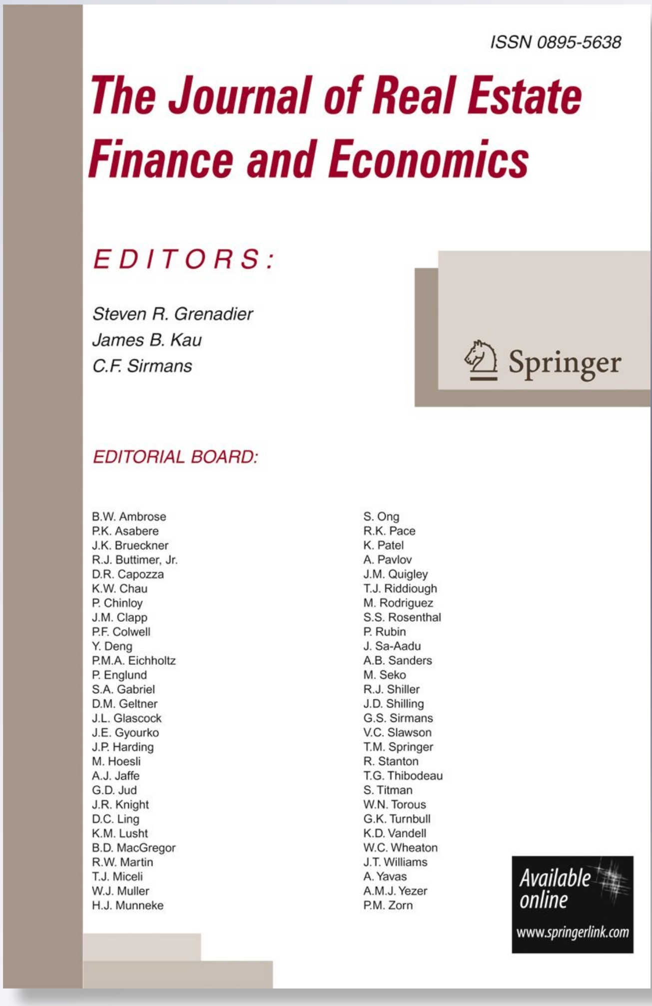
Home Price Risk, Local Market Shocks, and Index Hedging

DeForest McDuff

The Journal of Real Estate Finance and Economics

ISSN 0895-5638
Volume 45
Number 1

J Real Estate Finan Econ (2012)
45:212-237
DOI 10.1007/s11146-010-9255-2



Springer

J Real Estate Finan Econ (2012) 45:212–237
DOI 10.1007/s11146-010-9255-2

Home Price Risk, Local Market Shocks, and Index Hedging

DeForest McDuff

Published online: 23 May 2010

Abstract All real estate markets are local, or so the conventional wisdom goes. But just how local is local? I address this question empirically using over 75,000 repeat-sales transactions from a large suburban county of Washington D.C.. I construct and evaluate a variety of local home price indices defined by geography, price, and home type. I also calculate "house-specific" indices using locally weighted regressions with maximized kernel bandwidths. On the whole, local indices add a moderate amount of explanatory power relative to metropolitan indices. In my sample, the metropolitan index explains 50–75% of the variation in home price shocks, and local indices add 3–7% more. In an index hedging framework, homeowners should be willing to pay 5–10% to hedge with a local index versus a metropolitan index alone.

Keywords Housing · Home prices · Local markets · Hedging · Home price index

Introduction

All real estate markets are local, but how local is local? I target two aspects of this question using over 20 years of home transaction data from a Washington D.C. suburb.

First, I determine at what level home price shocks occur. I construct a variety of local housing indices using a repeat-sales methodology and evaluate their predictive power for individual home sales. In doing so, I learn about the size of local markets and the dimensions over which they are defined, including the relative importance of geography, home type, and price.

D. McDuff
Quant Economics, Inc., San Diego, CA, USA
e-mail: RDM@QuantEconomics.com

D. McDuff (✉)
Industrial Relations Section, Firestone Library, Princeton University, Princeton, NJ 08544, USA
e-mail: rmcduff@Princeton.edu

Table 1 Maryland 5-year migration rates from the 2000 U.S. Census

	Montgomery County, MD		State of Maryland	
Same house	428,462	50.94%	2,752,061	55.43%
Same county	189,097	22.48%	1,085,423	21.86%
Different county, same state	51,754	6.15%	465,100	9.37%
Different state	115,805	13.77%	514,875	10.37%
Elsewhere	55,967	6.65%	147,307	2.97%
Total	841,085	100.00%	4,964,766	100.00%
Population 5 years and over, 1995	841,085		4,964,766	
Population 5 years and over, 2000	813,460		4,945,043	
Net migration	(27,625)	−3.28%	(19,723)	−0.40%

The above table displays 5-year migration rates for Montgomery County, MD and the State of Maryland according to the 2000 United States Census

Second, I assess the value-added of local home price indices. I estimate a variance decomposition of home price shocks by city, local, and idiosyncratic components in order to determine the fraction of price movements attributable to each level. I frame the analysis in terms of a homeowner hedging his home price risk and ask how much he would be willing to pay to hedge with a local index versus hedging with a metropolitan index alone.

Results suggest a moderate value-added of local home price indices. In my sample, the metropolitan-level index explains 50–75% of the variation in home price movements, depending on the time interval between sales. Local indices add an additional 3–7%. A representative homeowner would be willing to pay 5–10% more to hedge with a local index versus a metropolitan index alone.

An important contribution of the paper is the development of optimal "house-specific" indices using locally weighted regressions.[1] I construct these indices by defining a local market centered at the home itself and extending outward to the nearest N homes. I optimize the kernel bandwidth over continuous distance measures of geography, home type, and price.

One advantage of using continuous distance measures is that the size of the market is not predetermined but is estimated from the data. Too large a bandwidth causes the market to be too heterogeneous, whereas too small a bandwidth contains too few transactions to construct a predictive index. I select the kernel bandwidth that maximizes the predictive power for actual home transaction prices.

