcote (2007), who establish that high initial land value shares are associated with high future appreciation rates and volatility.

The final contribution of the house price panel presented in this paper is the establishment of facts related to the stationarity of house prices in different parts of the city. The data show ZIP code-level house price series to be non-stationary at higher rates near the CBD and, holding distance to the CBD constant, in large cities. This is consistent with the notion that the elasticity of housing supply is higher in suburban ares. In an area with a highly elastic housing supply, a permanent housing demand shock is first capitalized into prices, but over time as quantities adjust, prices return to pre-shock levels (see Glaeser, Gyourko, Morales, and Nathanson, 2014). In contrast, near the CBD, where buildable sites are less available and regulation is presumably more onerous, a permanent demand shock can outpace supply responses, leading to permanent price increases.

Combined, these findings demonstrate a small sample of the many potential applications of a panel of highly disaggregated house price indices. When a local house price measure is necessary, current practice is to use in its place either a geographically aggregated index, a value measure that confounds house prices and quantities, or a proprietary index that may lack coverage and is unavailable to many potential users. While value proxies may be appropriate in certain narrow circumstances, such as when the characteristics of housing units in an area are identical and unchanging, bias introduced by these measures may have major consequences. For instance, using a home value index in place of a price index in a fast-growing area may introduce an upward bias in perceptions of appreciation, and using a city-level index in place of a local index when estimating current loan-to-value ratios for mortgages may introduce substantial error. Many applications and research opportunities already exist for an accurate, long-horizon panel of geographically disaggregated house price data, and we hope these indices will unlock promising insights.

References

Alonso, W. (1964). *Location and land use: Toward a general theory of land rent.* Harvard University Press.

Bailey, M. J., Muth, R. F., and Nourse, H. O. (1963). A regression method for real estate price index construction. *Journal of the American Statistical Association*, 58(304):933–942.

Billings, S. B. (2015). Hedonic amenity valuation and housing renovations. *Real Estate Economics*, 43(3):652–682.

Black, D., Gates, G., Sanders, S., and Taylor, L. (2002). Why do gay men live in San Francisco? *Journal of Urban Economics*, 51(1):54–76.

Blackley, D. M. and Follain, J. R. (1987). Tests of locational equilibrium in the standard urban model. *Land Economics*, 63(1):46–61.

Bostic, R. W., Longhofer, S. D., and Redfearn, C. L. (2007). Land leverage: Decomposing home price dynamics. *Real Estate Economics*, 35(2):183–208.

Brueckner, J. K. (1987). The structure of urban equilibria: A unified treatment of the Muth - Mills model. *Handbook of Regional and Urban Economics*, 2:821–845.

Calhoun, C. A. (1996). OFHEO house price indexes: HPI technical description. Working paper series, Office of Federal Housing Enterprise Oversight.

Calhoun, C. A., Chinloy, P., and Megbolugbe, I. F. (1995). Temporal aggregation and house price index construction. *Journal of Housing Research*, 6(3):419–438.

Campbell, J. Y., Giglio, S., and Pathak, P. (2011). Forced sales and house prices. *American Economic Review*, 101(5):2108–2131.

Case, K. E. and Shiller, R. J. (1987). Prices of single-family homes since 1970: New indexes for four cities. *New England Economic Review*, Sep/Oct:45–56.

Case, K. E. and Shiller, R. J. (1989). The efficiency of the market for single-family homes. *American Economic Review*, 79(1):125–137.

Clapp, J. M., Giaccotto, C., and Tirtiroglu, D. (1991). Housing price indices based on all transactions compared to repeat subsamples. *Real Estate Economics*, 19(3):270–285.

Coulson, N. E. (1991). Really useful tests of the monocentric model. *Land Economics*, 67(3):299–307.

Couture, V. and Handbury, J. (2016). Urban revival in America, 2000 to 2010. *Mimeo*.

David A. Dickey, W. A. F. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica*, 49(4):1057–1072.

Davis, M. A. and Heathcote, J. (2007). The price and quantity of residential land in the United States. *Journal of Monetary Economics*, 54(8):2595–2620.

Davis, M. A., Oliner, S. D., Pinto, E., and Bokka, S. (2016). Residential Land Values in the Washington, DC Metro Area: New Insights from Big Data. Working paper series, Social Science Research Network.

Deaton, A. and Muellbauer, J. (1980). An almost ideal demand system. *American Economic Review*, 70(3):312–326.

Depken, C. A., Hollans, H., and Swidler, S. (2009). An empirical analysis of residential property flipping. *Journal of Real Estate Finance and Economics*, 39(3):248–263.

Depken, C. A., Hollans, H., and Swidler, S. (2015). A low cost methodology for correcting the distressed sales bias in a downward spiraling housing market. *Journal of Real Estate Research*, 37(1):151–172.

Diamond, R. (2016). The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980-2000. *American Economic Review*, 106(3):479–524.

Doerner, W. M. and Leventis, A. (2015). Distressed sales and the FHFA house price index. *Journal of Housing Research*, 24(2):127–146.

Edlund, L., Machado, C., and Sviatchi, M. (2015). Bright minds, big rent: Gentrification and the rising returns to skill. Working Paper 21729, National Bureau of Economic Research.

Glaeser, E. (2011). *Triumph of the city: How our greatest invention makes us richer, smarter, greener, healthier, and happier*. Penguin.

