FHFA STAFF WORKING PAPER SERIES

A Flexible Method of House Price Index Construction using Repeat-Sales Aggregates

Justin Contat

and

William D. Larson*

June 2022
(Initial Version: March 2021)

Working Paper 21-01

FEDERAL HOUSING FINANCE AGENCY
Division of Research and Statistics
400 7th Street SW
Washington, DC 20219, USA

Working Papers prepared by staff of the Federal Housing Finance Agency (FHFA) are preliminary products circulated to stimulate discussion and critical comment. The analysis and conclusions are those of the authors alone, and should not be represented or interpreted as conveying an official FHFA position, policy, analysis, opinion, or endorsement. Any errors or omissions are the sole responsibility of the authors. References to FHFA Working Papers (other than acknowledgment) should be cleared with the authors to protect the tentative character of these papers.

*Please address correspondence to William Larson (william.larson@fhfa.gov). The authors thank seminar participants at UC Irvine and FHFA, Bob Avery, Alex Bogin, William Doerner, Lynn Fisher, and Marshall Reinsdorf for conversation, encouragement, and detailed comments. An earlier version of this working paper was released under the title “Transaction Composition and House Price Index Measurement: Evidence from a Repeat-Sales Aggregation Index.”

A Flexible Method of House Price Index Construction using Repeat-Sales Aggregates

Justin Contat, William D. Larson

FHFA Staff Working Paper 21-01

June 2022
(Initial Version: March 2021)

Abstract

The major issue which we address in this paper is the one-size-fits-all nature of the typical city-level house price index. In this vein, we make two contributions. First, we develop a new algorithm to ensure feasible estimation of geographically granular repeat-sales house price indices in cases of low transactions counts. This facilitates the estimation of a balanced panel of 63,122 Census tract-level repeat-sales house price indices (2010 definitions) at an annual frequency between 1989 and 2021, which we release alongside this paper. Second, we use these indices to estimate city-level house price indices that are robust to heterogeneous submarket appreciation and non-random sampling, two issues that confound classic approaches. Numerical simulations show this algorithm uncovers population indices even when house prices, quantities, and transaction sampling vary across locations and over time. This approach can be used in a flexible manner to calculate canonical price indices such as Lowe and Laspeyres, and more tailored summary indices on a variety of topics, including collateral valuation, climate risk assessment, or tracking changes to minority housing wealth over time.

Keywords: sampling, standard urban model, index number theory

JEL Classification: C43, R30, R32

Justin Contat
Federal Housing Finance Agency
Division of Research and Statistics
400 7th Street SW
Washington, DC 20219, USA
justin.contat@fhfa.gov

William D. Larson
Federal Housing Finance Agency
Division of Research and Statistics
400 7th Street SW
Washington, DC 20219, USA
william.larson@fhfa.gov

1 Introduction

Researchers, policymakers, and financial analysts frequently need estimates of the appreciation rate of houses. Some common uses for such estimates include valuing mortgage collateral, measuring changes to housing wealth, or tracking the price of housing services to estimate the elasticity of housing supply (e.g. Deng et al., 2000; Skinner, 1989; Green et al., 2005, respectively). Repeat-sales house price indices are one of the primary means by which appreciation rates are estimated, and these indices are, with some notable exceptions, typically available at the city (i.e. county or metropolitan statistical area) level.[1]

The major issue which we address in this paper is the one-size-fits-all nature of the typical city-level index estimation exercise. To estimate a city-level repeat-sales index, pairs of same-housing-unit transactions are used to calculate unit-level appreciation rates, with the index estimated by pooling all such transaction pairs within a city. As such, the index is subject to issues of within-city transaction composition and non-representative sampling (Gatzlaff and Haurin, 1997, 1998; Malone and Redfearn, 2020).

In this paper, we make two contributions. First, we construct a balanced panel of annual house price indices for single-family homes in 63,122 Census tracts between 1989 and 2021.[2,3] These Census tracts are mutually exclusive and exhaustive submarkets within 581 core-based statistical areas. This is the first study to feasibly estimate appreciation rates for so many Census tracts going so far back in time. To do so, we have constructed what we believe to be the largest database of home purchase transactions to date, including all purchase-money mortgages purchased or securitized by Fannie Mae or Freddie Mac, the Federal Housing Administration, and all purchases captured in publicly available county recorder rolls provided by CoreLogic. This results in over 63 million same-unit purchase

[1] The repeat-sales house price index has been one of the primary house price index construction methods since Bailey, Muth, and Nourse (1963), with seminal contributions by Case and Shiller (1989) and Calhoun (1996). The Federal Housing Finance Agency (FHFA) and Standard & Poor/Case-Shiller use repeat sales methods to calculate their flagship house price indices. Within-city indices have recently become available via Bogin, Doerner, and Larson (2019a), Zillow, and CoreLogic. These submarket indices have some limitations, however, because they are typically not exhaustive of all units in a city (i.e. missing neighborhoods), and/or do not span long time horizons (begin in the 2000s).

[2] This dataset is available at `https://www.fhfa.gov/PolicyProgramsResearch/Research/Pages/wp2101.aspx`. For a document addressing frequently-asked questions about these indices, see `https://www.fhfa.gov/PolicyProgramsResearch/Research/PaperDocuments/wp2101-faqs.pdf`.

[3] For comparison with other indices, the Bogin, Doerner, and Larson (2019a) dataset includes over 53,000 Census tracts, but only 23,000 go back to 1989. Zillow does not make tract-level values available.

pairs for the U.S. from 1975 through 2021 (transactions prior to 1989 can be sparse and serve as controls). To handle cases where traditional estimation is infeasible due to the small number of transactions in certain periods, we also develop a new algorithm to estimate tract-level appreciation rates.[4] This algorithm exploits an assumption regarding local homogeneity in appreciation rates to create fluid, period-specific aggregations of Census tracts, allowing feasible estimation of indices for all tracts in all time periods.

Our second contribution is to address the potential issue of non-representative sampling in city-level indices, while simultaneously introducing a method of constructing city-level indices that can be tailored to a particular use-case. Our insight is to treat a city as a basket of submarkets, similar to how a classic price index is created out of a basket of goods and services. In this manner, a city-level index can be calculated by aggregating mutually exclusive and exhaustive submarket indices using weights. The choice of weights is crucial, as it defines the target index: while sampling weights give a classic transaction-based city-level price index, the use of housing stock weights make the index representative of all existing homes in the city, and aggregate value weights result in the classic geometric Laspeyres formulation. This flexible approach can be used to generate novel index formulations too. For instance, using tract-level mortgage loan balances it is possible to calculate a city-level index that tracks the average change in collateral value for loans in the city. To track appreciation rates of minority-owned homes, the city-level index could be aggregated based on tract-level minority population shares. The point we are trying to make, however, is not to recommend any particular index formulation. Rather, this flexible approach, which we term a *repeat sales aggregation index* (RSAI), can be used to estimate *any* target city-level index rather than being constrained to the traditional sample-dependent rendition of the repeat-sales index.

There are striking differences in city-level house price indices calculated using different weights, both across time, and within periods for different index types. A necessary condition of differential city-level appreciation rates is non-constant appreciation rates in submarkets within the city (Malone and Redfearn, 2020). Accordingly, city-level index differences

[4] To validate this approach, we construct a theoretical monocentric city model in the vein of of Alonso (1964), Mills (1967), and Muth (1969). We then sample housing transactions from this city according to various decision rules, calculate RSAI and classic repeat-sales indices, and verify the RSAI recovers various target population price indices for the city, even when the prices, quantities, locations, and sampling of housing within the city change over time.

arise most commonly in large cities with sustained appreciation gradients, (Bogin, Doerner, and Larson, 2019a). There are five distinct eras in the United States between 1989 and 2021 where appreciation typically diverges and compound over time, 1989-2001, 2001-2005, 2005-2013, 2013-2019, and 2019-2021. For example of variation in differences over time, transaction sample-weighted indices underestimate appreciation in large cities compared to the Laspeyres (value weight) formulation by an average of about 0.20% per year between 1989 and 2001, and then by 0.35% per year between 2005 and 2013. This occurs because homes in tracts with higher-than-average appreciation rates are under-sampled relative to their proportion of home value in the city. This reverses in the years after 2013 and accelerates in the COVID-19 era, with sample-weighted indices overestimating appreciation by 0.40% per year through 2019, and 0.60% per year in 2020 and 2021. The 2013-2019 period has dramatic differences between weighting schemes, with sample-based indices showing an underestimate of appreciation for those calculated using the share of the non-white population (-1.20% per year) and housing stock (-0.25%), versus those calculated using housing value weights (0.40%), the share of the college-educated population (0.45%), and the share of Enterprise loan balances (0.52%).

Perhaps surprisingly, in the COVID-19 era, 2020 had only slight changes relative to prior years in terms of both average price gradients and sampling proportionality. However, 2021 faced a significant rotation of the appreciation gradient towards suburban locations, yet homes near the centers of cities were more-than-proportionally sampled, thus leading to an understatement of sample-weighted indices compared to those representing the value of the housing stock. In general, "drift" in index differences is often correlated within small time (5-10 year) windows, but these windows in turn have unpredictable correlations across other time windows in the same city. Thus, in order to protect against compounding differences, we recommend tailoring the weighting scheme used to the use case at hand, especially in large cities.

Our work is related to past efforts to address sampling variation in house price index construction. To address this concern in hedonic models, Hill and Melser (2008) develop a technique that estimates a hedonic model and imputes value estimates to untransacted units. However, a hedonic model requires a requires a rich set of housing unit characteristics for the entire housing stock, assumptions about independence of unobservable attributes, and spatial constancy of implicit prices. The issue of spatially varying implicit prices has been addressed by Carrillo and Yezer (2021), and necessitates the estimation of a large number of parameters leading to a dimensionality problem (e.g., the value of square footage is different for urban and rural environments). Gatzlaff and Haurin (1997, 1998) analyze selection bias introduced by only observing prices of transacted homes, first for hedonic and then repeat sales indices. Our work is perhaps most similar to Malone and Redfearn (2020, 2021), who construct city-level indices using a weighted average of submarket housing stock shares, termed "local pooling" indices in these papers. Our method encompasses theirs and extends it, in spirit, to other aggregation methods. We are also able to extend the aggregation approach to more cities to uncover general patterns in index divergence. Finally, their method of index construction is based on hedonic characteristics and thus requires large amounts of data for each area in question; ours is based on repeat-sales, requiring only transaction records.

The remainder of this paper is as follows. First, we briefly review repeat sales methodology, highlighting both decomposition along submarket lines and also the connection between classical index theory and repeat sales estimation. After that, we describe the RSAI and its implementation, pointing out the relationship between sampling schemes, submarket weights, and target indices. Next, we simulate a city and show that this new method recovers various target indices in a city under different non-proportional sampling schemes. We then take the estimator to real-world data and estimate appreciation rates for 581 core-based statistical areas (CBSA) in the United States between 1989 and 2021. Finally, we provide a brief discussion and conclude.

2 The Repeat Sales Aggregation Index

In this section, we begin with the classic Bailey, Muth, and Nourse (1963) (BMN) repeat-sales house price index and review two possible methods of estimation of a city-level index: a city-level pooled regression (as is standard) and a weighted summation of submarket appreciation rates. The latter aggregation approach, which has been largely ignored in the literature, involves treating a city as a basket of different submarkets in a traditional price index sense. Each submarket's change in price is given a basket weight and these are summed to arrive at a city-level index. We term this a "repeat-sales aggregation index" (RSAI) because it requires a new approach to estimate submarket repeat-sales to handle low transaction counts in some locations and time periods, and the aggregative nature of the index construction. This approach unlocks a number of important insights and possibilities regarding the construction and interpretation of repeat-sales indices.

In the RSAI the weights define the index. For some examples, a traditional pooled city-level index can be approximated by a RSAI that uses sample (i.e. transaction count) weights. Similarly, to estimate an index representing all housing units in a city, weights are assigned to the submarkets' shares of housing units. Using a submarket's share of aggregate value in a static base period gives a traditional geometric Lowe index, and time-varying initial-period value shares give a geometric Laspeyres index. Any particular weighting scheme gives an alternative "target" index.

There are two necessary conditions for target indices to differ. First, the weights assigned to each submarket must be different. If two different measures are linearly proportional in expectation across submarkets, then their associated (normalized to sum to unity) weights will be identical, leading to the same estimated index. Second, there must be heterogeneous (in expectation) appreciation across the different submarkets. If all submarkets appreciate at the same rate in expectation, then the choice of weights is immaterial and will all lead to the same estimated index.

To finish the section we describe our estimation procedure. The major difficulty in using this aggregation approach is that submarket appreciation rates are often infeasible to estimate using classic repeat-sales methods due to sparse transaction counts. We address this issue by developing an algorithm, motivated by urban economic theory, that creates period-specific aggregations of submarkets. This allows us to estimate an appreciation rate for each submarket for each period to facilitate aggregation under the least restrictive assumptions as necessary.

2.1 The repeat-sales index

Standard price index formulas (e.g. Jevons, Laspeyres, or Fisher) require matched pairs of transactions from consecutive time periods.[5] These matched pair observation are trivial to collect for homogeneous commodities like oil or corn, because these commodities are mostly identical and thickly traded. Housing, on the other hand, is both extremely heterogeneous and thinly traded, making it difficult to observe matched pairs of transactions for identical housing units in adjacent periods. BMN's major innovation was that price relatives from

[5] A "matched pair" is a set of two transactions of indistinguishable units at different points in time. This gives rise to a "price relative" which is the difference in price observed for the two transactions.

adjacent periods are not necessary to calculate an index. Rather, it is possible to calculate a Jevons (1863)-like index using matched pairs from any combination of periods.[6]

The now-common method of estimating a repeat-sales house price index is the regression model,

$$p_{it\tau} = \mathbf{D}'_{t\tau}\delta_{t\tau} + \epsilon_{it\tau} \tag{1}$$

where p is a (log) price for a transacted housing unit i, D is a dummy variable capturing the times of the consecutive transactions, $t\tau$ subscripts indicate differencing between t and τ (i.e. $p_{it\tau} = p_{it} - p_{i\tau}$ and so on), and ϵ is an IID error with mean zero and finite variance.[7] In practice, it has become common to model the error variance as a function of $t - \tau$, and use weighted least squares (Case and Shiller, 1989; Calhoun, 1996), though this is not required for consistency. The δ vector is estimated using all repeat-sales in the sample, and the (pooled) index $p^{pooled}_{t,t-1}$ is estimated using estimated parameters from equation (1):[8]

$$\hat{p}^{pooled}_{t,t-1} = \hat{\delta}_t - \hat{\delta}_{t-1} \equiv \hat{\delta}_{t,t-1} \tag{2}$$

This index formulation is typically used to represent changes to the price of a housing unit, the price of housing services, or the value of collateral on mortgages, to name several applications. But as we will soon show, each of these representations are only valid when certain conditions are met.