Maximizing over the kernel bandwidths suggests quite a narrow definition of local housing markets. The optimal geographical index, for example, covers just 10 square miles and contains only 3% of the county residences. Regression analysis

[1] Applying locally weighted regression techniques to home transaction data is not, by itself, a novel approach. See McMillen (2004) for an example and relevant references. To my knowledge, however, it has never been used to construct "house-specific" price indices, where the weighting scheme is unique to each home.

indicates that geography, home type, and price all add predictive power to local home price indices. The best local index is created by considering all three dimensions simultaneously.

Local housing markets are important given the high frequency of local moves. Table 1 presents migration statistics for Montgomery County, MD, the source of the home transaction data used in this paper. Between 1995 and 2000, nearly one-quarter of all households moved within the county versus just one-seventh moving across state lines.

A homeowner seeking to hedge home price risk benefits from a local index in two ways. First, a local index has less *basis risk* than a city-level index. Basis risk arises from the idiosyncratic portion of a home price shock that cannot be hedged with a market index. The higher the correlation between one's home and the index, the more a homeowner will hedge and the better the hedge will perform. Second, local indices allow homeowners to hedge *within-city* moves, which is not possible with city-level indices alone.

Ultimately, the value of local market indices depends upon the statistical properties of home prices in a given area. That is, what fraction of home price movements is attributable to the city component versus the local component versus the idiosyncratic component? The value of considering local markets is greater when the local component of home price shocks is larger.

In this paper, I find local markets to be of moderate importance. While local markets certainly remain a valuable consideration for home prices, their relevance is overshadowed by the larger influence of the city market component. Importantly, this paper examines the statistical properties of home prices in one metropolitan area only. While the methods can be generalized and applied to additional cities, caution is warranted in the broad applicability of this result.

The remainder of the paper proceeds as follows. "Literature" reviews the relevant literature. "Data" discusses the data. "Methodology" describes the methodology. "Results" presents the results. "Conclusion" concludes.

Literature

Academic literature studying home price indices dates back to Bailey et al. (1963), who first proposed a repeat-sales home price index calculation in a regression framework. Home price indices were popularized in the late 1980s when Case and Shiller (1987, 1989) refined the repeat-sales empirical methodology and estimated home price indices with home transaction data for several U.S. cities. The Case-Shiller methodology has subsequently become a generally accepted practice for estimating home price indices.

The repeat-sales methodology estimates home price appreciation for a given market by examining the same house sold in two subsequent periods. Implicit in the methodology is a constant-quality assumption, that is, that the home does not change in quality or characteristics over time. I review this methodology in more detail in "Methodology".

Since 2006, Standard & Poor's has published Case-Shiller home price indices for 20 U.S. Cities each month for futures trading on the Chicago Mercantile Exchange

(CME). Shiller and Weiss (1999) proposed a variety of risk reduction financial instruments—in particular, futures, options, and event-triggered derivatives—that might one day allow homeowners and housing-related business to hedge home price risk with such markets. As of this writing, the CME housing futures markets trade with low volume, leaving home price risk reduction largely a vision for the future.

An alternative approach to estimating home price indices uses hedonic regressions of home prices on home characteristics. Such regressions include time dummies or allow time-varying characteristic coefficients to define a price index. Some recent research has aimed to combine repeat-sales and hedonic approaches into single, hybrid-version indices.[2]

Another body of literature relevant for this paper focuses on housing submarkets, or smaller groupings of homes within a broader market. For example, a housing submarket might be a city within a country or a locality within a metropolitan area. Housing submarkets are important for understanding how home prices evolve over time and space.

Literature on housing submarkets generally takes one of two approaches. The first approach is to define submarkets in an "intelligent" way, with a priori information, perhaps by dividing up a region by census tract, school district, or zip code. Goodman and Thibodeau (2003) use all of these divisions—including a more flexible hierarchical model—to improve the predictive power of hedonic regressions of home prices on home characteristics.

Several studies extend this methodology by recognizing that housing submarkets span other home characteristics besides geography. For example, Goodman and Thibodeau (2007) point out that homes might be more appropriately grouped together based on dwelling size, price per square foot, or some other combination of home characteristics compared to using geography alone, again showing that such groupings improve hedonic price predictability.

The second approach provides more flexibility by allowing the data to statistically define submarkets. Dale-Johnson (1982), Maclennan and Tu (1996), and Bourassa et al. (2003) use factor analysis and clustering techniques to group common homes together. These papers estimate hidden factors of home clusters from data on home characteristics and/or sales transactions, organizing homes into subgroups based on similarity in price, land area, square footage, location, etc.

While I draw upon ideas from the literature, I take a novel approach on several dimensions. For calculating home price indices, I rely primarily on the Case-Shiller repeat-sales methodology. However, I use locally weighted regression techniques to vary the weights based on geography and other home characteristics. In other words, I incorporate cross-sectional characteristics typical of a hedonic regression into the repeat-sales methodology.

A large part of my methodology includes the construction of house-specific home price indices. To my knowledge, using locally weighted repeat-sales regressions in this context has not been done before. The combination of a repeated-sales technique

[2] See Bourassa et al. (2006) for overview and references.

and locally weighted regressions takes a priori information on home characteristics and allows the data to define the exact size of the market. In this way, I combine ideas from both approaches in the housing submarket literature.

Whereas previous approaches define housing submarkets based on cross-sectional characteristics, I implicitly group homes together based on *price co-movement*. Price co-movement is a useful variable for grouping homes because it identifies which homes are subject to the same supply and demand forces. Home A and Home B need not necessarily be on the same street or be constructed from the same material to co-move in price. In essence, I use price co-movement to define and ultimately estimate the size of a local housing market.

Data

This paper uses data from the Maryland Property Tax Assessment Database maintained by the Maryland Department of Assessments and Taxation. The State of Maryland uses the database for property tax assessments. The database includes cross-sectional home characteristics used for tax assessment purposes and the three previous sales transactions for each home.

I focus the analysis on Montgomery County, MD, the largest county in Maryland by population. The county is home to approximately 900,000 residents and 250,000 single-family homes. The county has substantial variation in home price and type. Sections of the county near the Washington D.C. border are mostly urban while other sections are quite rural.