Glaeser, E. L., Gyourko, J., Morales, E., and Nathanson, C. G. (2014). Housing dynamics: An urban approach. *Journal of Urban Economics*, 81:45–56.

Glaeser, E. L., Gyourko, J., and Saiz, A. (2008). Housing supply and housing bubbles. *Journal of Urban Economics*, 64(2):198–217.

Guerrieri, V., Hartley, D., and Hurst, E. (2013). Endogenous gentrification and housing price dynamics. *Journal of Public Economics*, 100:45 – 60.

Harding, J. P., Rosenblatt, E., and Yao, V. W. (2009). The contagion effect of foreclosed properties. *Journal of Urban Economics*, 66(3):165–178.

Harding, J. P., Rosenthal, S. S., and Sirmans, C. F. (2007). Depreciation of housing capital, maintenance, and house price inflation: Estimates from a repeat sales model. *Journal of Urban Economics*, 61(2):193–217.

Immergluck, D. and Smith, G. (2006). The impact of single-family mortgage foreclosures on property values. *Housing Policy Debate*, 17(1):57–79.

Leventis, A. (2009). The impact of distressed sales on repeat-transactions house price indexes. *Federal Housing Finance Agency Working Paper*.

McMillen, D. and Thorsnes, P. (2006). Housing renovations and the quantile repeat-sales price index. *Real Estate Economics*, 34(4):567–584.

McMillen, D. P. and Smith, S. C. (2003). The number of subcenters in large urban areas. *Journal of Urban Economics*, 53(3):321–338.

Mian, A. and Sufi, A. (2009). The consequences of mortgage credit expansion: Evidence from the u.s. mortgage default crisis. *The Quarterly Journal of Economics*, 124(4):1449–1496.

Mieszkowski, P. and Mills, E. S. (1993). The causes of metropolitan suburbanization. *The Journal of Economic Perspectives*, 7(3):135–147.

Miller, C. C. (2014). Where young college graduates are choosing to live. *New York Times*. URL: http://www.nytimes.com/2014/10/20/upshot/where-young-college-graduates-are-choosing-to-live.html.

Mills, E. S. (1967). An aggregative model of resource allocation in a metropolitan area. *American Economic Review*, 57(2):197–210.

Moretti, E. (2012). *The new geography of jobs*. Houghton Mifflin Harcourt.

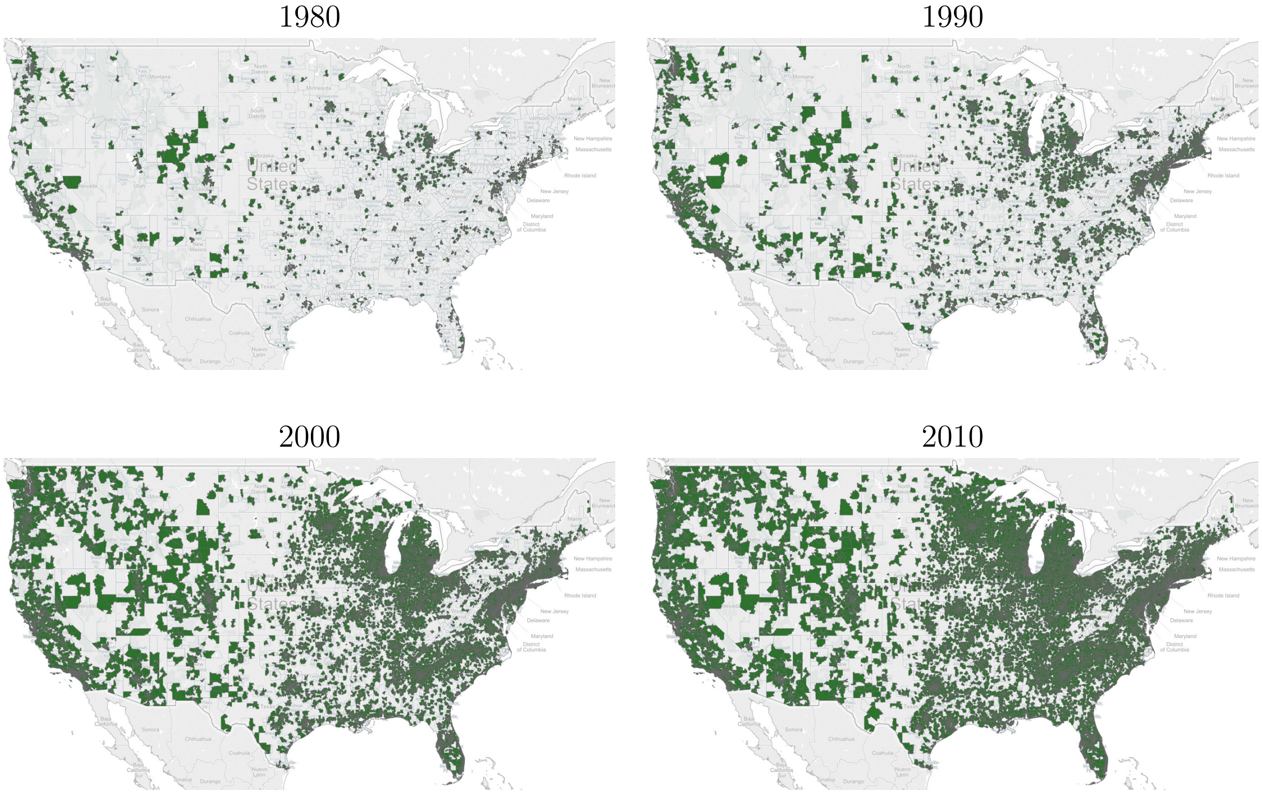
Muth, R. F. (1969). *Cities and Housing; the Spatial Pattern of Urban Residential Land Use.* University of Chicago Press.