There is an alternative formulation that can be used to estimate this index. As Bailey, Muth, and Nourse (1963) and Wang and Zorn (1997) show, estimation of the pooled index in equation (1) is not necessary; rather, the index can be calculated as a simple weighted average of price relatives from both adjacent and non-adjacent periods. In this vein, it is

[6] For example, to estimate appreciation from 2000 to 2001, one need not be restricted to looking at the rare case of homes that transacted in both 2000 *and* 2001. As we show later in more detail, as long as 1) a home transacted two or more times, and 2) at least one transaction occurred in either 2000 *or* 2001, the home would provide information for the estimation of the index between 2000 and 2001. Similarly, an alternate estimator for 2000 to 2001 appreciation is the difference between 1999 and 2001 appreciation and 1999 to 2000 appreciation. For a longer explanation of the Jevons-BNM link, as well as derivations of weighted average representations of these indices, see Appendix B. For more discussion on the particulars of the BNM estimator, see Wallace and Meese (1997).

[7] For a useful summary of violations of this assumption, see Bogin et al. (2019a, fn. 9). In general, issues tend to focus on depreciation of the housing unit, renovations, and distressed sales.

[8] Technically speaking the indices are typically exponentiated first, though for ease of notation we will not discuss this further.

also possible to split the city into a set of N mutually exhaustive and exclusive submarkets indexed by n, and then construct the aggregated city-level index $p^{a,s}$ as a weighted average of estimated submarket indices. Here the superscript a indicates aggregation (versus a pooled index) and s indicates a particular set of weights w_n^s. As we show in the appendix, $p_{t,t-1}^{pooled} = p_{t,t-1}^{a,Sample}$ under a particular set of weights, with

$$E[\hat{p}_{t,t-1}^{a,Sample}] = \sum_{n=1}^{N} w_n^{Sample} E[\delta_{nt,t-1}] \tag{3}$$

In this representation, the weights w^{Sample} sum to unity and represent a complicated formulation of transaction counts across periods.[9,10] This alternative formulation highlights a homogeneity assumption that is often implied and overlooked. The homogeneity assumption is that all submarket appreciation rates are equal to the city-level appreciation rate in expectation, $E[\hat{\delta}_{nt,t-1}] = \delta_{nt,t-1} = \delta_{t,t-1}$ $\forall$ submarkets n. If this assumption is not maintained, then (3) shows that estimated appreciation will be sensitive to changes in sample composition, implying too small or too large estimated appreciation depending upon how proportional the sample weights are to the target weights. But what if expected appreciation rates are different across submarkets? How would this affect our interpretation of a traditional city-level repeat-sales index, estimated using transactions that are pooled across the city?

2.2 Index differences

What are the conditions under which expected submarket appreciation rate differences may affect the interpretation of a pooled-transaction, city-level index? Let p represent some general "target index" calculated by aggregating submarket appreciation rates p_n using weights w_n. We drop the a superscript for the remainder of the paper unless we are comparing an aggregated index with a pooled index.[11]

[9] The precise sample weights are given in the appendix. We approximate these complicated weights with shares of half pairs, which perform very well in our simulated setting.

[10] Also note that this formulation uses appreciation rates and not index levels, as aggregating level indices with different base values amounts to changing the weighting scheme.

[11] This section arrives at a similar conclusion as Malone and Redfearn (2020, equation 12) regarding the necessary conditions for index differences.

$$p = \sum_{n=1}^{N} w_n p_n \tag{4}$$

Similarly let p' be another target index that uses different weights w'_n to calculate city-level appreciation using the same set of submarket appreciation rates. If the submarket price indices are same for both indices, then all differences in city-level indices are due to the choice of weights.

$$p - p' = \sum (w_n - w'_n) p_n \tag{5}$$

A necessary condition for two target indices to be different is that their weights must not be identical. However, this is less trivial than may appear at first glance because the weights should be interpreted as relative shares or measures of each submarket with respect to some population measure of interest. In practice it may be possible that two seemingly different metrics generate largely similar submarket weights, leading to very similar measured appreciation. For example, though aggregate value and sample weights are conceptually very different, they would produce the same index if homes transacted in a manner proportional to their value.

Another necessary condition for the two target indices to differ is that there must be differences in expected appreciation rates across submarkets. The choice of weights is irrelevant if all submarkets appreciate at the same rate. More formally, if $p_n = \tilde{p}$ for all n, any weighting scheme would give the same city-level index, as $\sum (w_n - w'_n) p_n = \tilde{p} \sum (w_n - w'_n) = 0$ because shares are defined to sum to unity.[12] When submarkets have heterogeneous appreciation rates, the sign of the difference is equal to the sign of the correlation between the weight differences and the submarket appreciation rates, $sign(p - p') = sign\,(corr(w_n - w'_n, p_n))$.[13] In other words, when a submarket has both higher appreciation and a higher weight relative to other submarkets, then the resulting aggregate will also tend be higher.

[12] This is also true more generally in expectation. If $E[p_n] = \tilde{p}$, then $E[\sum (w_n - w'_n) p_n] = \sum (w_n - w'_n) E[p_n] = \tilde{p} \sum (w_n - w'_n) = 0$.

[13] Two indices may still be equal by chance, either through the knife-edge result where the inner product is exactly 0 despite appreciation heterogeneity, or when the alternative weights are by chance equal to the weights used in the target index, e.g. under pps sampling (Balk, 2012, p. 182).

To summarize, the two necessary conditions for estimates of a house price index to be generally larger or smaller than a target index are (1) heterogeneous appreciation across submarkets and (2) weights that differ from the target index. Only when both conditions are met can there be differences between alternative target indices.

2.3 Index targets

Using aggregated indexes to measure appreciation allows one to easily adapt indices for different purposes by choosing appropriate alternative weighting scheme. Additionally, it is no longer necessary to assume that a city-level index, estimated using equation (1), represents a particular target index. Rather, weights can be chosen explicitly to estimate an intended target index, and where differences in appreciation can clearly be attributed to the different weighting schemes.

For example, suppose we (implictly) use sample weights by estimating a pooled sample, giving a city-level index $p^{a,Sample}$ as in equation (3). If our target index is the rate of appreciation of all homes in a city $p^{a,Unit}$, which can be represented using housing stock weights, then we will overestimate our target appreciation if submarket appreciation rates are positively associated with higher sampling frequencies. This specific issue has been addressed in the literature concerning sampling bias in repeat-sales indices (e.g. Gatzlaff and Haurin, 1998), but it is encompassed within our general aggregation framework. Hence $p^{a,Unit}$ can be computed directly, rather than (implicitly or explicitly) assuming representative sampling and/or homogeneous appreciation within the city and using a pooled index.

Other potentially useful alternatives within this framework include classic price index aggregation formulations, such as using rolling initial-period housing *value* shares as weights, giving a geometric Laspeyres formulation p^{gla}, or fixed initial-period housing value shares as weights, giving a geometric Lowe index formulation p^{gwa}. Both of these formulas are useful because they give an index that purports to represent the price of a continuous unit of housing services rather than the price of a housing unit.[14] This need is common in urban economics to assess producer and consumer optimization problems where the intensive margin of housing (e.g. interior square feet or quality) matters. Another possibility is using a share of mortgage loan collateral in each submarket as weights. This can be used to construct a city-level index that represents the average change in collateral value in a loan portfolio, a case that may be appropriate for analysts pricing mortgage-backed securities. For all applications in this paper, we use initial-period (i.e. t-1) weights, resulting in Laspeyres-like indices that isolate changes in prices from changes in weights during the period of the price relative, though other formulations are of course possible.

These are just several examples of different target indices. The key insight is the flexibility of the aggregation approach: it is no longer necessary to use a single sample-based estimator to represent various target indices. Various indices can be quickly and easily calculated directly using any desired weighting scheme. Of course, this assumes submarket house price indices are available. It turns out this is not a trivial problem. We now turn to the feasible estimation of submarket appreciation rates.

2.4 Estimating submarket price indices

It is difficult to estimate repeat-sales house price indices for small submarkets in all time periods. Inherent in all repeat-sales estimation is a tradeoff between aggregation bias and estimation error. Aggregation bias occurs when an index is estimated assuming homogeneity, but in reality, the entities in the sample are heterogeneous. Estimation error is statistical in

[14] Diewert and Shimizu (2017) argue that indices estimated in this manner are not necessarily price indices in the traditional sense, but rather "asset value price indices". There are a number of assumptions necessary to interpret a change in the value of a housing unit asset as a change in the flow price of housing quantities (also called services), as we typically think of items in a price index. Typically, the two most important are the assumption that the quantity of housing does not vary over time for the same housing unit, and that the rate of market capitalization is unchanging. These assumptions immediately follow from the simplest textbook representation of the dividend pricing hypothesis. This formulation posits the value of an asset as the sum of the discounted expected future flows, or $V = \int_{t=1}^{\infty} E_0[(1+r)^{-t} p(t)q(t)dt]$. This particular representation abstracts from the many things that make housing different than other assets, including depreciation, taxes, expectations, and other factors we would see in the traditional housing user-cost framework. See Piazzesi, Schneider, and Tuzel (2007) for a more realistic treatment. Then, assume a steady state where p and q do not change over time, reducing the valuation formula to that of an infinitely-lived annuity, $V = pq/r$. We can now readily see how the condition $\frac{dV}{V}/d\alpha = \frac{dp}{P}/d\alpha$ for some exogenous shifter α requires a constant q and r. In reality, these assumptions are typically violated in different ways, as repeat-sales do not adjust for natural obsolescence of housing structures, long-run interest rates are always evolving over time, and there are other myriad expectations and variables in the landowner's profit maximization function may change (McMillen, 2003; Tian, 2017; McMillen and Shimizu, 2020). However, these sources of variability are assumed away in order to proceed, as is standard, though we believe they are worth noting explicitly.

nature, and decreases with the number of observations in the sample. Balancing these two errors on the margin, perhaps formally with a loss function, determines the optimal aggregation level. For housing, using out-of-sample testing for annual indices, Bogin, Doerner, and Larson (2019b) find that Census tracts and ZIP-5 codes had nearly equivalent error in cities, with Census tracts having lower aggregation bias but higher estimation error, and ZIP codes having higher aggregation bias but lower estimation error. Indices measured by aggregating below (e.g. Census blocks) or above these levels (e.g. ZIP-3s or counties) were not on the error minimization frontier. Our choice of tract-level estimation agrees with the suggested aggregation levels proposed by Bogin, Doerner, and Larson (2019b).

Unfortunately, tract (or even ZIP-5)-level indices are not always feasible to estimate for all periods in all cities. To ensure all homes are represented by a submarket price index in the eventual city-level index construction, we use a unit of geography we term a *supertract*. Supertracts are aggregations of Census tracts. We maintain that it is reasonable to aggregate nearby Census tracts based on the isoutility condition in the classic monocentric city model of Alonso (1964), Mills (1967), Muth (1969), as summarized by Brueckner (1987). In this model, due to location substitutability, when prices change in one area, they will change at a similar rate in nearby areas. We maintain this same assumption: even if adjacent areas are have different home *values*, the appreciation rates will be very similar, meaning we are able to maintain a homogeneity assumption across nearby areas with minimal introduction of aggregation bias.

Aggregation is performed based on decision rules related to a minimum observation count and tract adjacency: if a tract does not meet a threshold number of observations, it is aggregated with the nearest tract into a supertract. If this supertract still does not have enough observations, it is aggregated into the nearest tract or supertract (based on centroid distance), and so on, until the threshold is met.[15] This ensures each geographic area for which we estimate an index has a minimum number of observations and that the indices cover areas representing all the housing units in the city. The assumption we make is that homes in the same supertract appreciate at the same rate in expectation.[16]

An aggregated city-level index $\hat{P}^a$ is constructed one year at a time using the following

[15] It is possible to increase the dimensionality of the comparison beyond centroid distances to include other tract-level housing unit or household characteristics. We leave this for future research.

[16] Lists of supertracts for each time are available upon request.

algorithm for each period $t = 1...T$. Assume a city level index (not in logs) is initialized such that $P^a_{t=0} = 1$.

1. Construct N_t supertracts for period t using period t and $t-1$ transaction counts.
2. Calculate classic BNM price indices for each supertract using the entire time sample.[17]
3. Capture $\hat{\delta}_{n,t}$ and $\hat{\delta}_{n,t-1}$ for each supertract, treating all other parameters as controls and discarding them.
4. Calculate or specify the desired supertract weights w_n.
5. Aggregate the rate of change at the city level as $\hat{p}^a_t = \sum_{n=1}^{N_t} w_n(\hat{\delta}_{n,t} - \hat{\delta}_{n,t-1})$.
6. Construct $\hat{P}^a_t = \hat{P}^a_{t-1} \times \exp \hat{p}^a_t$

Note that despite each repeat sales estimation generating price indices for all time periods, we must estimate a new vector of parameter for each year due to changing supertract definitions. This index is computationally intensive to estimate by historical standards, requiring estimating $T \times T \times \bar{N}$ parameters rather than T as in the pooled, city-level BNM specification. However, this algorithm ensures each housing unit is represented in each period of the index, and each submarket appreciation rate is estimated using a minimum number of transactions.

3 Synthetic Indices

Before estimating repeat-sales aggregation indices using real-world data, we examine properties of various target indices in a simulated laboratory setting. The model used to generate housing quantities, prices, and values in each simulated tract of the city draws from the long tradition of monocentric city models. The model in its modern form was developed over a series of seminal works by Alonso (1964), Mills (1967), and Muth (1969), summarized well by Brueckner (1987). The particular model used here is based on Larson and Yezer (2015), which is found in the appendix. The point of this model is not to closely mimic particular

[17] We estimate BNM (1963) indices rather than a more modern weighted estimator aimed at correcting heteroskedasticity because we do not wish to interfere with sampling. Assuming quality changes are constant per year in expectation, regardless of the holding period, the BNM estimator is consistent. Because our target index involves averaging over a large number of observations and tracts, any increased variance should have *de minimis* effect on the variability of the target index.

real-world cites or economic shocks. Rather, the goal is to show in a sterile environment how prices, quantities, values, cap rates, and changes to these factors over time can cause differences in measured price indices. Parameters are loosely calibrated for illustrative purposes to resemble a city at its long-run equilibrium in 1990 and then again at a new long-run equilibrium in 2020. We refer to these cities as the baseline (or period 0) and comparison (or period 1) cities.

This model generates plausible values of house prices, p, house quantities, h, and housing units n, for different areas of the city under alternative conditions. Another city is then simulated based on alternative parameterizations, and homes in the same location are compared; price differences are used to calculate pseudo appreciation rates, quantity differences are used to calculate changes in home size, and so on. The city is identical in every direction from the center of the city, so variables are expressed as a function of distance from the central business district, or CBD. Thus in our simulation, submarkets are distinguished by their distance from the center of the city.