The final data set contains 227,554 single-family homes and 75,947 repeat-sales pairs from 1985–2006. Following the lead of Case and Shiller (1987), I exclude sales which are not competitive or occur over too short a time interval.[3]

Table 2 displays the summary statistics for the cleaned dataset of repeat-sales transactions used throughout the paper. Cross-sectional variables in the data include land value, building value, assessment value, square footage, land area, year built, construction grade, and condition. I assign longitude and latitude for each home using an ArcGIS address-matching algorithm. Address matching identifies the exact coordinates of 88.1% of the addresses.

[3] The tax assessment database contains the price and date of the three previous transactions for each home. First, I link together consecutive sales at the same residence. This produces 121,210 repeat-sales pairs with complete price and date information. I consider only those transactions after 1985. I drop all sales that are not “arms-length,” indicating a non-competitive sale. For example, a sale from one family member to another would not be recorded as arms-length. I drop all repeat-sales that occur over less than 12 months since these are likely to be distressed sales. Next, I calculate an annualized return for each repeat-sales pair and drop observations beyond one-and-a-half standard deviations of the annual return distribution mean (2.6% of the sample) to eliminate homes which are most likely to have changed in quality. Finally, I drop all observations that cannot be located by address. The final sample contains 75,947 observations.

Because the data contain a maximum of three transaction per home, there is some concern of missing sales if a home has sold four or more times. A frequency plot suggests that this is unlikely to cause problems. Possible “missing” sales only affect 10% of homes built after 1995 and 20% of homes built after 1985. I calculate all indices from 1985–2006. Since the S&P/Case-Shiller indices are published only as early as 1987, I input the index values for these years using home appreciation in Montgomery County.

Table 2 Summary statistics

Variable	Units	*N*	Mean	SD	Min	Max
Land value	($1,000s)	75,947	274.2	144.5	15.0	2440.0
Building value	($1,000s)	75,947	245.2	174.4	23.3	4416.5
Total assessment	($1,000s)	75,947	519.4	286.7	87.2	6076.0
Square footage	(1,000s of sq ft)	75,947	1.80	0.86	0.42	17.65
Land area	(acres)	75,944	0.26	0.55	0.00002	30.45
Year built	(year)	75,947	1,974	18.4	1,900	2,006
Construction grade	(1=worst, 9=best)	75,929	4.39	0.75	2	9
Condition	(1=worst, 6=best)	75,841	3.07	0.26	1	5
Longitude	(degrees)	75,947	−77.14	0.10	−77.47	−76.90
Latitude	(degrees)	75,947	39.10	0.07	38.94	39.33
Price, 1st Sale	($1,000s)	75,947	223.5	165.8	2.0	5000.0
Price, 2nd Sale	($1,000s)	75,947	356.0	257.5	14.8	5325.0
Quarter, 1st Sale	(1=1,985-Q1)	75,947	37.5	21.3	1	84
Quarter, 2nd Sale	(1=1,985-Q1)	75,947	67.4	14.3	22	88
Partitions		N	Mean #	SD #	Min #	Max #
District		11	6,904	5,143	2,064	16,987
Zip Code		28	2,712	1,706	1,100	8,413
Home Type		11	6,904	1	6,903	6,906
Price Band		11	6,904	6	6,891	6,920

The repeat home sales from Montgomery County, MD come from the Maryland Department of Assessments and Taxation tax database. The database includes the 2006 tax assessment, a set of home characteristics, and the three previous sales transactions for every home in the county. Repeat sales pairs are dropped if they are closer than 12 months together, have an annualized price appreciation outside 1.5 standard deviations of the appreciation mean, or are missing longtitude/latitude data. The partition assignments are as follows. The District partition is provided in the data, with each district corresponding to roughly two high school districts. The Zip Code partition is also in the data, although the less common zip codes are lumped with the nearest common zip code to avoid calculating indices with too few homes. The Home Type and the Price Band partitions are constructed based on the first principal components over home characteristics (excluding location) and fitted prices from a hedonic price regression, respectively. The Home Type and Price Band partitions have eleven categories in order to match the econometric power of the District partition given by the data

I divide the sample into four different local market partitions based on district, zip code, home type, and price band. The 'District' partition is based on eleven districts coded in the data. These districts often overlap with the county's 19 high school districts, though the high school districts themselves are not directly coded.[4] The database also contains 'Subdivision' codes which I use to cluster standard errors. Figure 1 displays the full sample of 200,493 geocoded homes as well as the imputed district boundaries.

[4] I attempted to assign high school districts to the sample based on official maps from the Montgomery County Public Schools web site, calibrating longitude/latitude coordinates into computer pixel coordinates with Google Maps. The result was a noisy high school district variable with less predictive power than the 'District' variable provided in the data. Still, the two measures had substantial overlap. In the end, I use the coded district variable.

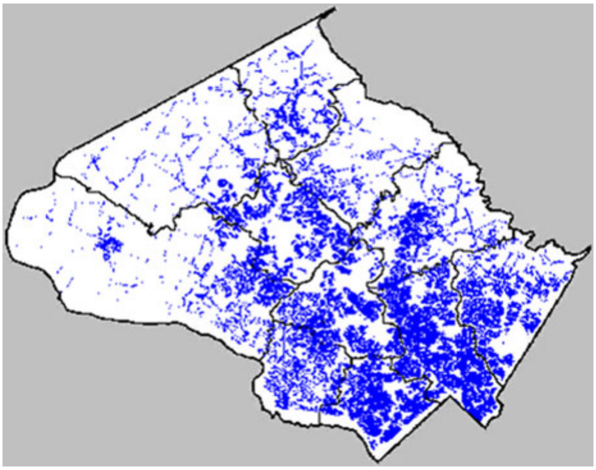


Fig. 1 Geocoded homes in Montgomery County, MD. The above figure plots the approximate locations of 200,493 single-family homes in Montgomery County, MD. The 'District' boundaries are drawn in black. Washington D.C. is located directly to the southeast of the county

The 'Zip Code' partition is created from zip code identifiers also in the data. I collapse small zip codes into larger ones based on location and create a full partition based on 30 zip code submarkets. The 'Home Type' partition is constructed based on the first principal component of the following home characteristics: square footage, land area, year built, construction grade, type of structure, number of stories, type of exterior, and maintenance condition. The homes are divided into eleven equally sized partitions based on the value of this principal component.[5] The 'Price Band' partition is based on the fitted values from a hedonic price regression on a similar set of characteristics used for the principal components.[6]

The full data contain over 20 years of home transactions. Figure 2 shows the inflation-adjusted price history for Washington D.C. and Montgomery County over the sample period. The data span an entire real estate cycle, so that there are periods of negative as well as positive price appreciation.