Saiz, A. (2010). The geographic determinants of housing supply. *The Quarterly Journal of Economics*, 125(3):1253–1296.

Voith, R. (1991). Transportation, sorting and house values. *Real Estate Economics*, 19(2):117–137.

Yinger, J. (1979). Estimating the relationship between location and the price of housing. *Journal of Regional Science*, 19(3):271–286.

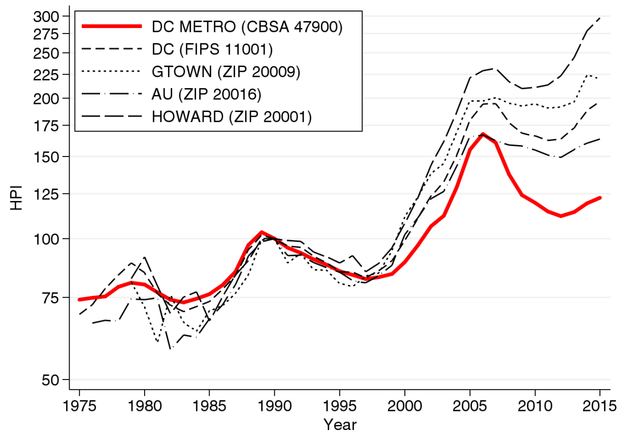
Figure 1: Coverage of 5-Digit ZIP Code HPIs



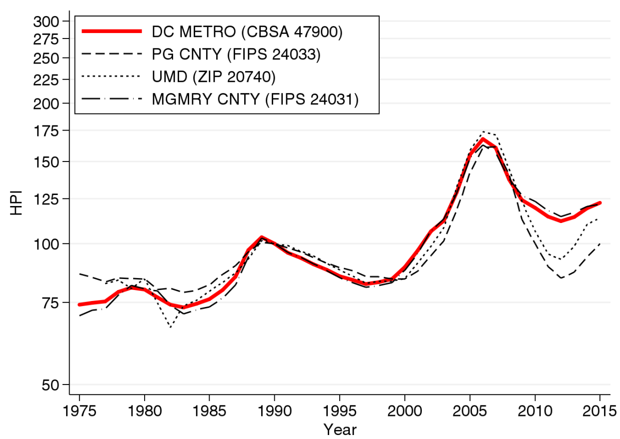
Note: Alaska and Hawaii are not shown here but both states have coverage.