Following Wang and Zorn (1997), we use the model to construct all values (i.e. the population) for both cities under full information. Afterwards, we introduce different sampling schemes to generate different types of samples from the population. Because we construct the spatial distribution of population in the simulation, we have the added benefit of observing differences in estimated appreciation rates that arise from differences in appreciation rates (comparison between two cities) with those differences that arise from different sampling schemes.

The first (baseline) city mimics a city of about 1 million households in around 1990. The house price gradient is downward sloping, the land price gradient is steeper than the house price gradient, and the quantity of housing services consumed per household is upward sloping, each in accordance with "Muth's (1969) equation" and the handbook chapter succinctly written by Brueckner (1987). The second (comparison) city is identical to the first, except for the introduction of a large CBD amenity and an increase in household income. This amenity is meant to illustrate the steep rise of center-city amenities between 1990 and 2020 (Couture and Handbury, 2020), and the income change represents export productivity gains. These combine to produce steeper house price gradients, giving an appreciation rate gradient between the first and the second city similar to that which was found empirically for large cities over the period by Bogin, Doerner, and Larson (2019a).[18] Figure 1 illustrates the house-price gradients (for the complete set of gradients, see appendix Figure A.1).

From this basic setup, we can introduce transactions by assuming sampling distributions for each house in the simulation to reflect real-world idiosyncracies in sampling. Figure 2 illustrates our three sampling schemes. These layers take us from the deterministic world of our computable general equilibrium model to the stochastic world of Wang and Zorn (1997).

We generate estimated appreciation rates under four different sampling regimes, each corresponding to a different weighting scheme, and two different estimators. The two estimators are the repeat-sales aggregation index and the pooled BMN (1963) index. Let $s_i = 1$ if pair i is used in the estimation, i.e. included in the sample. The four sampling distributions represent (1) the full population, $s_i = 1$ for each house i; (2) uniform 5% sampling, $s_i = I(d =< 0.05)$ where I is the indicator function and $d \sim U(0,1)$; (3) suburb-tilted sampling, $s_i = I(\Lambda \geq 0.5)$ where $\Lambda = 1/\exp(-0.95 + d + (k/\bar{k} - 1)/20)$, k is the distance from the CBD, and $\bar{k}$ is the mean distance across all housing units in period 0; (4) and CBD-tilted sampling with $\Lambda = 1/\exp(-0.95 + d - (k/\bar{k} - 1)/20)$. The associated weighting schemes for the latter three sample distributions are value shares (i.e. Laspeyres), sample shares, and housing unit shares, respectively. Importantly, we purposefully create samples where supertracts are needed to mirror their use with actual data. "Tracts" for our present purpose are defined as 0.01 mile annuli. We also introduce a residual $\ln \epsilon \sim N(0, 0.25)$ into the sampled price to reflect variation in transaction prices around a local market average price that can occur due to buyer/seller search dynamics (e.g. Genesove and Han, 2012).

These sampled transactions are then run through the algorithm in Section 2.4. Annuli with too few transactions are grouped with adjacent annuli until a threshold number of transactions is met. Each supertract index is then estimated with appreciation rates applied to component annuli. Then, city-level indices are calculated as weighted averages.

Table 1 shows four main results. First, the sample, unit-weighted, and value-weighted indices each give different appreciation rates. Second, both the value and unit-weighted indices are robust to different sampling regimes because weights depend upon only population charac-

[18] The city gives a new long-run equilibrium with each solution, and accordingly, housing quantities adjust too much to be realistic over a 30 year period.

teristics that are invariant to different sampling schemes. Third, as expected, the accuracy of the sample-based index (i.e. the typical pooled index) in achieving the other target indices is highly sensitive to the spatial sampling distribution. For instance, only when sampling shares mimic value shares by chance, as is the case in CBD-tilted sampling, does the sample-weighted index approach the target index. Finally, note the population sample-weighted repeat-sales aggregation index is essentially the classic pooled city-level repeat-sales index from BMN (1963), though with some small differences not shown due to rounding. Even under different sampling schemes and the introduction of estimation error, the two indices are virtually identical. This reiterates our previous point that typical city-level pooled indices are effectively aggregated indices with sample shares given by share of half-pairs as weights.

4 Data

The housing transactions data include all purchase-money mortgages acquired or guaranteed by Fannie Mae, Freddie Mac or FHA, and purchases that are present in CoreLogic's deed and recorder data. This data set contains over 63.3 million repeat-sales in the United States from 1975 through 2021. For each property, we have its Census tract as defined according to the 2010 Census, regardless of its tract identifier when the transaction occurred. We construct annual house price indices for 581 CBSAs (also referred to as "cities") from 1989 through 2021, using transactions and indices prior to 1989 as discardable controls.

Paired transaction prices are log-differenced to generate appreciation rates (price relatives) for individual homes. To eliminate transaction pairs that suggest substantial quality changes, we subject all price relatives to three filters. First, we remove any pair that transacts in the same 12 month period to eliminate possible "flips". Second, we remove any pair with a compound annual growth rate greater than $|30\%|$.[19] Finally, we remove any pair with cumulative appreciation greater than 10 times the prior sale or less than 75% below the prior sale.

Supertracts are then defined based on a minimum number of half pairs in each quarter of a two-quarter window. Half pairs are defined as in Bogin, Doerner, and Larson (2019a) as the sum of the number of price relatives with either the first or the second transaction in

[19] The compound annual growth rate is calculated as $(V_1/V_0)^{(1/(t_1-t_0))} - 1$.

a particular time period.[20] We set a cutoff of a minimum 40 half-pairs for each year in the two-year window. If a tract or supertract has fewer than 40 half-pairs in either year, the area is combined with the nearest area in accordance with the algorithm in Section 2.3.[21] If there is any year in which fewer than 40 half-pairs are observed in the entire city between 1989 and 2021, then the city is discarded from the sample, thus ensuring that there will be at least one supertract in each city for that range of years.[22] This cutoff is arbitrary, but tends to result in tract and city-level estimates with relatively little apparent noise. Repeat sales indices are then estimated for each supertract in the 581 cities using the Bailey, Muth, and Nourse (1963) (unweighted) formulation. This algorithm is computationally intensive, requiring estimation of 1,037,465 supertract-year regressions and over 47 million parameters in total. The resulting indices are then aggregated to the city level in various ways.

We calculate six city-level house price indices: 1) "Sample" based on sample (half-pair) weights, 2) "Value" based on the geometric Laspeyres formation using aggregate single-family housing value, 3) "Unit" based on the share of single-family housing units, 4) "UPB" based on the aggregate outstanding balances for Enterprise (Fannie Mae and Freddie Mac) loans, 5) "College" based on the college-educated population, and 6) "Non-White" based on ACS share of the non-white population. Indices 1-4 have weights that vary over time, and 5-6 have weights that are static and based on the 2010 Census, and thus represent a Lowe-like index formulation.[23]

[20] A half pair is a measure of the information contained within the price relatives. For instance, for an annual index, if there is a sequence of three transactions for a property in 1992, 1998, and 2018, there would be two repeat-sales, one for the 1992-1998 pair and another for the 1998-2018 pair. Thus this property would contribute one half-pair in 1992, two in 1998, and one in 2018.

[21] This number was calibrated on a somewhat ad-hoc basis to balance aggregation bias with estimation error. We leave further investigation for future research. Other values attempted were 25 and 100, which led qualitatively to similar results.

[22] In the case of a single supertract in the city, the pooled city-level index will be identical to the repeat-sales aggregation index. However, this is rare, and tends to occur only early in the sample and for small cities.

[23] Value and unit shares are calculated using the decennial Census in 1990, 2000, 2010, and the American Community Survey in 2017. UPB is only available starting in 2001 is based on internal FHFA data. For the Value index, in each supertract, the aggregate value is calculated using the identity $V_{nt} = p_{nt} \times q_{nt} = p_{nt} \times q_{nt}/h_{nt} \times h_{nt}$, where the median value approximates $p_{nt} \times q_{nt}/h_{nt}$. So, while the quantity of housing q_{nt} in tract n at time t is never directly observed, it can incorporated into the measure by using the quantity of homes h_{nt} instead. Value shares are then calculated as $s_{nt} = V_{nt}/\sum V_{nt}$. For years without shares, we impute using straight-line imputation. For years outside the interval [1990,2017], the trend is set to the most recent within-interval trend. An identical estimation procedure is used for non-base-year shares.

5 Results

5.1 Single-city illustration: Washington DC, CBSA

To illustrate the creation of the repeat-sales aggregation index for a large metropolitan city, we consider the Washington, DC CBSA. This area is dominated by the center state entity, the District of Columbia. There are also urban sub-centers on the borders with the District and in further suburban locations, owing in part to the District's building height limit and high levels of traffic congestion.

Using the supertract approach, we calculate annual appreciation rates for each of the 1,344 tracts in the CBSA from 1989 through 2021. These generate between 400 and 1,100 supertracts for each year, depending on the year (see appendix Figure A.2 for the time series). To illustrate the spatial appearance of supertracts and their associated house price indices, see Figure 3 for the supertracts within the District of Columbia. This figure shows traits which are common across areas and time periods, including the number of tracts per supertract, which is typically low, the shape of supertracts which is typically compact, and the appreciation rate of nearby supertracts, which is very similar.

We can see in Figure 4, panel (a) that there is substantial variation in the appreciation rate of housing across tracts in the Washington, DC CBSA, and that this variation is increasing over time. There are patterns, however, with the most immediately visible being the gradient as a function of the distance from the central business district (CBD). This gradient is shown in panel (b), with appreciation rates of approximately 4.7% per year at 1 mile from the CBD down to 3.4% at 20 miles from the CBD. Recall that this variation in appreciation rates across tracts is a necessary condition for differences in city-level indices under different weighting schemes.[24]

We then create six different indices out of these annual tract-level measures of appreciation. Figure 5, panel (a), shows these alternative tract weighting schemes for a single year (2021), depicted as the cumulative weight within various distances from the CBD. Here a higher cumulative weight for an index corresponds to more effective weight being placed to homes

[24]Note that for differences between indices, there need not be an appreciation gradient; for example, distance to CBD may not be as relevant for appreciation in smaller cities. All that is necessary for differences to exist is a correlation between the proportionality of sampling and appreciation rates. Distance to CBD is simply an illustrative and established covariate with appreciation rates.

closer to the CBD. This allows us to quickly identify how weights are tilted towards the center of the city versus the suburbs. For example, in 2021, the share of non-white population is the weighting scheme most tilted towards the CBD.

Panel (b) shows how the median distance from the CBD for each set of weights evolves over time. Both value and unit weights shift slowly as price gradients rotate and the quantity of housing units and services change. Sample weights, on the other hand, fluctuate dramatically from year to year. Note that the gradual suburbanization of the sample distribution is to a large degree a natural artifact of the suburbanization of housing units; as the median location of the population shifts, so too does the median transaction under proportional sampling. The UPB weights are fairly cyclical in their median location over the 2001-2020 housing cycles, moving first to a more suburban tilt in the run-up to the Great Recession, then falling afterwards until about 2010, then rising through the end of the sample.

Finally, we can see how the alternative weighting schemes and non-zero appreciation gradients interact to cause variation in city-level indices. Figure 6 highlights a sixteen year range where there are major differences in both gradients and sampling. In 2005, each index is initialized to be equal to 1, but by 2009 the College and Value (Laspeyres) indices are substantially higher, and the non-white index is substantially lower. By 2021, each index value is nearly identical again, with the exception of UPB, which is the lowest. Thus, if one were to use a city-level index to estimate changes in housing wealth by non-white homeowners, a very different picture would result if one used a traditional Sample-based index versus one tailored to track the fortunes of predominantly non-white areas, which experienced a bigger bust from 2005-2011, but then greater appreciation between 2011 and 2021.

5.2 Results within and across cities

When we extend the analysis to the 581 CBSAs over which we are able to estimate the repeat-sales aggregation indices, several patterns emerge. First, we show that it is likely for differences across index formulations to exist simply by looking at aggregates: the combination of city average appreciation gradients with changes in city average sampling locations over time suggests sample-based indices may differ from population-based target indices. Next, we show that there are differences in measured appreciation rates across index formulations, and these differences can be large and may accumulate over lengthy time windows, including boom or bust periods of a housing cycle. Differences are most likely in large cities because locations are less substitutable within-city, suggesting the possibility of non-random variation in appreciation rates that may be correlated with transaction frequency. Finally, we explore two recent periods to illustrate how sampling and appreciation rates may interact to cause differences in indices (the Great Recession), or be, perhaps surprisingly, rather benign (COVID-19 period).

Is it possible to predict differences in indices across various formulations simply by looking at national aggregates? Figure 7 replicates the findings of Bogin, Doerner, and Larson (2019a) from 1989 through 2021, where large cities, on average, have steep appreciation gradients, unlike small or medium-sized cities. Based on this finding alone, we can surmise that city-level prices generated using the standard sample-based formulation may depart from other target indices in large cities. This is made more likely when we observe the distance of the median-weight Census tract in various weighting schemes and time periods. Figure 8 shows the median weight distance from the CBD has been growing in medium and large cities over the 30 year period for many weights, including Value, Unit, and UPB, reflecting increasing suburbanization over the period.[25] The location of the median transaction varies much more than the other weights over time, indicating the distribution of within-city transaction patterns is not constant over time.

But do differences across index formulations, in fact, exist? In order to summarize the extent of differences between various indices, we perform a number of exercises related to within-city differences in appreciation rates generated by alternative weights. We set the Sample index as the reference, and calculate pairwise differences in measured appreciation rates as $100 * (\ln P_{end}^{Sample} - \ln P_{start}^{Sample} - (\ln P_{end}^{W} - \ln P_{start}^{W}))$, where W is some weighting scheme such as Unit or Value. Negative (positive) values of this measure indicate that the use of sample weights results in understatement (overstatement) of appreciation relative to the chosen target, i.e. that homes in high appreciating areas were under(over)-sampled in proportion to their target shares. After estimating some trial models relating the size of the difference to city size by year, we observe there are five periods over the last 32 years where annual appreciation rate differences are of similar sign and tend to accumulate over the period. These ranges are 1989-2001, 2001-2005, 2005-2013, 2013-2019, and 2020-2021.

[25] "Large city" is defined as a CBSA (2020 definitions from Office of Management and Budget) with component counties where the number of housing units in 1990 was greater than 1 million. A "medium city" is between 250,000 and 1 million housing units. A "small city" is less than or equal to 250,000 units. There are 19 large cities, 50 medium cities, and 512 small cities.