[5] I create eleven submarkets for both the Home Type and Price Band partitions to make them comparable to the District partition coded in the data. Experimenting with other sizes suggests that anywhere from 5 to 15 submarkets yields similar results.

[6] The hedonic model regresses price at the time of sale on all available home characteristics for the 24,352 homes sold in 2005 and 2006. The right-hand side variables include: ln(square footage), ln(land area), (year built), (year built)^2, and dummy variables for quarter of sale, construction grade, type of structure, number of stories, type of exterior, and maintenance condition. Districts and zip codes are intentionally omitted in order to make the price prediction over home characteristics without using geography. I use the estimated model to predict the value of every home in the sample for Q4-2006. The regression has an R^2 equal to 0.515 (0.558 for a regression that includes geographical dummies), which is less explanatory power than even the City Index in Table 4. As a validation check, the correlation between the predicted home prices and the official tax assessment value used by the State of Maryland is 0.826 for the homes included in the regression and 0.868 for the full sample. As before, the final partition also contains eleven submarkets.

Methodology

Local Home Price Indices

The most fundamental computation in this paper is a home price index regression, proposed by Bailey et al. (1963) and popularized by Case and Shiller (1987). The computation uses variation in time and price of repeated home sales to estimate average home price appreciation for a market over time.

Estimating a home price index begins with a simple model of home prices, where the price (in logs) of any home i in market j at time t is the sum of an individual component (α_i), a market component (β_{jt}), and an error term (ε_{ijt}):

$$\log P_{ijt} = \alpha_i + \beta_{jt} + \varepsilon_{ijt}. \tag{1}$$

Log price changes between t_0 and t_1 therefore do not contain a home i fixed effect:

$$\Delta \log P_{ij(t_0,t_1)} = \log P_{ijt_1} - \log P_{ijt_0} = \beta_{jt_1} - \beta_{jt_0} + \varepsilon_{ijt_1} - \varepsilon_{ijt_0}. \tag{2}$$

The market component time series (β_{jt}) can be estimated with repeat sales pairs in a regression framework. With N homes and T time periods, the dependent variable is a $N\times 1$ matrix of log price changes and the independent variable is a $N \times T$ matrix of sales periods:

$$\underbrace{\Delta \log P}_{N\times 1} = \underbrace{Z}_{N\times T} \times \underbrace{\beta}_{T\times 1} + \underbrace{\varepsilon}_{N\times 1}. \tag{3}$$

Each row of Z contains -1 for the first sale, $+1$ for the second sale, and 0 for all other time periods:

$$Z = \underbrace{\begin{bmatrix} 0 & -1 & 0 & \dots & 0 & 1 & 0 & \dots & 0 & 0 & 0 \\ -1 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 1 & 0 \\ \dots & & & & & & & & & & \\ 0 & 0 & 0 & \dots & 0 & -1 & 0 & \dots & 0 & 0 & 1 \end{bmatrix}}_{N\times T}. \tag{4}$$

The regression estimates β as a $T\times 1$ vector representing the home price index over the time period covered by the data.[7]

[7] Case and Shiller (1987) point out that sales pairs with longer times between them will have larger errors, on average, if the error term (ε_{ijt}) has a random walk component. They correct for this heteroskedasticity by estimating a second-stage regression of the squared error terms on the time between sales and rerun the first-stage regression using the square roots of the fitted values from the second-stage as weights. I adopt the convention throughout the paper in order to effectively down-weight repeat sales over longer periods of time. Still, the results are not sensitive to this adjustment.

There is some discussion in the literature regarding index revision and contract settlement using repeat-sales indices in financial markets. Index revisions occur when previously published index estimates are revised based on new data. See Clapham et al. (2005), Baroni et al. (2008), and Deng and Quigley (2007) for details and references. Although revision biases may be substantial in some contexts, I choose not to address them directly as they distract from the ultimate aim of examining within-market home price distributions. Still, the implication is that this paper estimates slightly more precise local market indices than might be possible in real time.

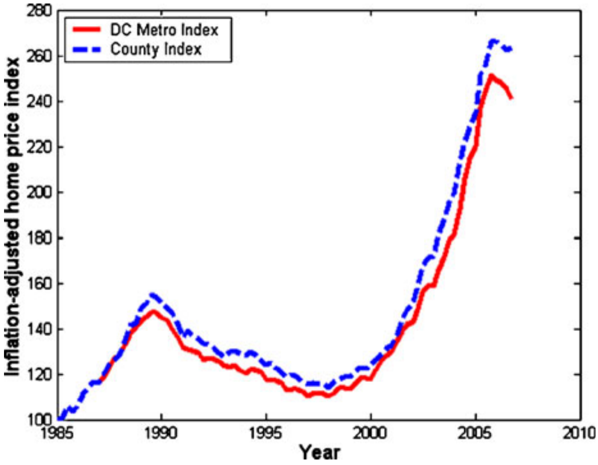


Fig. 2 Inflation-adjusted home price indices, 1987–2006. The above figure plots inflation-adjusted home price indices for the Washington D.C. Metropolitan Area and Montgomery County, MD. The former is published by Standard & Poor's and the latter is calculated using the full data set of 75,497 repeat home sales. Both are adjusted by the Consumer Price Index as published by the Bureau of Labor Statistics. As seen in the figure, the data span an entire housing cycle, for which price peaks occurred in 1989 and 2006

For a given market, I perform a variance decomposition by regressing changes in log home prices on changes in log home price indices:

$$\Delta \log P_i = \alpha + k^{city}\Delta\beta_i^{city} + k^{county}\Delta\beta_i^{county} + k^{local}\Delta\beta_i^{local} + \varepsilon_i, \qquad (5)$$

where β_i^{city} denotes the Washington D.C. Metro Index, β_i^{county} denotes the County Index, and β_i^{local} denotes the relevant Local Index (all in logs). Each observation represents the price change between two consecutive sales at home i on the left-hand side and the index change over the same time interval on the right-hand side. Finally, the k's represent the regression coefficients.