Figure 2: House Price Indices: Select Washington, D.C. MSA Submarket Indices

(a) District of Columbia



(b) Maryland



(c) Virginia

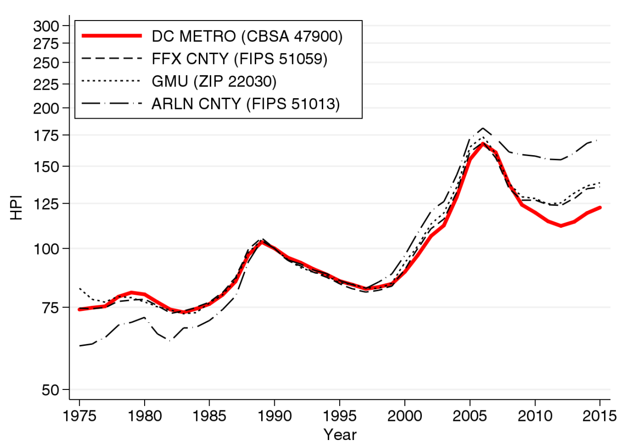
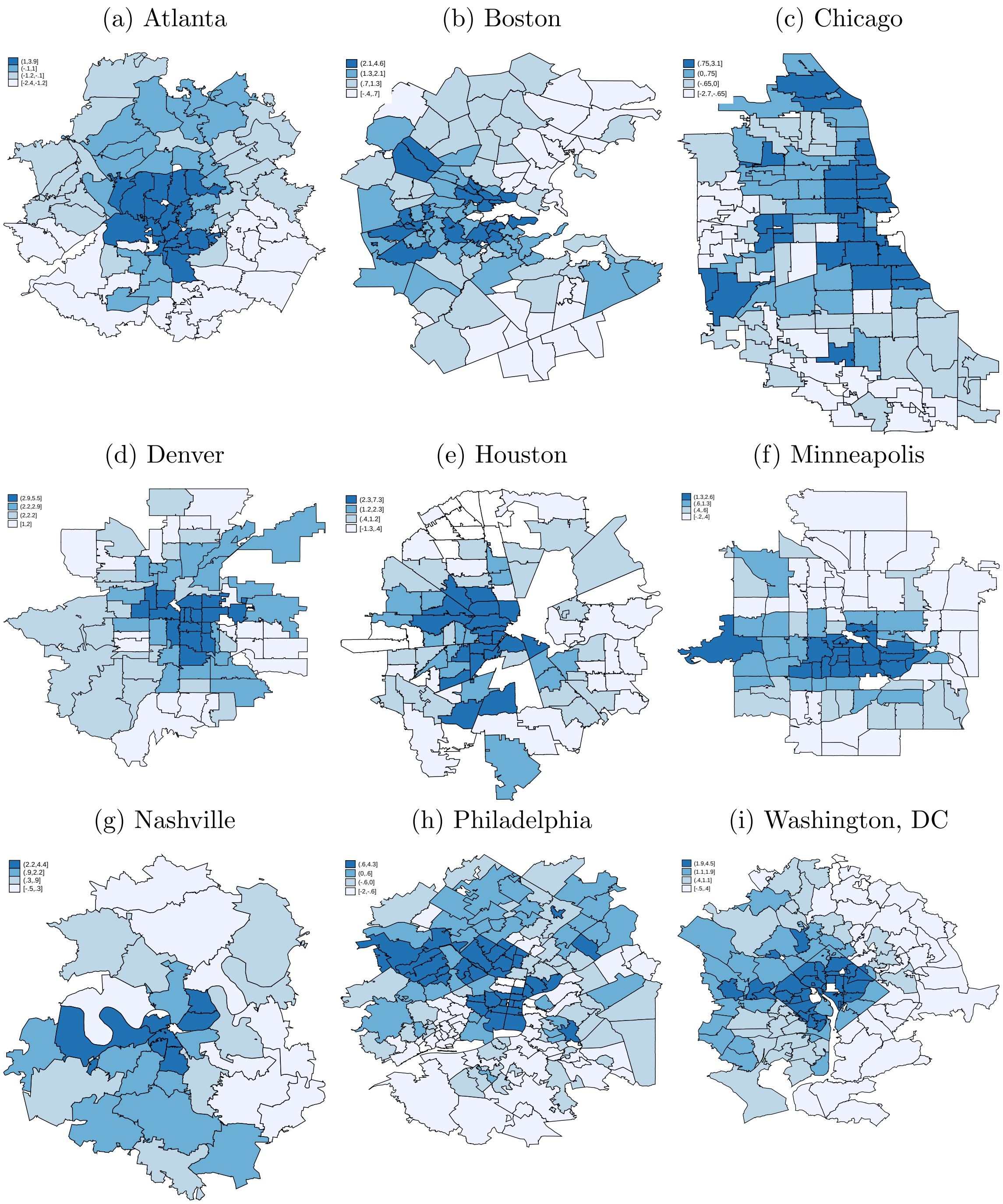
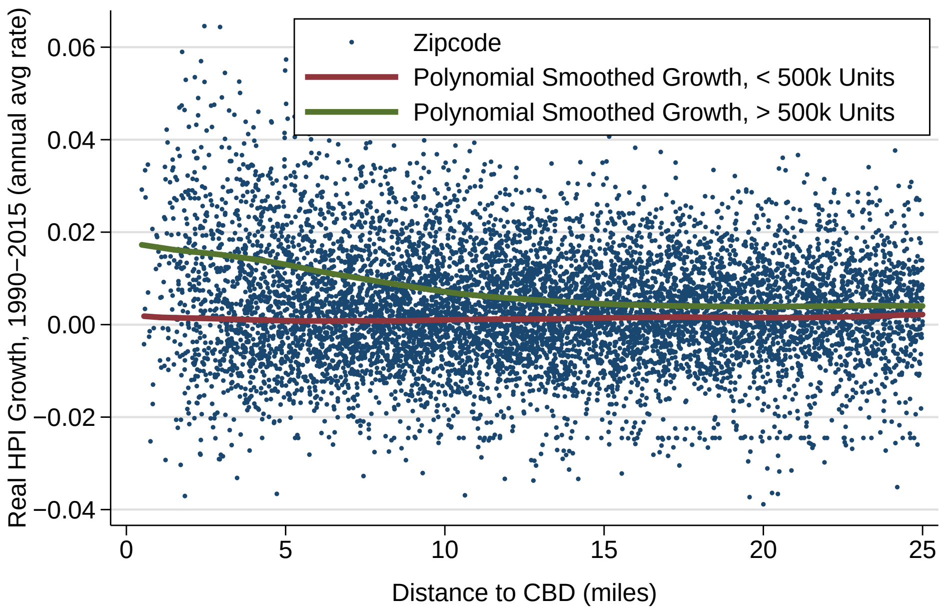