Kernel densities of cumulative appreciation rate differences for the unit-based index over respective periods are shown in Figures 9, and over the full sample are shown in Figure 10. Distributions for small and medium cities are typically centered around zero. We see in the 1989-2001 period that large cities tend to have a negative gap, indicating that sample-based indices are generating price indices that are consistently lower than that which would be represented by the full stock of single-family housing units. This is due to the interaction of a negatively sloped appreciation gradient with disproportional sampling in suburban areas. In the 2001-2005 period, both large and small areas are centered around 0 but with higher variance in large cities. In the 2005-2013 period all size classes have a positive difference, indicating high-appreciation homes are more-than-proportionally sampled. The 2013-2019 and 2019-2021 periods are mirror images of the prior 8-year period: areas that were low-appreciation were over-sampled during the Great-Recession, but in the following decade, high-appreciation homes were over-sampled. On a net basis, large cities have a negative skew and medium/small cities have a slight positive skew over the 30 year period, as shown in Figure 10.

Table 2 list all of the large cities in the sample, along with annual average appreciation differences across each grouping of years, and the percent of cities with overstated appreciation during the respective time frames. Most sample-based indices for large cities understated appreciation relative to measures using the single-family housing stock as weights, with 4 out of 19 large cities experiencing overstated appreciation in excess of 1% between 1989-2021, and 9 out of 19 experiencing understated appreciation in excess of -1%. The most extreme periods were 1989-2001 and 2013-2019. The 2001-2005 and 2019-2021 periods were more middling, with some areas facing positive gaps and some negative. The most extreme gaps were the two large Texas cities of Houston and Dallas, each with a gap in excess of -7 percentage points.

We can see from Table 3 that gaps that emerge within cities can be correlated across time windows, and thus can be reinforcing or offsetting depending on the time period. For instance, Sample vs Unit index gaps that emerge between 1989 and 2001 are largely uncorrelated with future gaps, as are gaps that emerge in the 2001-2005 period. In contrast, when one index diverges from the other within a city in the 2005-2013, the gap tends to reduce in the years that follow, with correlations of -0.31 and -0.25 in successive windows. Also, while in many respects, the broad 2013-2019 price increases are similar in geography to 2019-2021, gaps between price indices are perhaps surprisingly uncorrelated.

In the same manner by which we treat tracts as sub-units and aggregate up to cities, it is possible to aggregate cities to national values.[26] When we aggregate appreciation rates for cities within size classes to large, medium, and small-city averages, we observe cross-city trends in index differences over time. Figure 11 shows each of these aggregates by city size from 1990-2021. We can see clearly from these figures how small and medium cities that had high rates of appreciation also transacted more-than-proportionately compared to the other weighting schemes considered. Accordingly, national metrics that simply pool together transactions will give potentially misleading rates of appreciation across the time period in question. For instance, for small cities, sample-based indices are between 9 and 20 percentage points above the others by the end of the sample. In large cities, it is less clear–rates of rates of transaction were not necessarily associated with rates of appreciation, with some weighting schemes giving higher or lower rates of appreciation, depending on the weight and the time period.

But how do cumulative differences for individual cities compare with one another? The cells in Table 4 show cumulative differences between indices over a particular window, with asterisks giving statistical significance of the hypothesis that a mean is zero. This table suggests that an average index in each city size class did a good job at documenting the run-up in house prices between 2001 and 2005, as there is little difference between index types in any geography. Beyond this period, however, there are systemic differences across index types across different city size classes. For example, the difference between the Value (Laspeyres) and Unit indices give us some sense if high versus low valued units are appreciating at higher rates; when the Value index gives higher appreciation, such as the 2013-2021 period, that means higher valued units are appreciating at higher rates than low-valued units, suggesting rising housing wealth inequality. Additionally, the large magnitudes for the non-white population-weighted index suggests non-white homeowners faced bigger drops in prices between 2005-2013, but much higher appreciation between 2013 and 2019. Overall, however, differences reported in Table 4 are much smaller than Figure 11 and sometimes have opposite

[26]Note this is similar to FHFA's national house price index, which is a housing stock-weighted average of census division indices, and the S&P Case-Shiller composite indices, which aggregate by weighting by each city's aggregate housing stock value from the previous decennial census. See and Calhoun (1996) and `https://www.spglobal.com/spdji/en/documents/methodologies/methodology-sp-corelogic-cs-home-price-indices.pdf` (accessed May 2022), respectively.

signs. This occurs because while cities may have proportional weights within-city, they may not be proportional across cities, and this lack of proportionality occurs in a manner that is correlated with appreciation.

Finally, to dig deeper into within-city cross-period correlations in differences, we turn to visual inspection of appreciation and sampling gradients in the Great Recession and COVID-19 periods. Figure 12 shows a steep appreciation gradient in 2009, combined with undersampling of near-to-CBD homes, implying sample-based indices show lower appreciation than those representing all units. In the COVID-19 period, it has been widely documented that households sought to increase housing consumption due to mobility restrictions and the widespread adoption of telework, increasing relative suburban demand (e.g. Brueckner et al., 2021). We can see this in the rotation of the house price gradient for 2021 but not 2020. In fact, both appreciation gradients are similar in 2020 to both 2019 and 2018. Meanwhile, transactions shifted towards the suburbs in 2020, then became roughly proportional in 2021, at least on the distance-from-CBD dimension. Now, recall that for index differences to exist, tract-weighted differences and and appreciation rates need to be correlated. Combined, these results suggest minor index differences due to this spatial dimension; 2020 had a flat appreciation gradient but suburb-tilted sampling, while 2021 had an upward-sloped gradient but with more proportional sampling. On a net basis, there is no major difference between the average Sample index and the Unit index in either 2020 or 2021, for any city size class, suggesting standard transactions-driven indices produced by Case-Shiller and FHFA for the period were representative of the housing stock.[27]

Overall, these results highlight the need for bespoke indices of house price appreciation depending on the use case.[28] Using the tract-level indices we have developed, along with use-defined weights, it is possible to create a house price index that is tailored to the task at hand. Sample-based indices, which are typically used as a one-size-fits-all solution for collateral valuation, housing wealth calculations, prices of housing services, measures of affordability, and so on, can give misleading estimates that can accumulate over both small and large time intervals. Accumulation is most likely in large cities because of their often-

[27] For 2020, the small city average is {Sample, Unit}={7.2%, 7.1%}, medium is {7.8%, 7.7%}, and large is {6.2%, 6.3%}. For 2021, the small city average is {14.9%, 14.4%}, medium is {15.2%, 15.1%}, and large is {11.8%, 11.8%}.

[28] Some additional metrics could include climate or disaster risk weights, housing tenure, school quality, land cover, amenity, or other socioeconomic indicators.

times non-zero appreciation gradients. Sustained appreciation gradients can often lead to a non-zero correlation between weight differences and appreciation rates, leading to difference accumulation.

6 Conclusion

While granular submarket indices have been available for some time, this paper is the first to create a balanced panel of thousands of Census tract appreciation rates for over three decades. These tracts are exhaustive of all housing units in 581 CBSAs, facilitating a new house price index estimation approach based on aggregation of submarket appreciation rates. Using simulations, we have shown this new approach produces indices that are identical to existing indices created by pooling all repeat-transactions within a city.

This aggregation approach allows for the straightforward construction of custom city-level house price indices. Rather than using a sample-based index to represent many different populations, we show how to construct city-level indices tailored to a particular population. Some examples include the stock of all housing units in a city, the aggregate home value (giving a Laspeyres index), loan balances, or individual demographic characteristics. We expect these new tract-level indices and our aggregation approach to provide a framework for novel research programs and allow policymakers to better track house price movements in summary form.

References

Alonso, W. (1964): "Location and land use. toward a general theory of land rent." *Location and land use. Toward a general theory of land rent.*

Altmann, J. L. and J. S. DeSalvo (1981): "Tests and Extensions of the Mills-Muth Simulation Model of Urban Residential Land Use," *Journal of Regional Science*, 21, 1–21.

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

Balk, B. M. (2012): *Price and quantity index numbers: models for measuring aggregate change and difference*, Cambridge University Press.

Bertaud, A. and J. K. Brueckner (2005): "Analyzing building-height restrictions: predicted impacts and welfare costs," *Regional Science and Urban Economics*, 35, 109 – 125.

Bogin, A., W. Doerner, and W. Larson (2019a): "Local house price dynamics: New indices and stylized facts," *Real Estate Economics*, 47, 365–398.

Bogin, A. N., W. M. Doerner, and W. D. Larson (2019b): "Missing the mark: Mortgage valuation accuracy and credit modeling," *Financial Analysts Journal*, 75, 32–47.

Brueckner, J., M. E. Kahn, and G. C. Lin (2021): "A new spatial hedonic equilibrium in the emerging work-from-home economy?" Tech. rep., National Bureau of Economic Research.

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," *OFHEO Working Paper Series*.

Carrillo, P. E. and A. M. Yezer (2021): "What Do Hedonic House Price Estimates Tell Us When Capitalization Rates Vary?" *Mimeo*.

Case, K. and R. Shiller (1989): "The Efficiency of the Market for Single-Family Homes," *American Economic Review*, 79, 125–37.

Couture, V. and J. Handbury (2020): "Urban revival in America," *Journal of Urban Economics*, 119, 103267.

Deng, Y., J. M. Quigley, and R. Van Order (2000): "Mortgage terminations, heterogeneity and the exercise of mortgage options," *Econometrica*, 68, 275–307.

Diewert, E. and C. Shimizu (2017): "Alternative approaches to commercial property price indexes for Tokyo," *Review of Income and Wealth*, 63, 492–519.

Gatzlaff, D. H. and D. R. Haurin (1997): "Sample selection bias and repeat-sales index estimates," *The Journal of Real Estate Finance and Economics*, 14, 33–50.

——— (1998): "Sample selection and biases in local house value indices," *Journal of Urban Economics*, 43, 199–222.

Genesove, D. and L. Han (2012): "Search and matching in the housing market," *Journal of Urban Economics*, 72, 31–45.

Goetzmann, W. N. (1992): "The accuracy of real estate indices: Repeat sale estimators," *The Journal of Real Estate Finance and Economics*, 5, 5–53.

Green, R. K., S. Malpezzi, and S. K. Mayo (2005): "Metropolitan-specific estimates of the price elasticity of supply of housing, and their sources," *American Economic Review*, 95, 334–339.

Hill, R. J. and D. Melser (2008): "Hedonic imputation and the price index problem: an application to housing," *Economic Inquiry*, 46, 593–609.

Jevons, W. S. (1863): *A Serious Fall in the Value of Gold Ascertained: And Its Social Effects Set Forth*, E. Stanford.

Larson, W., F. Liu, and A. Yezer (2012): "Energy footprint of the city: Effects of urban land use and transportation policies," *Journal of Urban Economics*, 72, 147–159.

Larson, W. and A. Yezer (2015): "The energy implications of city size and density," *Journal of Urban Economics*, 90, 35–49.

Malone, T. and C. L. Redfearn (2020): "To measure globally, aggregate locally: Urban land, submarkets, and biased estimates of the housing stock," *Real Estate Economics*.

——— (2021): "Land Leverage, Housing Submarkets, and the Performance of Aggregate Housing Price Indices," *Mimeo*.

McDonald, J. F. (2009): "Calibration of a monocentric city model with mixed land use and congestion," *Regional Science and Urban Economics*, 39, 90 – 96.

McMillen, D. and C. Shimizu (2020): "Decompositions of house price distributions over time: The rise and fall of Tokyo house prices," *Real Estate Economics*.

McMillen, D. P. (2003): "The return of centralization to Chicago: using repeat sales to identify changes in house price distance gradients," *Regional Science and Urban Economics*, 33, 287–304.

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

Muth, R. (1969): *Cities and housing: the spatial pattern of urban residential land use*, University of Chicago Press.

Muth, R. F. (1975): "Numerical solution of urban residential land-use models," *Journal of Urban Economics*, 2, 307–332.

Piazzesi, M., M. Schneider, and S. Tuzel (2007): "Housing, consumption and asset pricing," *Journal of Financial Economics*, 83, 531–569.

Skabardonis, A. and R. Dowling (1997): "Improved speed-flow relationships for planning applications," *Transportation Research Record: Journal of the Transportation Research Board*, 1572, 18–23.

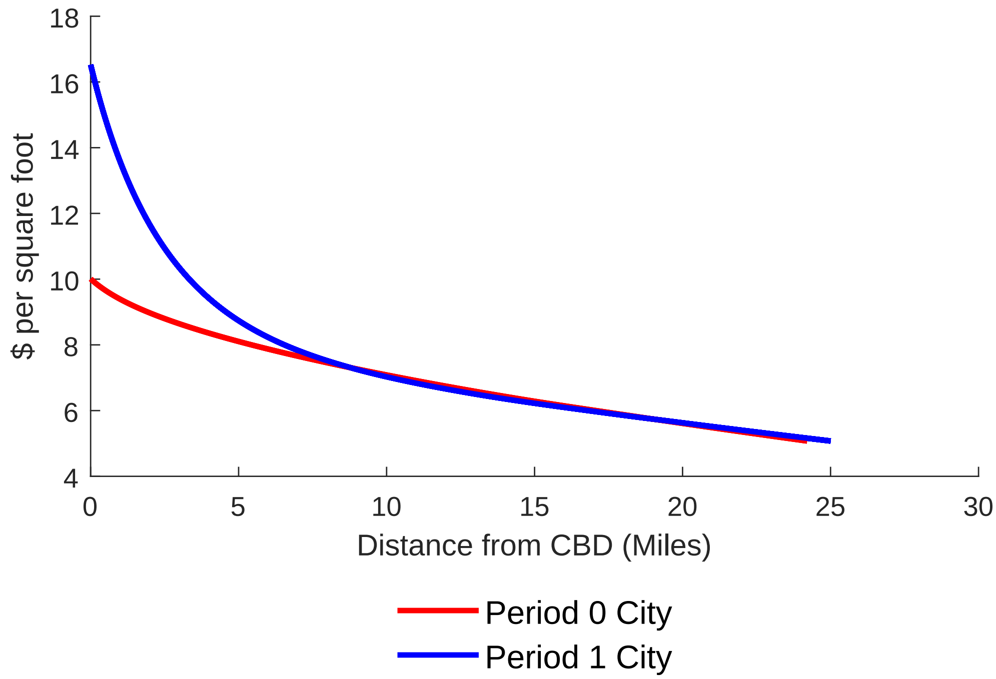
Skinner, J. (1989): "Housing wealth and aggregate saving," *Regional Science and Urban Economics*, 19, 305–324.

Tian, C. Y. (2017): "Are CAP Rates Constant within Housing Markets: Theory and Empirical Evidence," *Real Estate Finance*, 24, 64–76.

Wallace, N. E. and R. A. Meese (1997): "The construction of residential housing price indices: a comparison of repeat-sales, hedonic-regression, and hybrid approaches," *The Journal of Real Estate Finance and Economics*, 14, 51–73.