In order to isolate the effects of each index, I subtract the city and county indices from the local indices on the right-hand side. That is, the local index represents the movement in the local market above or below the broad market.[8] This improves the interpretability of the regression results since local markets are naturally correlated with the broader markets.

The explained variation in prices represents market risk which can be hedged, and the unexplained variation in prices represents idiosyncratic home price risk which cannot. In world where all home prices move one-for-one (in percentage terms) with the index, all price movements represent market risk and the regression has a perfect fit. In a world where price movements within a distribution are entirely random, the index reveals only the mean price change and the regression has no explanatory power.

[8] I also experimented with regressing the local index on the county and city indices and using the residual as an orthogonal market movement measure. The results were similar yet more difficult to interpret. The advantage of subtracting the county and city indices is that the regression results can be easily interpreted as if the full county and city indices were included.

Performing a variance decomposition using Eq. 5 allows for comparisons between indices. Local indices with more information orthogonal to broad market indices will add more fit to the regression. The greater the explained variance from the local index, the greater the value of a local index. The coefficients indicate the best composite index as a weighted sum of the raw indices. Such an index has identical predictive power to having all indices separately.

In order to avoid a spurious, mechanical relationship in the above regression, I utilize leave-one-out cross-validation to separate the index calculation from its evaluation. In other words, each house-specific index is calculated using the nearest N homes, *not including the home of interest*. I use this technique throughout—even the full sample index is calculated N times, each instance leaving out a single home—so that indices are never evaluated with the same homes used to create them.[9]

I construct and evaluate two main types of home price indices: partition indices and house-specific indices. Partition indices are constructed with Eq. 3 using a subsample based on one of four variables: district, zip code, home type, and price band. The methodology for house-specific indices is reviewed in the following section.

House-specific Indices

The creation of "house-specific" indices is one of the main innovations in this paper. House-specific indices utilize locally weighted regressions to estimate a price index for a specific home in the sample.

Locally weighted regressions, pioneered by Fan (1992), minimize weighted mean-squared errors from regressions that weight observations based on a kernel, a kernel bandwidth, and a continuous distance variable. I adopt the notation of Deaton (1997) in the framework below.

Locally weighted regressions assigned each home a weight based on distance over d dimensions (x^d), the type of kernel (K), and the kernel bandwidth (h). The weight for home i in the calculation of home *0*'s index is given by:

$$\theta_i\left(x_0^d, N\right) = \frac{1}{h(N)} K\left(\frac{\left\|x_0^d - x_i^d\right\|}{h(N)}\right). \tag{6}$$

[9] The regression coefficient will mechanically equal 1 when the regression contains a single index and the same homes are used to construct the evaluate it. For this reason, leave-one-out cross-validation is a necessary choice. I have experimented with leaving all homes in the index construction, but this issue is especially problematic for kernel bandwidth maximization since problems are magnified in small samples. The "optimal" local indices occur at several hundred homes rather than several thousand since using the same home in construction and evaluation leads to substantial predictive power in small samples.

Still, I have experimented with constructing and testing the indices without leave-one-out cross-validation. That is, I include the home in the question to both create and evaluate the house-specific index. Results yield substantially "smaller" local markets since econometric power vastly improves when including the home in small samples. However, it would be incorrect to do so since it overestimates the predictive power of the local market.

I have also experimented with a 90% sample and 10% testing sample structure as an alternative to leave-one-out cross-validation. The results are similar in magnitude, but the econometric precision is greatly reduced. Thus, I present results for the leave-one-out methodology only.

The distance function $\left(\left\|x_0^d - x_i^d\right\|\right)$ provides a measure of similarity between home *0* and home *i* based on measures such as geography or descriptive home characteristics. For example, weights based exclusively on price differential uses a distance function in a single dimension: $\left\|x_0^d - x_i^d\right\| = |\widehat{p}_0 - \widehat{p}_i|$, where $\widehat{p}_i$ is the fitted value of home *i* from a hedonic price regression. The *h*(*N*) notation indicates that the bandwidth varies to include exactly N homes.[10] Using $\theta_i\left(x_0^d, N\right)$ as weights, the house-specific home price index for home *0* is a weighted version of the Case-Shiller index regression from Eq. 3:

$$\widehat{\beta}\left(x_0^d, \theta\left(x_0^d, N\right)\right) = \left[Z'\theta\left(x_0^d, N\right)Z\right]^{-1} Z'\theta\left(x_0^d, N\right)(\Delta \log P). \tag{7}$$

I calculate four types of house-specific indices: (1) a Nearest Neighbor Index based on geographical location (x_N), (2) a Nearest Type Index based on home type characteristics (x_T), (3) a Nearest Price Index based on price (x_P), and (4) a 3-Dimensional Index using a composite distance over all three variables.

For the Nearest Neighbor Index, geographical distance is calculated "as the crow flies" using longitude and latitude coordinates. For the Nearest Price Index, price is based on the fitted hedonic price model described in "Data", where the distance between homes is the absolute price differential: $\left\|x_0^d - x_i^d\right\| = |\widehat{p}_0 - \widehat{p}_i|$.

For the Nearest Type Index, home type is based on the first three principal components from: square footage, land area, year built, construction grade, condition, type of structure, number of stories, type of exterior, and maintenance condition. The distance between two homes is the spherical distance in the first three components, normalized to have the same variance.

Finally, the 3-Dimensional Index calculates the weighted squared distance over all three dimensions: $\left\|x_0^d - x_i^d\right\| = \sqrt{\|x_{N0} - x_{Ni}\|^2 + A_T\|x_{T0} - x_{Ti}\|^2 + A_P\|x_{P0} - x_{Pi}\|^2}$. A_T and A_P are coefficients that assign relative weights to home type and price differences, respectively, with A_N set to *1*. In three dimensions, this kernel is an ellipse where A_T and A_P denote the relative skew along each axis. I normalize x_N, x_T, and x_P to be in standard deviation terms to give some interpretation to the coefficients. I also allow for a down-weighting by fraction δ to simulate the effect of a home being "down the street but in the other district." The 3-Dimensional index captures local shocks over the nearest N homes that are similar in distance, type, and price simultaneously. Whereas single-dimensional kernels have one degree of freedom: *hN*, the 3-Dimentional Index has four: *h*(*N*), δ, A_T and A_P.