Figure 3: Annual Average Real Appreciation in Select Cities, 1990 to 2015



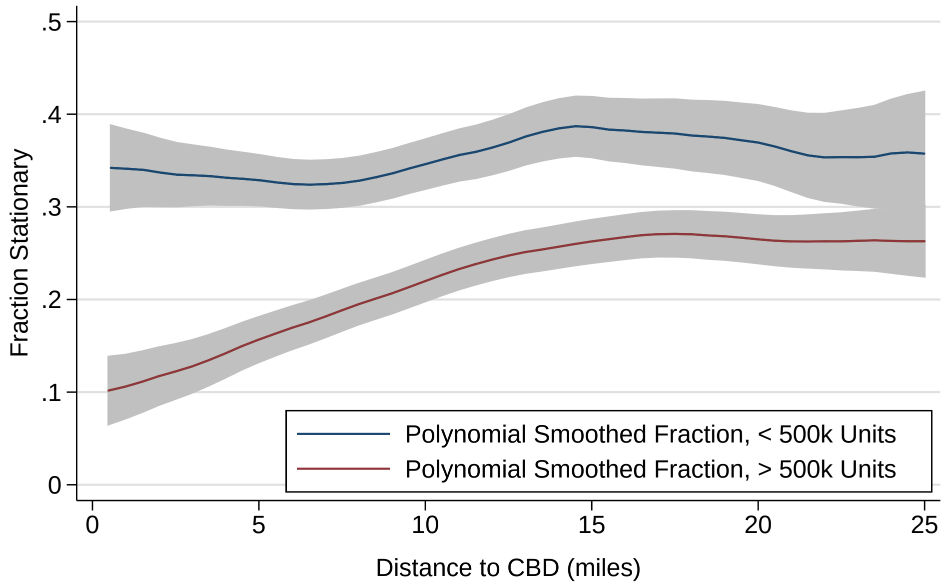
Note: In instances when ZIP code coverage begins after 1990, we calculate annual average appreciation over a shorter period. Maps only include ZIP codes within 20 miles of the CBD.

Figure 4: Spatial Distribution of Annual Average Appreciation in U.S. Cities, 1990-2015



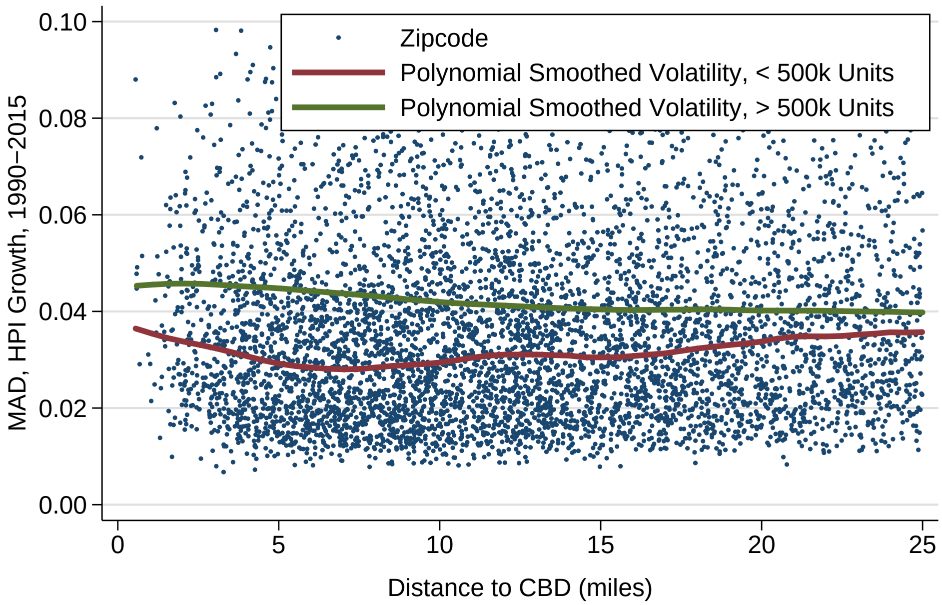
Note: 24 zipcodes have annual average appreciation rates greater than 6% or lower than −4%. These are omitted from the scatterplot but are reflected in the curves.

Figure 5: Spatial Distribution of House Price Stationarity



Note: Figure depicts the smoothed fraction of all zip codes at a given distance from the CBD where the null of a unit root is rejected at a 90% level using Augmented Dicky−Fuller tests with 2 lags, a constant term, and no trend.

Figure 6: Spatial Distribution of Appreciation Volatility in U.S. Cities, 1990-2015



Note: 28 zipcodes have MADs greater than 10%. These are omitted from the scatterplot but are reflected in the lowess curve. MAD stands for Median Absolute Deviation.

Table 1: Publicly Available House Price Indices in the United States

Index	Geography	Count	Frequency	Start
S&P/Case-Shiller	CBSA	20	Monthly	1987
FHFA	CBSA	401	Quarterly	1975
Freddie Mac	CBSA	367	Quarterly	1975
FHFA	ZIP3	885	Quarterly	1995
Zillow (Value)	ZIP5	12,988	Monthly	1996
This Paper:	CBSA	914	Annual	1975
	County	2,742	Annual	1975
	ZIP3	879	Annual	1975
	ZIP5	17,936	Annual	1975

Note: "CBSA" stands for Core Based Statistical Area, which includes both MSAs and MicroSAs. "ZIP3" refers to an area defined by the first three numbers in a ZIP code while "ZIP5" are smaller areas within each ZIP3 area, denoted by a 5-digit identifier.