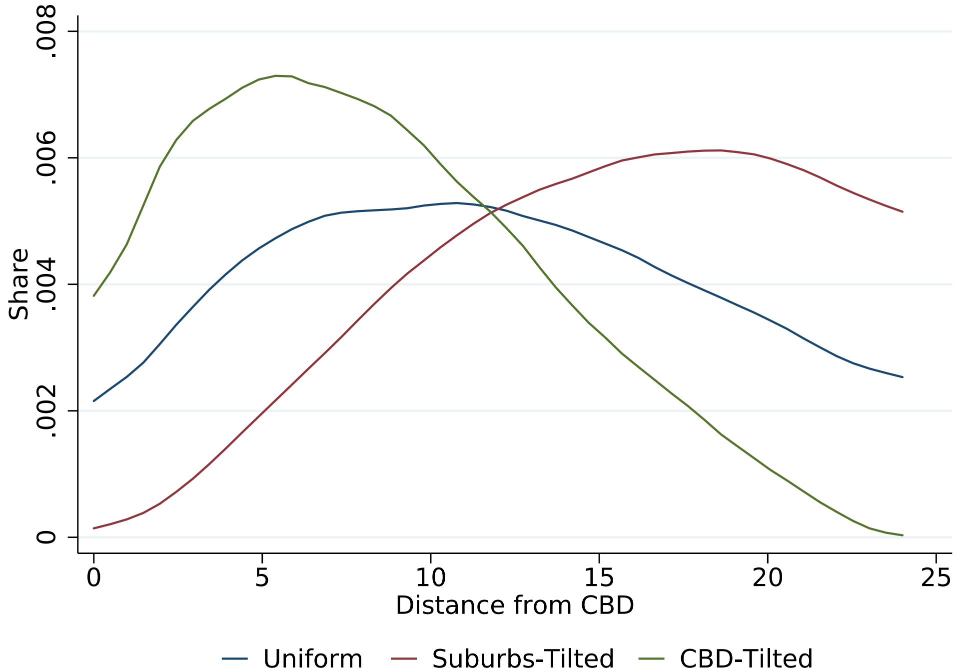
Wang, F. T. and P. M. Zorn (1997): "Estimating house price growth with repeat sales data: What's the aim of the game?" *Journal of Housing Economics*, 6, 93–118.

Figure 1: Synthetic House Price Gradients



Notes: In the simulation, housing quantity is reduced to a single dimension scaled to resemble the square feet of interior space. Accordingly, the price gradients are in terms of the price of this unit measure. The simulation model is described in the appendix.

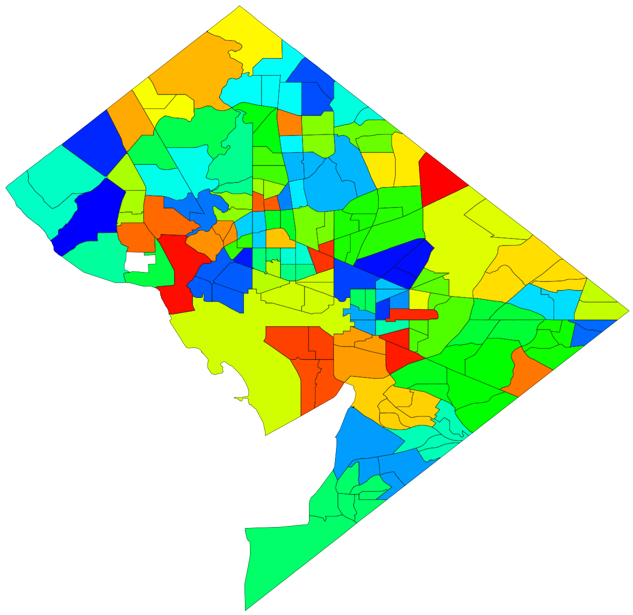
Figure 2: Sampling Assumptions from Simulation



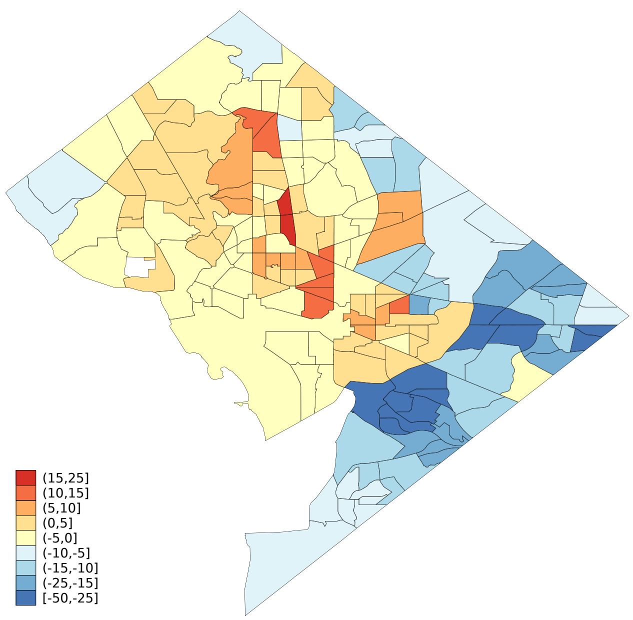
Notes: Assumptions regarding sampling are described in the text, and set to give approximately 5% of housing units in the city sampled twice (both period 0 and 1).

Figure 3: Washington, DC Supertracts in 2010

(a) Supertracts



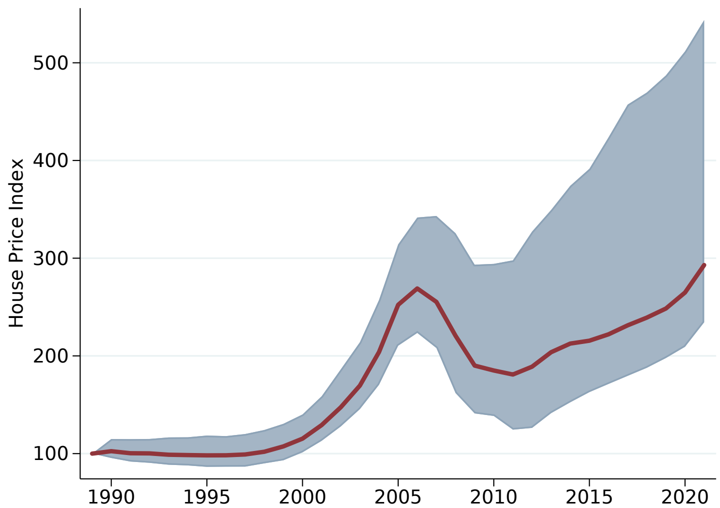
(b) Appreciation



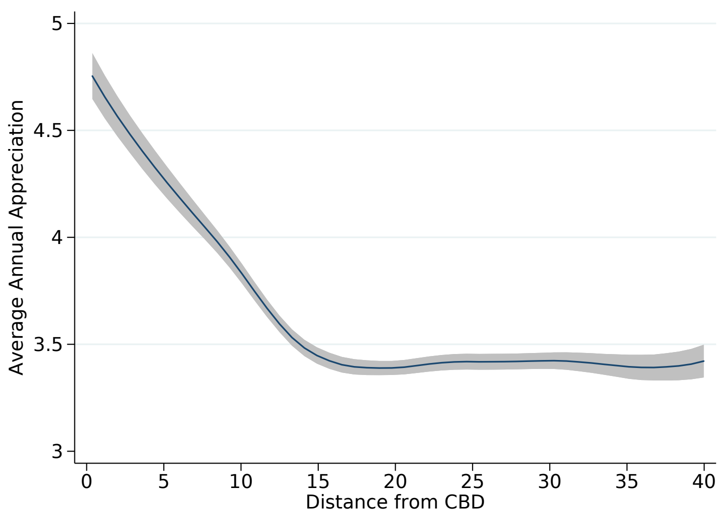
Notes: This figure presents supertracts in Washington, DC in 2010, with borders showing individual tracts. In panel (a), supertract colors are random so some adjacent supertracts may have similar colors by chance. Panel (b) shows appreciation rates for the above supertracts.

Figure 4: Washington, DC CBSA Tract-Level House Prices

(a) House Price Indices



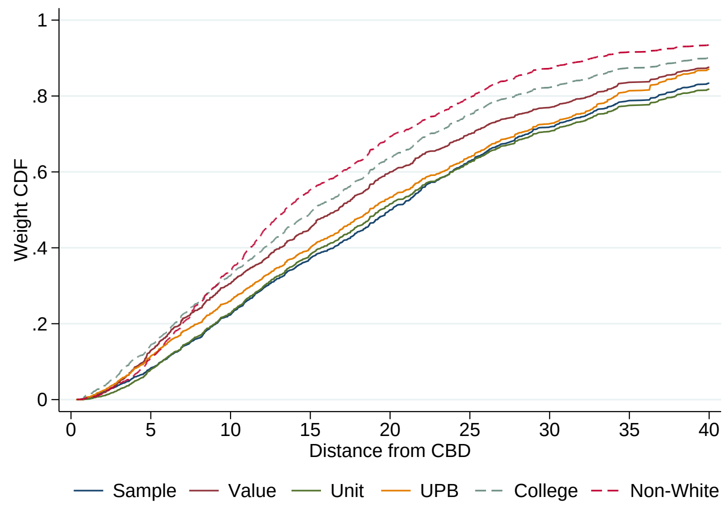
(b) The Appreciation Gradient, 1989-2021



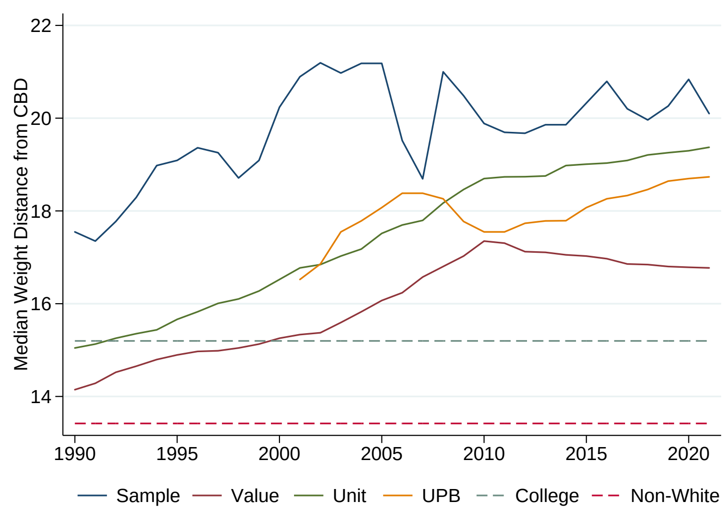
Notes: Indices are calculated using methods described in the text. Each uses the same tract-level appreciation rates calculated using the supertract approach. City-level indices are calculated by aggregating based on tract-level shares of aggregate city-level home value. In panel (a) shaded areas are the 5th and 95th percentile index values in the particular time period, and the red line is the median. In panel (b) gradients are calculated using STATA's polynomial smoother, with a bandwidth of 1 mile.

Figure 5: Washington, DC CBSA, Weights

(a) Alternative Weights, 2021

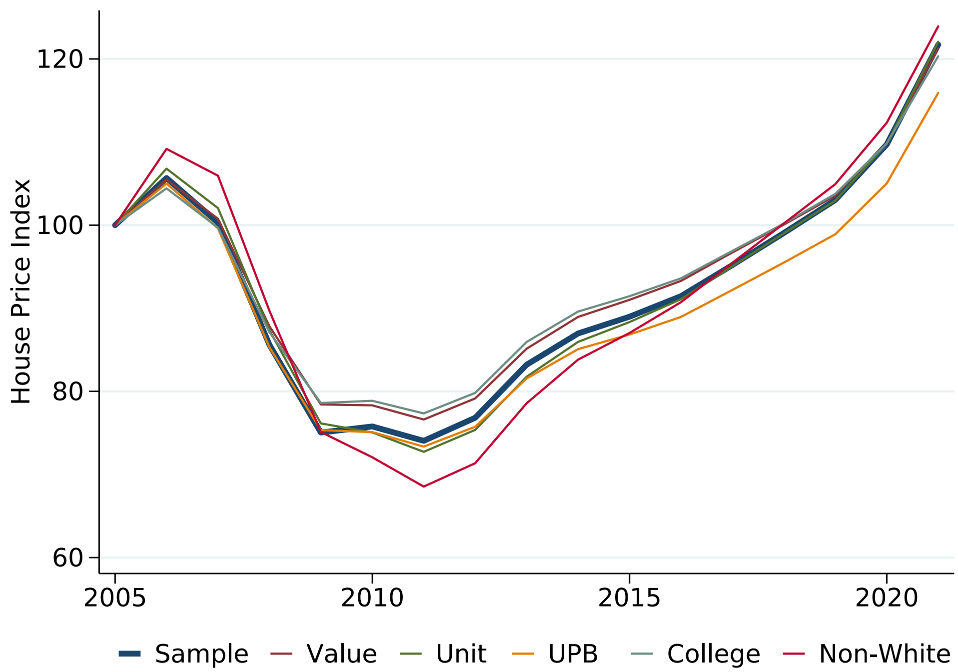


(b) Alternative Weights, Median Location over Time



Notes: The location of the median weight is calculated by ordering tracts by distance from the CBD, determining which tract crosses (or is equal to) the 50th percentile weight in a particular time period, and capturing the distance of this tract from the CBD. This value is re-calculated and tracked over time. The dotted lines indicate weights that are based on 2010 Census levels, and hence do not vary over time.