Kernel Bandwidth Maximization

Kernel bandwidth maximization for local price indices selects the bandwidth that maximizes the predictive power of the Eq. 5. That is, I select a particular distance

[10] Varying the bandwidth based on the number of homes is one of two ways to proceed. The choice is whether to have bandwidths defined by fixed values, such as distance in miles, or to allow the bandwidth to vary with home density. In the end, I choose the latter to achieve econometric consistency in index evaluation.

All locally weighted regressions in this paper use epanechnikov kernels which decrease observation weights smoothly over distance. Indicator kernels were also considered and tested but ultimately omitted due to lower overall performance.

measure and optimize the bandwidth relative to that measure. Too wide a bandwidth makes "local" shocks not so local anymore, and too narrow a bandwidth introduces too much noise to the index.

I reduce the curse of dimensionality by parameterizing the distance measure $\left(\left\|x_0^d - x_i^d\right\|\right)$ and performing a grid search over one dimension at a time. For example, the Nearest Price index parameterizes distance as $\left\|x_0^d - x_i^d\right\| = |\widehat{p}_0 - \widehat{p}_i|$. and finds the N that maximizes the predictive power of $\widehat{\beta}\left(x_0^d, \theta\left(x_0^d, N\right)\right)$. For the 3-Dimensional Index, I perform a gradient parameter search over all parameters simultaneously.

A typical calculation proceeds as follows. First, I choose the desired index type and the number of homes in the index construction. One-by-one, I create a unique home price index for each of the 75,497 homes in the sample. Each index has the same number of homes that are included with positive regression weight. I calculate the main regression in Eq. 5 and record the R^2. This process is repeated for many bandwidths until a maximum is found.

Index Hedging

A better understanding of the statistical properties of local housing markets has important implications for hedging home price risk with a real estate market index.

Consider a representative homeowner or prospective homebuyer planning to buy or sell a home at some future date. He can remain unhedged to his home price risk by simply participating in the market. Or he can hedge home price risk by shorting the index if he is long housing or going long the index if he is short housing.

Mechanically, the hedging strategy works as follows. A homeowner hoping to sell his home can enter into a futures contract that pays off if home prices decline. If prices rise, he makes money on his house but loses money on his futures contract. If prices fall, he loses money on his house but makes money on his futures contract. In either case, his exposure to home price risk is reduced. The same logic and strategy can be applied to prospective homeowners buying into a market.

The effectiveness of such a hedging strategy depends on the statistical correlations between housing markets and individual home prices. With a perfectly correlated index, homeowners can fully hedge and remove home price risk entirely. But because home values do not move one-for-one with the hedge, a hedged homeowner reduces his *home price risk* but incurs *basis risk* in the process.

The possibility of hedging with futures markets for metropolitan indices is available in practice via markets on the Chicago Mercantile Exchange. As of 2009, futures contracts trade on the CME for ten metropolitan areas for time intervals up to 5 years ahead. The availability of such indices is promising, but low volume on these markets suggests hesitations in participation. Basis risk is one possible explanation; that is, the indices may be too different from any individual homeowner's market to be worthwhile.

The benefit of local market indices depends on the improvement in correlation that occurs from using local transactions. A more highly correlated hedge will benefit homeowners in two ways: (1) basis risk is reduced for any given amount of hedging, and (2) homeowners will purchase larger hedges.

I model the value of index hedging with a simple, single-period representative agent problem with two types of assets: (1) homes and (2) futures contracts price indices.[11] The agent owns 1 unit of housing and hedges with k units of the index. After choosing how much to hedge, his home receives shock ε and his hedge receives shock ν. His wealth process is given by $\log w = \log w_0 + \varepsilon + k\nu$. The correlation between ε and ν is known and given by $coor(\varepsilon,\nu) = \rho$. I assume that the hedge is actuarially fair, so that the mean of ε equals the expected home price appreciation and the mean of ν equals zero.

In order to determine a closed-form solution, I assume the agent has Constant Relative Risk Aversion (CRRA) utility function over his wealth: $U(w) = -w^{1-\lambda}/(1-\lambda)$, where $\lambda>1$ is the coefficient of relative risk aversion. I assume that all shocks are in logs and have normal distributions. Thus, the agent maximizes the mean of his log wealth distribution minus one half of the variance times the risk aversion parameter: $\max\left[\mu_{\log w} - (\lambda-1)\sigma^2_{\log w}/2\right]$. Under this formulation, the agent seeks only to minimize the variance of his wealth since his choice of hedge does not affect the mean of his wealth, given by $\log w = N(\log w_0 + \mu_\varepsilon, \sigma^2_\varepsilon + k^2\sigma^2_\nu + 2k\sigma_\varepsilon\sigma_\nu\rho)$.

A regression framework provides the solution for the optimal hedging strategy. The variance of log w is minimized when $k^* = -\rho\sigma_\varepsilon/\sigma_\nu = -\text{cov}(\varepsilon,\nu)/\text{var}(\varepsilon)$, which equals the coefficient from a regression of ε on ν. The optimal hedge is negative when ρ is positive since the agent increases his hedge when the value of his home has more correlation with the index.

The agent reduces the variance of his wealth equal to one minus the R^2 from the regression used to compute k^*, so that $\text{var}(\log w|k^*) = \sigma^2_\varepsilon(1-\rho^2)$. Although k^* decreases with σ_ν, $\text{var}(\log w|k^*)$ does not since the agent adjusts the size of his hedge proportionally to the market shock variance. The quadratic reduction in wealth variance results from two multiplicative effects: (1) the hedge performs better for any amount purchased and (2) the agent buys more of it.[12]

[11] I focus on futures contracts exclusively rather than consider alternative financial instruments,—most notably put options discussed in Shiller and Weiss (1999)—for two reasons. First, futures contracts are the most natural hedge for a homeowner wishing to reduce price risk; homeowners still take on home price exposure when hedging with put options. Second, the main qualitative results should carry over to other derivative contracts.