Table 2: House Price Index Counts

		Start Date					
Sample	Number	Pre 1980	1981–1985	1986–1990	1991–1995	1996–2000	Post 2000
ZIP in single CBSA	12,965	4,140	1,496	2,383	2,236	1,150	1,560
Center-City (<5 Miles from CBD)	1,474	596	180	312	202	79	105
Mid-City (5-15 Miles from CBD)	4,306	1,730	508	760	655	277	376
Suburbs (15-25 Miles from CBD)	3,274	964	382	591	591	319	427
Exurbs (>25 Miles from CBD)	3,895	845	424	719	785	472	650
ZIP in multiple CBSAs	1,445	11	9	70	300	383	672
ZIP not in a CBSA	3,526	296	145	450	966	732	937
Total ZIP codes	17,936	4,447	1,650	2,903	3,502	2,265	3,169
ZIP3s	879	717	44	71	23	13	11
Counties in MSA	1,142	681	83	118	146	69	45
Counties in MicroSA	614	188	76	158	124	42	26
Counties not in CBSA	986	22	16	112	293	277	266
Total Counties	2,742	891	175	388	563	388	337
MSAs	381	351	22	5	3	0	0
MicroSAs	533	192	75	144	103	14	5
Total CBSAs	914	543	97	149	106	14	5
State-Less-CBSAs	45	27	11	5	2	0	0

Note: An index is calculated for any area with more than 100 repeat sales over the sample, with an index start year once a threshold of 25 half-pairs (HP) is met. We define the term "half-pairs" as a transaction occurring in either t or τ. For instance, if a home sells in 1978 and 1990, the transaction pair will contribute one half-pair to the count for 1978 and one half-pair for 1990. If the home sells a third time in 1998, the home is involved in two transaction pairs, contributing one half-pair in each of 1978 and 1998, and two half-pairs in 1990. The amount of unique information in the HP count is thus between HP and HP/2.

Table 3: Appreciation Regressions, Large Cities (>500k Units)

	LHS Variable: Δ HPI (log)						
Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Covariate Set	Empty	Transportation	Structure	Labor Force	Earnings	Housing Demand	All
Distance from CBD (log, miles)	-0.152***	-0.0930***	-0.140***	-0.131***	-0.169***	-0.108***	-0.0725***
	[0.0203]	[0.0227]	[0.0205]	[0.0220]	[0.0212]	[0.0200]	[0.0195]
% Driving		-0.542***					-0.649***
		[0.0913]					[0.0869]
Commute Time (minutes)		-0.00168					0.000749
		[0.00363]					[0.00314]
Rooms			0.0302**				0.0406***
			[0.0129]				[0.0111]
Single-family Det.			-0.0921**				0.231***
			[0.0446]				[0.0466]
% 20+ Unit Structure			0.266***				-0.0609
			[0.0800]				[0.0593]
% Mobile Home			-0.117				0.195***
			[0.0817]				[0.0698]
% in Labor Force				-0.654***			-0.571***
				[0.179]			[0.118]
% Receiving Ret. Inc.				-0.916***			-1.653***
				[0.239]			[0.202]
Median Income (log)					0.105**		0.0346
					[0.0457]		[0.0572]
% Receiving Public Asst.					0.0448		-0.191
					[0.300]		[0.241]
Unemployment Rate					0.716		-0.0163
					[0.465]		[0.628]
% with Children						-0.388***	-0.843***
						[0.0765]	[0.163]
% with Married Head						-0.123*	-0.199
						[0.0693]	[0.121]
Constant	0.404***	0.696***	0.233***	0.942***	-0.705	0.495***	0.993*
	[0.0519]	[0.125]	[0.0795]	[0.146]	[0.482]	[0.0488]	[0.565]
Observations	2,742	2,742	2,742	2,742	2,742	2,742	2,742
CBSAs	38	38	38	38	38	38	38
R-squared	0.167	0.237	0.195	0.205	0.177	0.203	0.364

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The left-hand side variable is the log difference in real house prices between 1990 and 2015, de-meaned using the CBSA index. Standard errors are adjusted based on clustering at the CBSA level.

Table 4: Volatility Regressions, Large Cities (>500k Units)

	LHS Variable: Median Absolute Deviation, Δ HPI (log)						
Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Covariate Set	Empty	Transportation	Structure	Labor Force	Earnings	Housing Demand	All
Repeat Sales (log)	-0.00720***	-0.00714***	-0.00710***	-0.00674***	-0.00575***	-0.00663***	-0.00674***
	[0.00144]	[0.00140]	[0.00146]	[0.00138]	[0.00161]	[0.00145]	[0.00156]
Distance from CBD (log, miles)	-0.00450***	-0.00310**	-0.00337***	-0.00381***	-0.00172	-0.00243*	-0.00297*
	[0.000981]	[0.00136]	[0.00113]	[0.00119]	[0.00119]	[0.00124]	[0.00163]
% Driving		-0.00766					0.0142
		[0.00612]					[0.00875]
Commute Time (minutes)		-0.00028					-0.000275*
		[0.000188]					[0.000148]
Rooms			-0.00443***				-0.00205*
			[0.000974]				[0.00119]
Single-family Det.			0.0058				0.00163
			[0.00419]				[0.00403]
% 20+ Unit Structure			-0.0155**				0.00285
			[0.00718]				[0.00923]
% Mobile Home			-0.00136				0.00553
			[0.00753]				[0.00615]
% in Labor Force				-0.0254*			-0.00351
				[0.0132]			[0.0148]
% Receiving Ret. Inc.				-0.0159			0.0532**
				[0.0181]			[0.0235]
Median Income (log)					-0.00121		0.00913*
					[0.00339]		[0.00467]
% Receiving Public Asst.					0.018		0.000638
					[0.0276]		[0.0359]
Unemployment Rate					0.132***		0.127***
					[0.0406]		[0.0334]
% with Children						0.0474***	0.0577***
						[0.00780]	[0.0130]
% with Married Head						-0.0369***	-0.0391***
						[0.00594]	[0.0126]
Constant	0.0717***	0.0802***	0.0912***	0.0861***	0.0575**	0.0668***	-0.0354
	[0.0118]	[0.0125]	[0.00962]	[0.0167]	[0.0269]	[0.0119]	[0.0448]
Observations	2,742	2,742	2,742	2,742	2,742	2,742	2,742
CBSAs	38	38	38	38	38	38	38
R-squared	0.089	0.091	0.106	0.093	0.123	0.115	0.14