Figure 6: Washington, DC CBSA-Level Indices



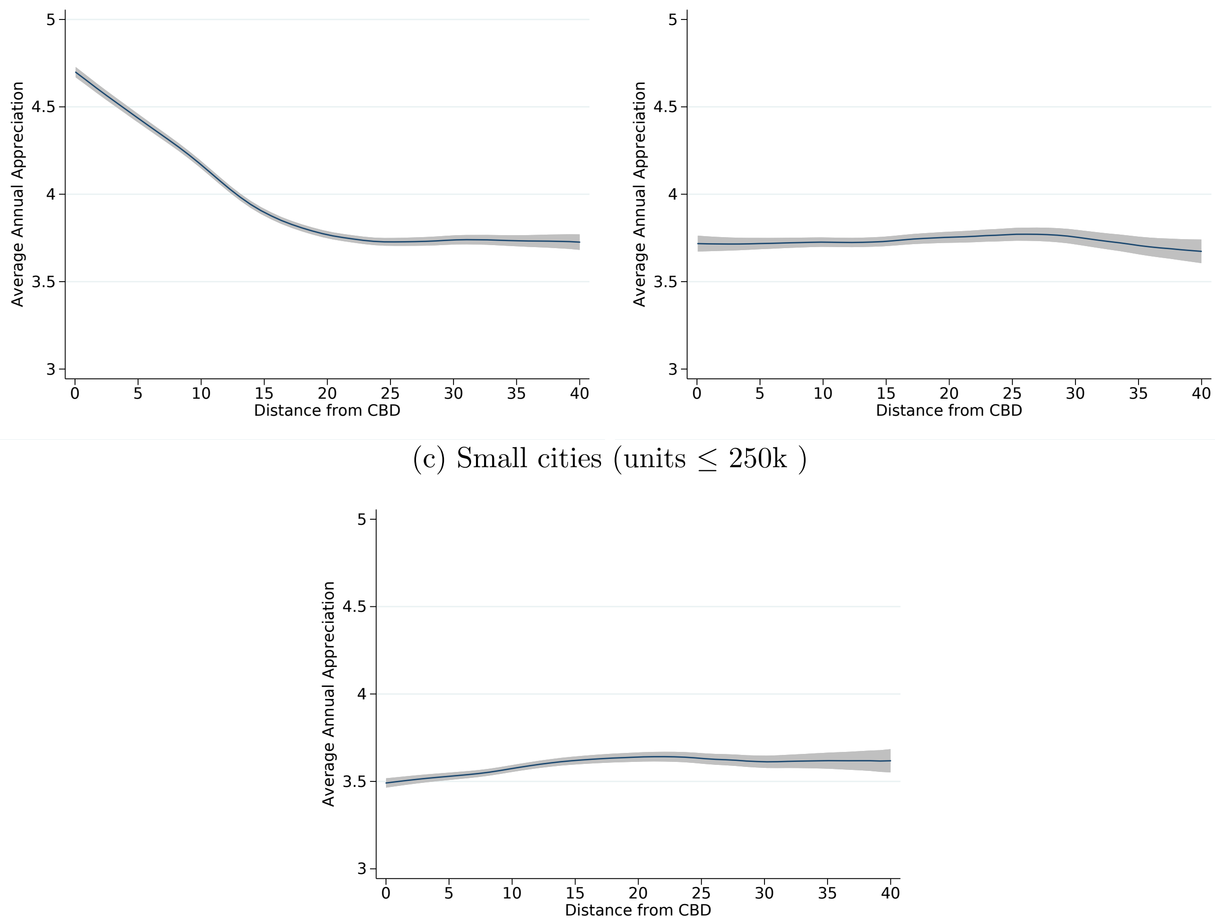
Notes: Indices are calculated using methods described in the text. Each uses the same tract-level appreciation rates calculated using the supertract approach. City-level indices are calculated by aggregating based on the tract-level within-city shares shown in the legend. See text for further details.

Figure 7: Tract appreciation gradients, 1989-2021

(a) Large cities (units >1 million)

(b) Medium cities (250k > units ≤ 1 million)

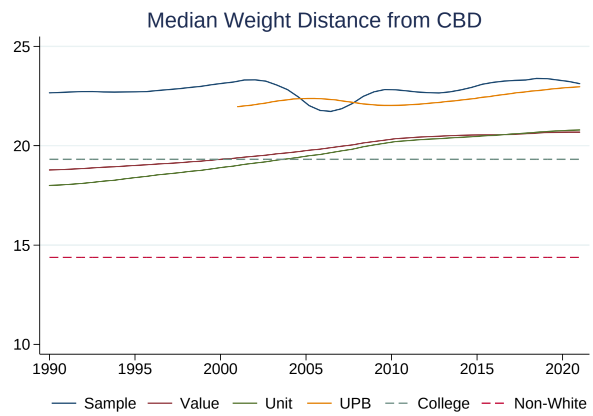
(c) Small cities (units ≤ 250k)



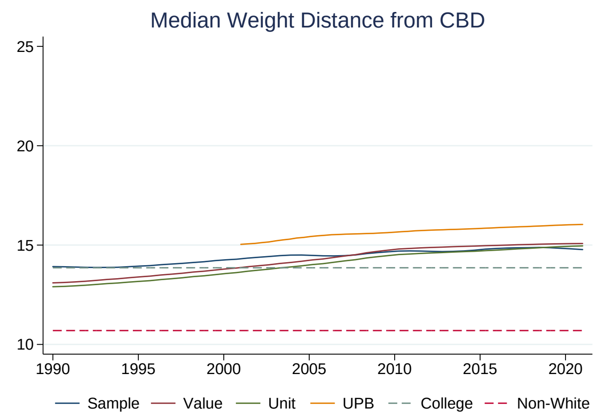
Notes: Each figure presents tract-level annual average appreciation rates between 1989 and 2021. Gradients are calculated using STATA's polynomial smoother, with a bandwidth of 1 mile. Shaded regions are the 90% confidence interval of the estimated mean at a particular value.

Figure 8: Alternative Weights, Median Location over Time

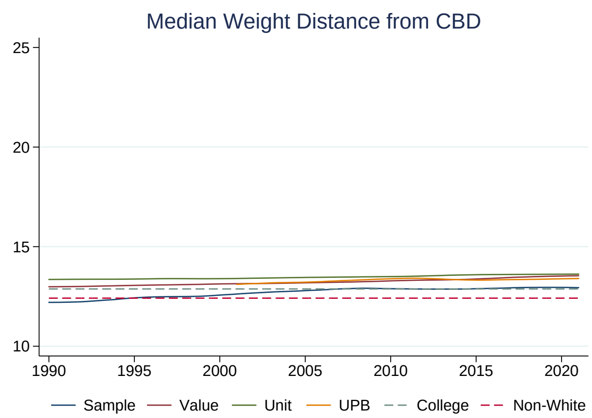
(a) Large cities (units >1 million)



(b) Medium cities (250k > units ≤ 1 million)



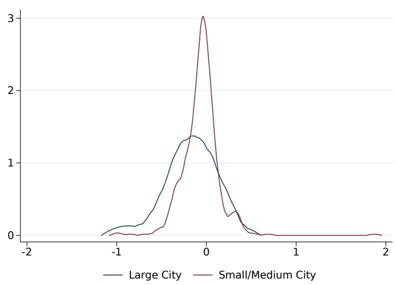
(c) Small cities (units ≤ 250k)



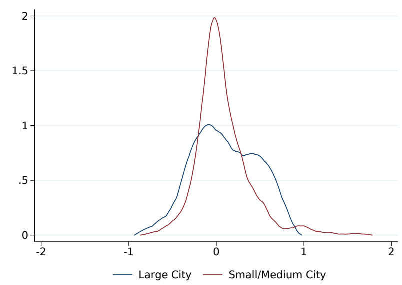
Notes: The location of the median weight is calculated by ordering tracts by distance from the CBD, determining which tract crosses (or is equal to) the 50th percentile weight in a particular time period, and capturing the distance of this tract from the CBD. This value is re-calculated and tracked over time. The dotted lines indicate weights that are based on 2010 Census levels, and hence do not vary over time.

Figure 9: Average annual appreciation differences

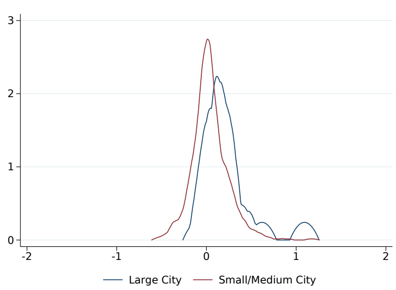
(a) 1989 through 2001



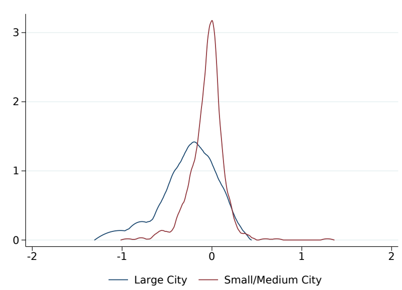
(b) 2001 through 2005



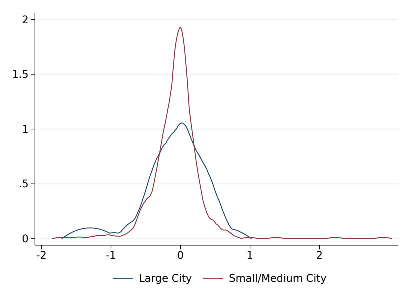
(c) 2005 through 2013



(d) 2013 through 2019

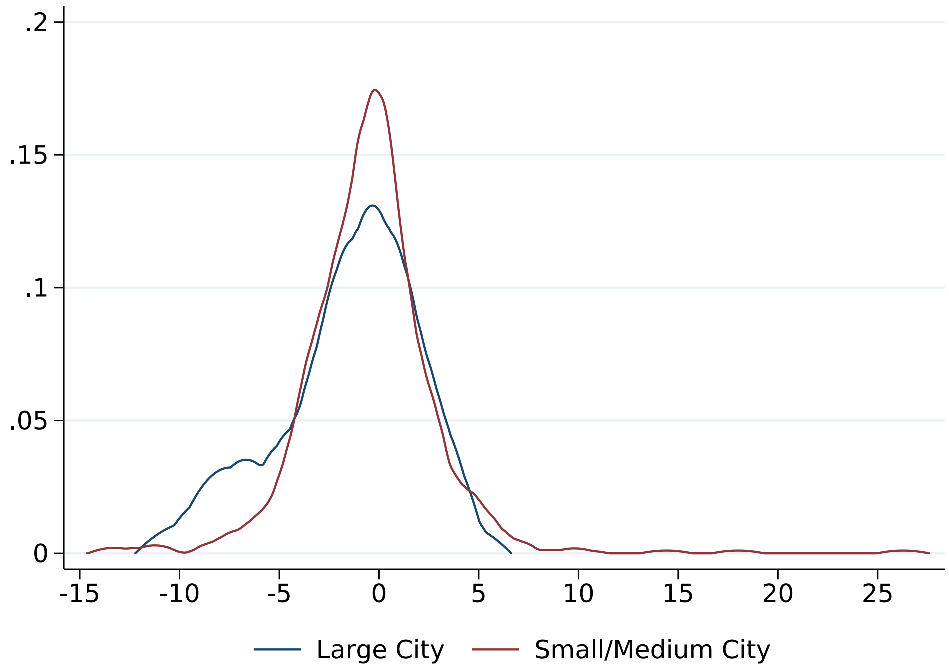


(e) 2019 through 2021



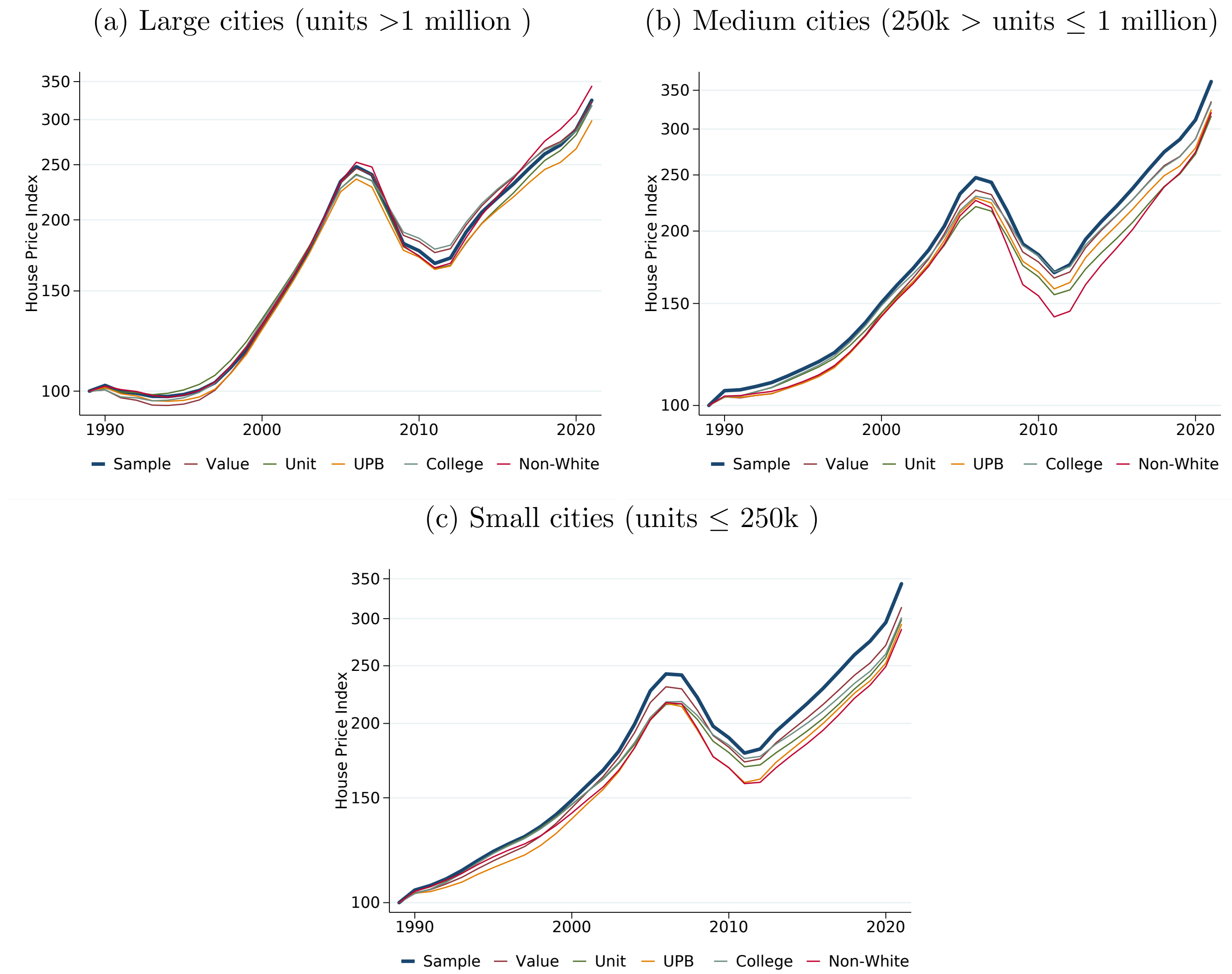
Notes: Average annual appreciation difference is defined as $100 * (\ln P_{t1}^{Sample} - \ln P_{t0}^{Sample} - (\ln P_{t1}^{Unit} - \ln P_{t0}^{Unit}))/(t_1 - t_0 + 1)$, where P^{Unit} is a unit-weighted index calculated using the repeat-sales aggregation index algorithm, P^{Sample} uses sample (half-pair) weights, and $t0$ and $t1$ define the window noted in the subfigure title.

Figure 10: Cumulative appreciation differences, 1989-2021



Notes: Cumulative difference is defined as $\ln P_{2021}^{Sample} - \ln P_{1989}^{Sample} - (\ln P_{2021}^{Unit} - \ln P_{1989}^{Unit})$, where P^{Unit} is a unit-weighted index calculated using the repeat-sales aggregation index algorithm, and P^{Sample} uses sample (half-pair) weights.