[12] The same regression setup and solution applies to the general case where a homeowner hedges multiple homes at the same time. That is, a homeowner is naturally long his own home and short all the homes to which he might move. A previous version of this paper considered the case where $\log w = \log w_0 + \sum_{i=0}^N a_i\varepsilon_i + \sum_{i=0}^N k_i\nu_i$, where the coefficients a_i represent the agent's exposure to house i (positive for currently owned homes and negative for possible future homes). The optimal hedging strategy is still given by the coefficients of a regression of $\sum_{i=0}^N a_i\varepsilon_i$ on ν_0, ν_1, … ν_N, and the fraction of overall wealth reduction remains one minus the R^2. Due to space considerations, I remove the analysis from the current version of the paper.

The value of index hedging equals the certainty equivalent of the hedged distribution minus the certainly equivalent of the unhedged distribution. Thus, the value of hedging with an index that has correlation parameter ρ is given by:

$$\begin{aligned} Value\left(\rho, \sigma^2_{\log w}, \lambda\right) &= CE\left(\mu_{\log w}, (1-\rho^2)\sigma^2_{\log w}, \lambda\right) - CE\left(\mu_{\log w}, \sigma^2_{\log w}, \lambda\right) \\ &= \rho^2 \frac{(\lambda-1)}{2} \sigma^2_{\log w}. \end{aligned} \tag{8}$$

Equation 8 indicates that the value of hedging with a perfectly correlated index (ρ=1) equals $(\lambda - 1)\sigma^2_{\log w}/2$. The value of an imperfect index depends on the correlation parameter. If an agent is willing to pay *$X* to remove all of his home price risk, he would pay *$X* × ρ^2 to hedge with an index that has correlation ρ. Intuitively, a completely independent hedge (ρ=0) has no value at all.

Results

Kernel Bandwidth Maximization

Figure 3 displays the kernel bandwidth maximization process using locally weighted regressions. The y-axis plots the R^2 of changes in log home prices regressed on changes in the log home price indices from Eq. 5. The x-axis plots the number of homes used to construct the index from Eq. 7. As discussed in

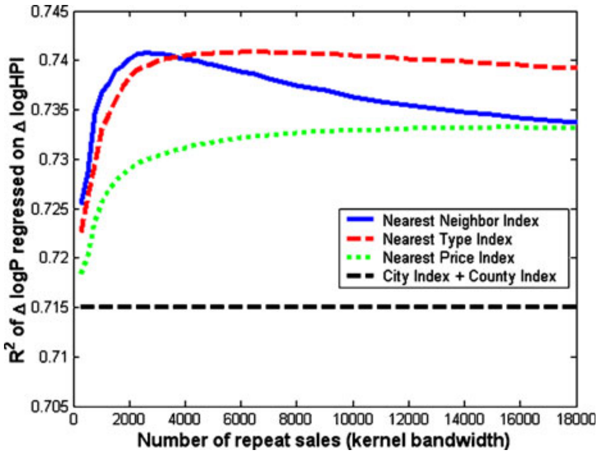


Fig. 3 Maximizing the home price index correlation over the kernel bandwidth. The above graph plots the R^2 when regressing changes in log home prices on changes in the log home price indices as a function of the number of repeat sales included in the index calculation. The indices are calculated using an epanechnikov kernel over the nearest *N* homes, so that homes just inside the cutoff have a small but positive weighting. The nearest *N* homes are determined by geographical distance for the Nearest Neighbor Index, the Euclidean distance over the first 3 principal components of home characteristics for the Nearest Type Index, and the price difference based on a fitted hedonic model for the Nearest Price Index. In all cases, the bandwidth adjusts to have exactly *N* homes in the index calculations. All regressions additionally contain the City Index, the County Index, and a constant on the right-hand side

Table 3 Optimal bandwidths for the continuous kernel indices

	Partition indices							Continuous indices					
	City	County	District	Zip	Type	Price	R^2	Nearest neighbor		Nearest type		Nearest price	
								N	R^2	N	R^2	N	R^2
1	.	.	.	.	.	.	.	4,000	0.739	7,500	0.741	15,000	0.733
2	y	y	.	.	.	.	0.715	2,500	0.741	6,500	0.741	15,000	0.733
3	y	y	y	.	.	.	0.729	2,250	0.741	4,250	0.745	14,000	0.739
4	y	y	.	y	.	.	0.730	2,500	0.741	4,250	0.745	14,000	0.740
5	y	y	.	.	y	.	0.728	2,500	0.746	5,000	0.741	13,000	0.733
6	y	y	.	.	.	y	0.731	2,250	0.746	4,000	0.742	15,000	0.733
7	y	y	y	y	y	y	0.740	2,250	0.747	3,250	0.747	1,250	0.742

The above table regresses changes in log home prices on changes in log home price indices for the entire sample of 75,497 repeat home sales. The first R^2 column represents the baseline regression that does not include a continuous index. The subsequent columns add either the Nearest Neighbor Index, the Nearest Type Index, or the Nearest Price Index, all of which are created with an epanechnikov kernel over the nearest N homes. The number of included homes represents the maximum R^2 achieved by searching over the possible values for N as in Fig. 3

"House-specific Indices", the Nearest Neighbor Index weights homes by geographical distance, the Nearest Type Index by home characteristics, and the Nearest Price Index by price differential.

The maximization calculates the value of local information orthogonal to the broad market since β^{city} and β^{county} are included on the right-hand side of Eq. 5. When N approaches zero, the regression approaches R^2=0.715, equal to the explanatory power of β^{city} and β^{county} alone, since the local index is too narrow. The regressions also approach R^2=0.715 as N goes to infinity since the local index no longer contains local information. A maximum occurs between these two extremes where the value of local information is the largest.

The maximums occur at quite locally defined markets. The optimal Nearest Neighbor Index utilizes just N=2,500 repeat-sales pairs. This corresponds to roughly 10 square miles (2% of the county land area) and 7,500 single-family homes (3% of the total number of residences) for the median home.[13] The maximum fit for the Nearest Neighbor Index is R^2=0.741. The Nearest Type Index achieves roughly the same maximum value (R^2=0.741) but uses a local market roughly 2.5 times as large (N=6,500) as the Nearest Neighbor Index. The Nearest Price Index does not achieves a lower maximum than the other two R^2=0.733 indicating that searching over price alone provides less predictive power than geography or home type. Creating the best Nearest Price Index requires a local market that spans roughly $200,000 and contains 20% of the county (N=15,000) for the average home price index (Table 3).