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The left-hand side variable is the ZIP code-specific median absolute deviation of the annual appreciation rate between 1990 and 2015, de-meaned using the CBSA index. Standard errors are adjusted based on clustering at the CBSA level.

Appendix

Table A-1: Appreciation Regressions, All Cities

	LHS Variable: Δ HPI (log)						
Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Covariate Set	Empty	Transportation	Structure	Labor Force	Earnings	Housing Demand	All
Distance from CBD (log, miles)	-0.0726***	-0.0437***	-0.0614***	-0.0609***	-0.0771***	-0.0388***	-0.0366***
	[0.0131]	[0.0111]	[0.0121]	[0.0129]	[0.0138]	[0.0104]	[0.0105]
% Driving		-0.541***					-0.642***
		[0.0575]					[0.0594]
Commute Time (minutes)		-0.00078					0.000215
		[0.00156]					[0.00164]
Rooms			0.0173*				0.0285***
			[0.00939]				[0.00847]
Single-family Det.			-0.0661*				0.208***
			[0.0359]				[0.0334]
% 20+ Unit Structure			0.305***				0.0242
			[0.0655]				[0.0547]
% Mobile Home			-0.0891*				0.132***
			[0.0468]				[0.0506]
% in Labor Force				-0.558***			-0.407***
				[0.104]			[0.0735]
% Receiving Ret. Inc.				-0.744***			-1.143***
				[0.137]			[0.117]
Median Income (log)					0.0428		0.0273
					[0.0292]		[0.0357]
% Receiving Public Asst.					0.0283		-0.360**
					[0.179]		[0.154]
Unemployment Rate					0.373		0.268
					[0.240]		[0.325]
% with Children						-0.368***	-0.728***
						[0.0596]	[0.103]
% with Married Head						-0.137***	-0.0629
						[0.0446]	[0.0851]
Constant	0.188***	0.548***	0.0859	0.659***	-0.271	0.315***	0.756**
	[0.0327]	[0.0702]	[0.0532]	[0.0924]	[0.303]	[0.0381]	[0.349]
Observations	5,163	5,163	5,163	5,163	5,163	5,163	5,163
CBSAs	423	423	423	423	423	423	423
R-squared	0.061	0.154	0.098	0.098	0.064	0.115	0.266

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The left-hand side variable is the log difference in real house prices between 1990 and 2015, de-meaned using the CBSA index. Standard errors are adjusted based on clustering at the CBSA level.

Table A-2: Volatility Regressions, All Cities

	LHS Variable: Median Absolute Deviation, Δ HPI (log)						
Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Covariate Set	Empty	Transportation	Structure	Labor Force	Earnings	Housing Demand	All
Repeat Sales (log)	-0.00527***	-0.00479***	-0.00490***	-0.00476***	-0.00392***	-0.00481***	-0.00439***
	[0.000786]	[0.000801]	[0.000865]	[0.000770]	[0.000929]	[0.000809]	[0.000950]
Distance from CBD (log, miles)	-0.00476***	-0.00342***	-0.00305***	-0.00419***	-0.00284***	-0.00222***	-0.00213***
	[0.000608]	[0.000706]	[0.000691]	[0.000654]	[0.000694]	[0.000722]	[0.000768]
% Driving		-0.0193***					0.00483
		[0.00497]					[0.00543]
Commute Time (minutes)		-0.00016					-3.84E-05
		[0.000132]					[0.000124]
Rooms			-0.00441***				-0.00149**
			[0.000727]				[0.000722]
Single-family Det.			0.000838				0.00351
			[0.00281]				[0.00273]
% 20+ Unit Structure			-0.0123**				0.00184
			[0.00603]				[0.00808]
% Mobile Home			-0.0177***				-0.00641
			[0.00443]				[0.00475]
% in Labor Force				-0.0302***			-0.00416
				[0.00796]			[0.00966]
% Receiving Ret. Inc.				-0.0276**			0.0287**
				[0.0111]			[0.0117]
Median Income (log)					-0.00141		0.00383
					[0.00189]		[0.00325]
% Receiving Public Asst.					0.0402**		0.018
					[0.0166]		[0.0220]
Unemployment Rate					0.0880***		0.0617***
					[0.0241]		[0.0232]
% with Children						0.0393***	0.0405***
						[0.00512]	[0.00863]
% with Married Head						-0.0429***	-0.0385***
						[0.00405]	[0.00824]
Constant	0.0546***	0.0658***	0.0737***	0.0741***	0.0470***	0.0564***	0.00717
	[0.00661]	[0.00730]	[0.00561]	[0.0101]	[0.0157]	[0.00670]	[0.0250]
Observations	5,163	5,163	5,163	5,163	5,163	5,163	5,163
CBSAs	423	423	423	423	423	423	423
R-squared	0.062	0.068	0.080	0.068	0.096	0.091	0.104

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The left-hand side variable is the ZIP code-specific median absolute deviation of the annual appreciation rate between 1990 and 2015, de-meaned using the CBSA index. Standard errors are adjusted based on clustering at the CBSA level.