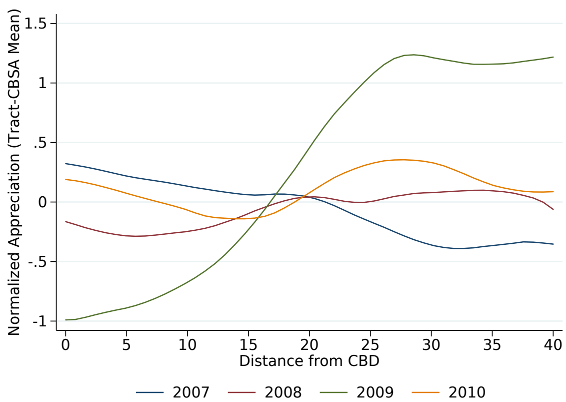
Figure 11: City Size Indices, 1989-2021



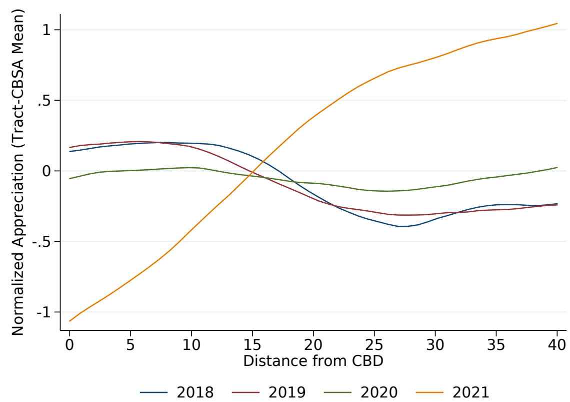
Notes: Indices in each city size category are calculated by aggregating city-level indices using weights shown in the legend. A large city is defined as a CBSA (2020 definitions from Office of Management and Budget.) with component counties where the number of housing units in 1990 was greater than 1 million. A medium city is between 250,000 and 1 million housing units. A small city is less than 250,000 units.

Figure 12: Appreciation and Sampling Gradients, Great Recession and COVID periods

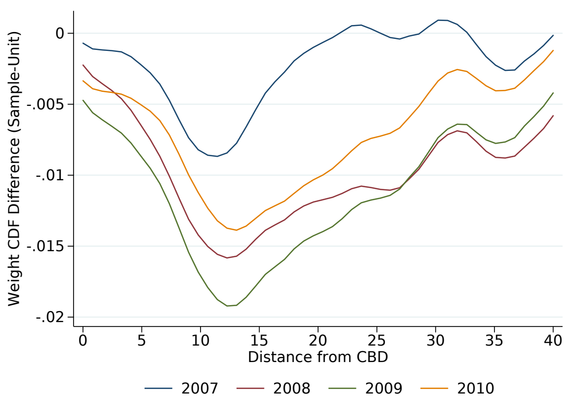
(a) Appreciation Gradient, Great Recession



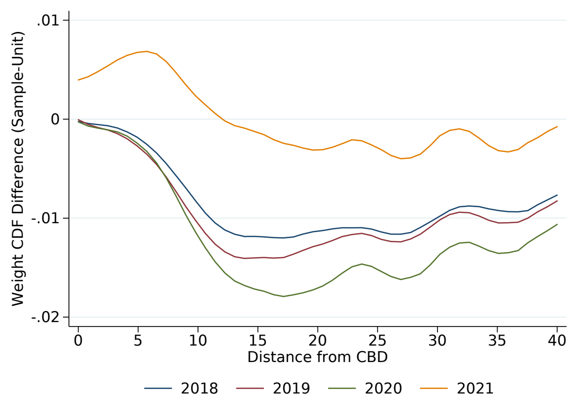
(b) Appreciation Gradient, COVID-19



(c) Sample vs Units Gradient, Great Recession



(d) Sample vs Units Gradient, COVID-19



Notes: Each figure presents tract-level gradients (scatterplot omitted). Gradients are calculated using STATA's polynomial smoother, with a bandwidth of 1 mile. The sample is all tracts in the 581 CBSAs. Panels (a) and (b) give the difference between the tract's appreciation rate and that of the CBSA in the particular year. Panels (c) and (d) give the difference between the Sample and Unit Distance-from-CBD CDFs. When the line is above 0, this indicates that homes within a particular distance are more-than-proportionally sampled in the particular year.

Table 1: Simulated City-Level Appreciation Rates

Estimator	RS Aggregation Index			BMN (1963)
Sampling	Value	Sample	Units	Sample
Population	7.8%	3.7%	6.5%	3.7%
Uniform	7.9%	3.7%	6.5%	3.7%
Suburb-Tilted	7.5%	0.7%	6.2%	0.7%
CBD-Tilted	7.8%	6.8%	6.4%	6.8%

Notes: Appreciation rates are generated by simulating period 0 and period 1 cities under identical conditions, except the period 1 city has higher income and a CBD amenity, which generates positive average appreciation. Each "tract" is defined as a 0.01 mile annulus. Sampling of transactions and prices are then performed, tracts are aggregated into supertracts, and the supertract indices are estimated and applied to the annuli. The city indices are then calculated as various weighted averages.

Table 2: Cumulative appreciation differences, large cities, Sample vs Unit weights

		Average Annual Difference					Cumulative Difference
CBSA	CBSA Name	1989-2001	2001-2005	2005-2013	2013-2019	2019-2021	1989-2021
26420	Houston-The Woodlands-Sugar Land, TX	-0.9%	-0.4%	0.2%	-0.1%	0.2%	-9.0%
19100	Dallas-Fort Worth-Arlington, TX	-0.5%	-0.6%	0.1%	-0.1%	0.1%	-7.1%
16980	Chicago-Naperville-Elgin, IL-IN-WI	-0.4%	-0.1%	0.6%	-0.6%	-1.3%	-6.3%
12060	Atlanta-Sandy Springs-Alpharetta, GA	-0.2%	-0.1%	0.2%	-0.4%	-0.5%	-4.2%
45300	Tampa-St. Petersburg-Clearwater, FL	-0.2%	0.1%	0.2%	-0.4%	-0.2%	-2.8%
33100	Miami-Fort Lauderdale-Pompano Beach, FL	-0.4%	0.0%	1.1%	-1.0%	-0.6%	-2.6%
35620	New York-Newark-Jersey City, NY-NJ-PA	0.0%	0.2%	0.1%	-0.7%	0.6%	-1.9%
42660	Seattle-Tacoma-Bellevue, WA	-0.2%	-0.1%	0.1%	0.1%	0.0%	-1.3%
33460	Minneapolis-St. Paul-Bloomington, MN-WI	0.0%	-0.1%	0.0%	0.0%	-0.1%	-1.3%
37980	Philadelphia-Camden-Wilmington, PA-NJ-DE-MD	-0.1%	0.5%	0.0%	-0.2%	-0.1%	-0.9%
19820	Detroit-Warren-Dearborn, MI	-0.4%	0.6%	0.2%	0.0%	-0.3%	-0.8%
47900	Washington-Arlington-Alexandria, DC-VA-MD-WV	-0.1%	0.4%	0.2%	-0.3%	-0.2%	-0.3%
38060	Phoenix-Mesa-Chandler, AZ	-0.3%	0.5%	0.4%	-0.4%	0.1%	0.1%
38300	Pittsburgh, PA	-0.3%	0.4%	0.2%	0.1%	-0.2%	0.4%
31080	Los Angeles-Long Beach-Anaheim, CA	0.2%	-0.1%	0.2%	-0.4%	0.3%	1.1%
14460	Boston-Cambridge-Newton, MA-NH	0.3%	-0.2%	0.0%	-0.2%	0.2%	1.8%
41860	San Francisco-Oakland-Berkeley, CA	0.1%	0.6%	-0.1%	-0.2%	0.3%	1.9%
40140	Riverside-San Bernardino-Ontario, CA	0.1%	-0.2%	0.4%	-0.3%	0.3%	2.2%
41180	St. Louis, MO-IL	0.1%	0.5%	0.0%	0.2%	-0.3%	3.4%

Notes: Average annual appreciation difference is defined for a city as $100 * (\ln P_{t1}^{Sample} - \ln P_{t0}^{Sample} - (\ln P_{t1}^{Unit} - \ln P_{t0}^{Unit}))/(t1 - t0 + 1)$, where P^{Unit} is a unit-weighted index calculated using the repeat-sales aggregation index algorithm, P^{Sample} uses sample (half-pair) weights, and $t1$ and $t0$ are in the column headers. Large cities are defined as CBSAs with over 1 million housing units in 1990.

Table 3: Within-city appreciation difference correlations, Sample vs Unit weights

	1989-2001	2001-2005	2005-2013	2013-2019	2019-2021
1989-2001	1.00	0.15	-0.10	0.06	0.19
2001-2005	0.15	1.00	0.05	0.07	-0.10
2005-2013	-0.10	0.05	1.00	-0.31	-0.25
2013-2019	0.06	0.07	-0.31	1.00	-0.07
2019-2021	0.19	-0.10	-0.25	-0.07	1.00

Notes: Average annual appreciation difference is defined for a city as $(\ln P_{t1}^{Sample} - \ln P_{t0}^{Sample} - (\ln P_{t1}^{Unit} - \ln P_{t0}^{Unit}))/(t1 - t0 + 1)$, where P^{Unit} is a unit-weighted index calculated using the repeat-sales aggregation index algorithm and P^{Sample} uses sample (half-pair) weights, and $t0$ and $t1$ are listed in the row/column headers. Correlations are calculated between appreciation difference pairs for the same city over time.

Table 4: Average Annual Appreciation Differences

	1989-2001	2001-2005	2005-2013	2013-2019	2019-2021	1989-2021
Value (Laspeyres)						
Large City	-0.196***	0.182*	-0.347***	0.403***	0.595***	-0.019
Medium City	-0.112***	0.012	-0.263***	0.287***	0.276***	-0.033***
Small City	-0.053***	0.007	-0.052***	0.049***	0.063***	-0.018***
Unit						
Large City	-0.171***	0.101	0.212***	-0.254***	-0.077	-0.047**
Medium City	-0.083**	0.158***	0.150***	-0.174***	-0.249***	-0.020
Small City	-0.058***	0.066***	0.030***	-0.041***	-0.040***	-0.015***
UPB						
Large City	0.004	0.076	-0.202*	0.519***	0.560***	0.096***
Medium City	0.033	-0.045	-0.341***	0.483***	0.445***	0.040**
Small City	0.000	0.008	-0.065***	0.076***	0.104***	0.006
College						
Large City	-0.168***	0.256**	-0.579***	0.451***	0.918***	-0.030
Medium City	-0.071***	-0.039	-0.399***	0.330***	0.424***	-0.042***
Small City	-0.020***	0.023*	-0.086***	0.079***	0.106***	-0.004
Non-White						
Large City	-0.105	0.082	0.797***	-1.195***	-0.179	-0.064
Medium City	0.172**	0.280*	0.922***	-1.071***	-1.005***	0.063*
Small City	-0.020	0.165***	0.219***	-0.171***	-0.220***	0.023***

Notes: Average Annual appreciation difference is defined for a city as $100 * (\ln P_{t1}^{Sample} - \ln P_{t0}^{Sample} - (\ln P_{t1}^{W} - \ln P_{t0}^{W}))/(t1 - t0 + 1)$, where P^W is a weighted index noted in italics calculated using the repeat-sales aggregation index algorithm ($W = Value, Unit, \ldots$) and P^{Sample} uses sample (half-pair) weights. Large cities are defined as CBSAs with over 1 million housing units in 1990. Medium cities are between 250,000 and 1 million housing units. Small cities have less than 250,000 housing units. There are 19 large cities, 50 medium cities, and 512 small cities.

A A Closed-City Numerical Urban Simulation

The model is a rendition of the standard monocentric city approach of Alonso (1964), Mills (1967), and Muth (1969), the seminal model linking housing production, land use, and housing demand in a spatial, general equilibrium framework. Muth (1975) produced the first numerical simulation of the standard urban model, beginning a tradition that includes works by Altmann and DeSalvo (1981), Bertaud and Brueckner (2005), Larson, Liu, and Yezer (2012), and many others.

Underpinning this model is the endogeneity of both producer and household behavior with regards to profits, utility, and location. Each household within a city can move freely until an equilibrium is reached where utility is the same for all households and there is no incentive to relocate. Similarly, housing producers achieve a zero profit equilibrium where there is no incentive to produce either less or more housing at each location in the city. Land is the only fixed asset and thus is the only input able to achieve economic rents. The analysis is long-run in nature, meaning prices and quantities are expressed in terms of steady-state flows. Population and earnings are fixed within the model; altering model parameters causes changes to utility as well as other outputs.

1.1 Land

The city lies on a featureless plane. An exogenous central business district (CBD) takes up the area of a circle with a radius 1 mile. Firms in the CBD demand N workers, with each job paying W wages. The city is isotropic in relation to the center of the city, meaning all variables can be expressed radially in terms of k.

Land outside the CBD is endogenously allocated to either agriculture or residential use. Agriculture pays a fixed land rent p_L^A per acre and does not affect the city in any way other than to provide a reservation rental price. Between the CBD and the agricultural hinterland is the residential district, where $k_{CBD} < k < k^*$, and k^* is the endogenous edge of the city. In the residential district, a fixed fraction of the land area is allocated to roads, $\bar{\rho}$, and housing $\bar{\theta}$. The remainder is reserved for other uses outside the model.

1.2 Housing Production

Housing services Q in each location are produced using structure (S) and land inputs (L) by a perfectly competitive constant returns industry according to a CES production function:

$$Q(k) = A\left[\alpha_1 S(k)^\rho + \alpha_2 L(k)^\rho\right]^{1/\rho} \tag{6}$$

where, α_1 and α_2 are distribution parameters, and the elasticity of substitution is $1/(1-\rho)$. Land prices are endogenously determined and structure prices are fixed and equal to unity. Housing is not durable in this model, indicating a long-run interpretation of the simulation.

1.3 Households

Household utility is CES in housing and the numeraire consumption. In addition to the numeraire good, which is purchased directly, the household also receives an additive numeraire consumption amenity based on its proximity to the CBD. In equilibrium each household achieves the reservation utility.

$$\bar{U} = [\beta_1(y(k) + a(k))^\eta + \beta_2 q(k)^\eta]^{1/\eta} \tag{7}$$

where a is the amenity, q is housing consumption, and y is the numeraire consumption good. β_1 and β_2 are distribution parameters, and the constant elasticity of substitution between housing and the numeraire is given by $1/(1-\eta)$. The amenity a is expressed as a negative exponential with the CBD at its center. The per-household amenity at radius k is thus

$$a(k) = a_{CBD}e^{-\eta k} \tag{8}$$

Each household inelastically supplies a single worker with a fixed labor supply and receives an income W. The household's budget constraint is

$$W(k) = y(k) + r(k)q(k) + T(k) \tag{9}$$

where T is the sum of both time and out-of-pocket commuting cost, r is the rental price of housing services, and q is the quantity of housing services consumed, each varying with the distance from the center of the city. The calculation of home values requires assuming a rate of market capitalization, with $V(k) = r(k)q(k)/\delta$, where δ is the per-period discount rate.