Table A-3: Appreciation Regressions, Small Cities (<500k Units)

	LHS Variable: Δ HPI (log)						
Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Covariate Set	Empty	Transportation	Structure	Labor Force	Earnings	Housing Demand	All
Distance from CBD (log, miles)	-0.0104	-0.00498	-0.00934	-0.00737	-0.0134	-0.00235	0.000434
	[0.0107]	[0.0104]	[0.0107]	[0.0104]	[0.0104]	[0.00921]	[0.00869]
% Driving		-0.174**					-0.329***
		[0.0682]					[0.0584]
Commute Time (minutes)		-0.00082					0.000318
		[0.00159]					[0.00108]
Rooms			0.00451				0.0223***
			[0.00904]				[0.00772]
Single-family Det.			0.104***				0.226***
			[0.0384]				[0.0482]
% 20+ Unit Structure			0.338**				0.193
			[0.132]				[0.123]
% Mobile Home			-0.0129				0.07
			[0.0496]				[0.0584]
% in Labor Force				-0.256***			-0.277***
				[0.0654]			[0.0852]
% Receiving Ret. Inc.				-0.253***			-0.656***
				[0.0805]			[0.0908]
Median Income (log)					-0.0309		0.00909
					[0.0322]		[0.0277]
% Receiving Public Asst.					-0.514***		-0.519***
					[0.176]		[0.167]
Unemployment Rate					0.102		0.316
					[0.198]		[0.196]
% with Children						-0.385***	-0.549***
						[0.0615]	[0.0863]
% with Married Head						0.105	-0.0267
						[0.0691]	[0.0824]
Constant	0.0253	0.167**	-0.0863	0.232***	0.376	0.0807**	0.389
	[0.0243]	[0.0716]	[0.0563]	[0.0681]	[0.339]	[0.0402]	[0.244]
Observations	2,421	2,421	2,421	2,421	2,421	2,421	2,421
CBSAs	385	385	385	385	385	385	385
R-squared	0.003	0.011	0.021	0.015	0.017	0.044	0.138

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The left-hand side variable is the log difference in real house prices between 1990 and 2015, de-meaned using the CBSA index. Standard errors are adjusted based on clustering at the CBSA level.

Table A-4: Volatility Regressions, Small Cities (<500k Units)

	LHS Variable: Median Absolute Deviation, Δ HPI (log)						
Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Covariate Set	Empty	Transportation	Structure	Labor Force	Earnings	Housing Demand	All
Repeat Sales (log)	-0.00330***	-0.00273***	-0.00269***	-0.00280***	-0.00189***	-0.00291***	-0.00198***
	[0.000534]	[0.000519]	[0.000526]	[0.000540]	[0.000493]	[0.000518]	[0.000514]
Distance from CBD (log, miles)	-0.00569***	-0.00436***	-0.00431***	-0.00530***	-0.00413***	-0.00301***	-0.00320***
	[0.000824]	[0.000872]	[0.000971]	[0.000804]	[0.000798]	[0.000975]	[0.000980]
% Driving		-0.0388***					-0.0161**
		[0.00849]					[0.00784]
Commute Time (minutes)		-0.000240**					2.83E-05
		[0.000116]					[0.000112]
Rooms			-0.00534***				-0.00087
			[0.000873]				[0.000954]
Single-family Det.			-7.66E-05				0.00421
			[0.00306]				[0.00401]
% 20+ Unit Structure			-0.0169**				0.000525
			[0.00858]				[0.00925]
% Mobile Home			-0.0160***				-0.00883
			[0.00478]				[0.00613]
% in Labor Force				-0.0333***			0.00909
				[0.00620]			[0.00976]
% Receiving Ret. Inc.				-0.0254***			0.0207**
				[0.00824]			[0.0104]
Median Income (log)					-0.00645***		-0.00491
					[0.00204]		[0.00305]
% Receiving Public Asst.					0.0469***		0.0294*
					[0.0140]		[0.0155]
Unemployment Rate					0.0295		0.0242
					[0.0233]		[0.0232]
% with Children						0.0267***	0.0144
						[0.00593]	[0.00900]
% with Married Head						-0.0428***	-0.0183*
						[0.00641]	[0.0104]
Constant	0.0389***	0.0669***	0.0632***	0.0606***	0.0869***	0.0463***	0.0822***
	[0.00525]	[0.00872]	[0.00643]	[0.00723]	[0.0213]	[0.00531]	[0.0255]
Observations	2,421	2,421	2,421	2,421	2,421	2,421	2,421
CBSAs	385	385	385	385	385	385	385
R-squared	0.048	0.065	0.069	0.056	0.082	0.074	0.089

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The left-hand side variable is the ZIP code-specific median absolute deviation of the annual appreciation rate between 1990 and 2015, de-meaned using the CBSA index. Standard errors are adjusted based on clustering at the CBSA level.