The number of households (also, housing units) in the city is N, which is equal to the integral of the density of households from the CBD to the edge of the city at radius $\bar{k}$. Density, in turn, is simply the ratio of housing production per unit of land from the housing producer's problem, divided by housing consumption per household, $D(k) = [Q(k)/L(K)]/q(k)$.

$$N = \int_{k_{CBD}}^{\bar{k}} 2\pi\theta k D(k)dk \tag{10}$$

Similar calculations are possible for the number of housing units within particular annuli.

1.4 Commuting

Commuting speeds are endogenous with respect to the commuter flow-road ratio. Roads are exogenous and distributed uniformly across the city as a fraction of available land, allowing the quantity of roads in a given annulus to be expressed as follows:

$$R(k) = 2\pi\bar{\rho}k \tag{11}$$

The commuting speed function is then:

$$\upsilon(k) = \frac{1}{\upsilon_0 + \upsilon_1 M(k)^{\upsilon_2}} \tag{12}$$

where $M(k) = \overrightarrow{N}(k)/R(k)$ is ratio of traffic to roads through annulus k. The traffic through annulus k is the total number of households N minus the number of households inside radius k, or $\overrightarrow{N}(k) = N - \int_{CBD}^{k} N(\kappa)d\kappa$. The parameters υ_0 and υ_1 are defined such that at CBD, $\upsilon(k_{CBD}) = \upsilon_{min}$, and at the edge of the city, $\upsilon(k^*) = \upsilon_{max}$ with υ_2 calibrated to achieve reasonable curvature.[29]

Commuting costs for CBD commuters include the fixed costs of owning a vehicle m_0, depreciation of the vehicle and tires linearly related to the miles traveled m_1, fuel costs of travel, and the opportunity cost of travel. The fuel cost of travel is related to the velocity of the automobile according to $G(\upsilon(k))$, and the price of gasoline p_g. The time cost of commuting is a fixed share τ of nominal income net of any compensating differentials, $\tau\bar{w}$, where $\bar{w}$ is set to the nominal wage in a baseline simulation.

$$T(k) = m_0 + m_1 k + p_g \int_0^k \frac{1}{G(\upsilon(M(\kappa)))} d\kappa + \tau\bar{w} \int_0^k \frac{1}{\upsilon(M(\kappa))} d\kappa \tag{13}$$

1.5 Solving the model

The model is solved numerically following Larson, Liu, and Yezer (2012), who rely on the methods of Muth (1975), Altmann and DeSalvo (1981), and McDonald (2009) with the addition of endogenous traffic congestion. Initialized wage and house price values at the edge of the CBD allow for the determination of a house price gradient as a function of commuting costs and the change in amenity. Then, gradients for housing demand, land

[29] This speed function is well-established in the literature and is called the "Bureau of Public Roads Congestion Function" (BPR). There are other alternatives such as the Akçelik function, but according to Skabardonis and Dowling (1997), the BPR is a fairly accurate and tractable speed function.

prices, the structure-land area ratio, and household density are found recursively.

The model must satisfy an equilibrium condition. At the radius where the population target is met, the land price per acre at the edge of the city k^* must be equal to the agricultural reservation price of land per acre. If the residential land price is too high (low), the CBD house price is decreased (increased) and the simulation is re-run.

Simulation outputs are found below under a city where in period 0 household income is \$50,000, and in period 1 it is \$70,000. The CBD consumption amenity in period 0 is zero, and in period 1 it is \$20,000. All other model parameters are identical.

B Price Index Relations

To understand the contribution of Bailey, Muth, and Nourse (1963) (BMN) requires going back further into the price index literature. Typically, price indices are calculated using *price relatives*, or prices of matched pairs in subsequent periods. This is typically easy to do for standardized commodities such as oil or agricultural goods. Jevons (1863) via Balk (2012) introduces a simple geometric average price index, $P_t^J = \prod_{i=1}^{I} \left(P_{it}/P_{i,t-1}\right)^{1/I}$ where P^J over $i \in I$ goods that is the simple average of all observed price relatives between periods t and t-1, $P_{it}/P_{i,t-1}$. This expression can be readily estimated by taking logs (denoted by lowercase) and regressing the price relatives on a constant term.[30]

$$p_{ikt,t-1} = \delta_{t,t-1} + \epsilon_{ikt,t-1} \tag{14}$$

The major innovation of BMN is that matched pairs in adjacent periods are not necessary for the construction of a price index. Rather, a large sample of matched pairs from *any* time periods (i.e. repeat sales) can be used in equation 3 to estimate the adjacent period appreciation rates $\delta_{t,t-1}$. This method increased the number of matched pairs that could be used to estimate an index by an order of magnitude, ushering in a new era of price index measurement, with notable contributions in housing, computers, automobiles, and other heterogeneous, composite commodities. Wang and Zorn (1997) highlight the BNM estimator from equation 3 is actually a weighted-average of mean appreciation rates between both adjacent and non-adjacent periods, $\bar{p}_{t\tau}$.[31] This point is important, and is the seed of the solution to the issue we raise next.

To break the pooled sample index into submarket indexes requires more care. If we let $f(n_{01}, n_{02}, n_{12}) = \frac{n_{01}(n_{02}+n_{12})}{n_{01}n_{02}+n_{01}n_{12}+n_{02}n_{12}}$ and $1 - f(n_{01}, n_{02}, n_{12}) = \frac{n_{02}(n_{01}+n_{12})}{n_{01}n_{02}+n_{01}n_{12}+n_{02}n_{12}}$, then after some algebra we can rewrite the estimated indices as

[30]Under a constant error variance between periods 0 and 1, the Jensen's inequality correction—often called a Goetzmann (1992) correction in the housing literature—drops out.

[31]In the two period case, let $n_{t\tau}$ be the number of matched pairs where the first transaction is in time t and the second is in time τ. Then,

$$\widehat{\delta_{1,0}} = \frac{n_{01}(n_{02}+n_{12})}{n_{01}n_{02}+n_{01}n_{12}+n_{02}n_{12}}\overline{p_{01}} + \frac{n_{02}n_{12}}{n_{01}n_{02}+n_{01}n_{12}+n_{02}n_{12}}(\overline{p_{02}} - \overline{p_{12}})$$
$$\widehat{\delta_{2,1}} = \frac{n_{02}(n_{01}+n_{12})}{n_{01}n_{02}+n_{01}n_{12}+n_{02}n_{12}}\overline{p_{02}} + \frac{n_{01}n_{12}}{n_{01}n_{02}+n_{01}n_{12}+n_{02}n_{12}}(\overline{p_{01}} + \overline{p_{12}})$$

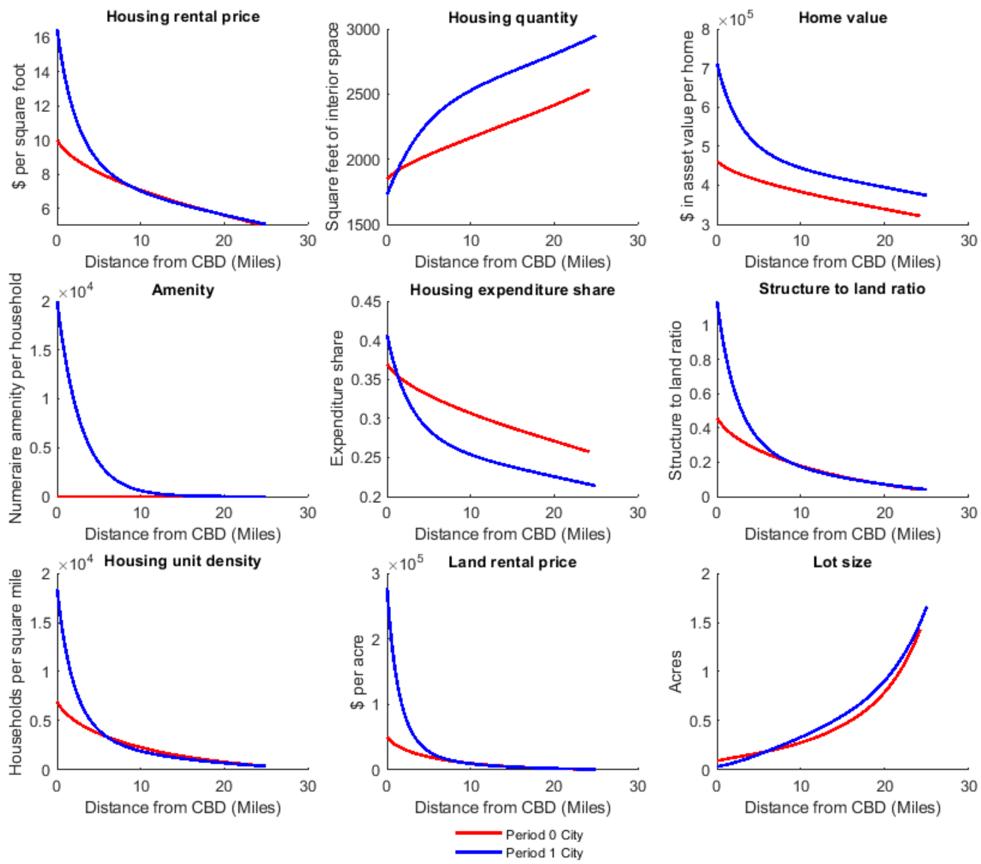
$$
\begin{aligned}
\widehat{\delta_1} &= f\overline{y_{01}} + (1-f)(\overline{y_{02}} - \overline{y_{12}}) \\
&= f\left(\frac{n_{01}^A}{n_{01}} \times \overline{y_{01}^A} + \frac{n_{01}^B}{n_{01}} \times \overline{y_{01}^B}\right) + \cdots \\
&= \left(\sum_{k=A,B} \frac{n_{01}^k}{n_{01}} \times f \times \overline{y_{01}^k}\right) + \cdots \\
&= \left(\sum_{k=A,B} \frac{n_{01}^k}{n_{01}} \times \frac{f}{f^k} \times f^k \times \overline{y_{01}^k}\right) + \cdots \\
&= \left(\sum_{k=A,B} \frac{n_{01}^k}{n_{01}} \times \frac{f}{f^k} \times \widehat{\delta_1^k}\right) + \cdots \\
&= \sum_{k=A,B} \left(\frac{\frac{n_{02}+n_{12}}{n_{01}n_{02}+n_{01}n_{12}+n_{02}n_{12}}}{\frac{n_{02}^k+n_{12}^k}{n_{01}^k n_{02}^k+n_{01}^k n_{12}^k+n_{02}^k n_{12}^k}}\right) \times \widehat{\delta_1^k} + \cdots \\
&= \sum_{k=A,B} \theta^k \times \widehat{\delta_1^k} + \cdots
\end{aligned} \tag{15}
$$

where the weight θ^k for tract k is a complicated function of the sample shares. The weight is the proportional to the inverse of the ratio of the proportion of tract k's homes sold in periods 0 and 1 relative to other time periods to the proportion of all homes sold in period 0 and 1 relative to other time periods, multiplied by tract k's share of homes that transacted in periods 0 and 1 relative to all other tracts. Intuitively, if more homes in tract k are sold in periods 0 and 1 relative to other tracts, then tract k receives a larger weight for this appreciation measure. Alternatively, if in tract k homes sold disproportionately more in periods 0 and 1 than in other time periods (relative to the average across all tracts), then θ^k will be larger.

The above derivation shows that is is possible to write a pooled index as a weighted average of the submarket indices, though the weights are very complicated. In our paper, we use each tract's proportion of half-pairs as an estimate of θ.

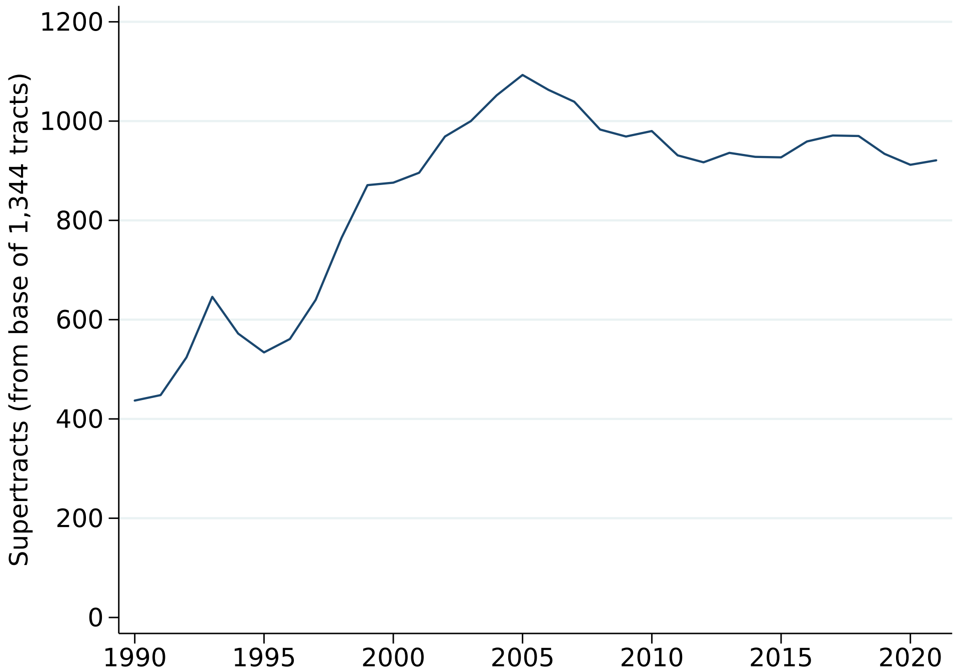
C Tables and Figures

Figure A.1: Simulation Outputs



Notes: This figure presents outputs from the simulation described in the appendix.

Figure A.2: Washington, DC CBSA, Supertract Counts



Notes: This figure presents the number of supertracts exceeding 40 half-pairs in both t and $t-1$ for each current period t, shown in the figure.

Table A.1: Comparison with the BDL Dataset

Decade	Tracts (this paper)	Tracts (BDL Dataset)	Tracts (Intersection)	Appreciation Correlation
1990-1999	63,122	23,664	17,494	0.43
2000-2009	63,122	46,003	33,517	0.76
2010-2019	63,122	53,181	37,413	0.51

Notes: The comparison dataset is the "BDL" dataset from Bogin et al. (2019a), version 2021. Appreciation is measured as annual log-differences.