[13] One disadvantage of using a kernel bandwidth that adjusts with home density is that I report the local market size as a median or mean rather than as a fixed value for all homes. Still, the drawbacks are outweighed by the econometric consistency gained by using the same number of observations in each index calculation, as reviewed in "House-specific Indices".

Table 4 performs the bandwidth maximization exercise for a variety of index combinations. Row 1 shows the maximization exclusively with a local index, dropping β^{city} and β^{county} as regressors. Row 2 displays the maximums from Fig. 3. The subsequent rows represent different maximums that depend on other indices that are included. For example, Row 5 indicates that when combined with a simple partition index based on home type, the Nearest Neighbor Index achieves quite a good fit R^2=0.746. The best index combinations are those which capture both spatial and home type information.

Finally, I capture local market shocks across geography, home type, and price simultaneously by maximizing the 3-dimensional kernel described in "House-specific Indices". The maximized parameters are: N=6,000, δ=0.575, A_T=0.9, and A_P=1.2. The index achieves a fit of R^2=0.750 when β^{city} and β^{county} are included. The optimal N is larger than the Nearest Neighbor Index but it spans three dimensions. The δ value of 0.575 indicates that homes in different districts provide about half as much predictive power for homes in the same district. Interestingly, price variation gets the most weight in the 3-dimensional index calculation even

Table 4 Summary of home price shock variables

	Variable	N	Full sample		Time series, 1-year	
			Mean	SD	Mean	SD
1	Actual Price Shocks	.	0.447	0.396	0.079	0.090
2	City Index	.	0.428	0.327	0.072	0.073
3	County Index	.	0.014	0.030	0.003	0.018
4	District Index	.	0.001	0.059	0.000	0.047
5	Zip Code Index	.	0.002	0.072	0.001	0.058
6	Home Type Index	.	0.001	0.056	−0.001	0.034
7	Price Band Index	.	0.002	0.060	−0.001	0.035
8	Nearest Neighbor Index	500	−0.005	0.153	0.003	0.144
9	Nearest Neighbor Index	2,500	0.001	0.087	0.002	0.067
10	Nearest Type Index	500	−0.007	0.151	0.002	0.135
11	Nearest Type Index	6,500	−0.001	0.067	0.002	0.039
12	Nearest Price Index	500	−0.005	0.143	0.000	0.136
13	Nearest Price Index	15,000	−0.001	0.054	−0.001	0.029
14	3-Dimensional Index	6,000	−0.002	0.081	0.002	0.053

The above table summarizes home price shocks for all 75,947 homes used in the paper. The first two columns make no adjustment for time of sale or time between sales. The second two columns are 1-year time series values. All shocks are in logs, so that 0 indicates no price change. The City Index refers to the published S&P/Case-Shiller index for Washington, D.C., whereas all other indices are computed from the data. The County Index subtracts changes in the City Index, and all other indices subtract changes in the City Index and the County Index. Indices 4–7 are equally weighted indices based on the partitions described in Table 2. Indices 8–14 are weighted indices using an epanechnikov kernel over the nearest N homes. The 3-Dimensional Index utilizes a 3-dimensional epanechnikov kernel over distance, type, and price. Details on the construction of these indices can be found in "Methodology" though it performed the worst as a single-variable index. In other words, price matters much more when the market is already limited on other dimensions.

Local Price Indices

Table 4 displays summary statistics for the home price and index shocks. Local market shocks have higher variances than broad market shocks due to higher market volatility and increased measurement error. I include house-specific indices with $N=500$ to illustrate the lack of econometric precision when an index is constructed using too few homes.

Table 5 presents the raw correlations between the indices. All local indices subtract off the city and county components, so that the residual has very little correlation with the broad market. The local index correlations are high enough correlations to suggest that they contain similar information but low enough correlations so as to contain orthogonal information as well.

Table 6 presents the main regressions from Eq. 5 of changes in log home prices on changes in log home price indices for partition indices. Each column represents a different combination of local indices included on the right-hand side. As discussed, I include the City Index in full, the County Index as a residual of the City Index, and all local indices as residuals of both the City Index and the County Index. The "Net City Index" and the "Net County Index" present the relative loadings when every index is included in full.

The fit of the regressions in Table 6 provide information about how strongly each index relates to home price shocks. The City Index by itself achieves a pretty high fit already ($R^2=0.711$). Adding the County Index improves the predictive power only slightly ($R^2=0.715$). Adding each of the local partition indices in columns (3)–(6) provide modest improvements: District Index ($R^2=0.729$), Zip Code Index ($R^2=0.730$), Home Type Index ($R^2=0.728$), and Price Band Index ($R^2=0.731$).

The coefficients in Table 6 provide information about how the explanatory power is distributed across indices. The indices are certainly correlated with one another (as in Table 5), so including them simultaneously tells about the relative importance of each. The low coefficients on the Net City Index and the Net County Index indicate that local indices absorb most of the predictive power. The fact that the coefficients on the various local indices are so similar in column (7) indicates that each local index represents a different aspect of local housing markets.

Table 7 presents the same analysis for house-specific indices. As before, each column represents a different combination of right-hand side regressors. On the whole, house-specific indices demonstrate significant improvement over the simpler partition indices. The fits of the optimal indices are given by: Nearest Neighbor Index ($R^2=0.741$), Nearest Type Index ($R^2=0.741$), Nearest Price Index ($R^2=0.733$), and 3-Dimensional Index ($R^2=0.750$). Including all indices simultaneously achieves the best fit ($R^2=0.756$).

As before, the coefficients in Table 7 provide information about how the explanatory power is distributed across indices. The low coefficients on the $N=500$ indices demonstrate the empirical imprecision of calculating an index with such few values. In the final column, the 3-Dimensional Index absorbs the most predictive power by a large magnitude, indicating that it is perhaps the best measure of a local market.

Table 5 Index correlations

	Variable	N	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	Actual Price Shocks	.	1.00													
2	City Index	.	0.84	1.00												
3	County Index	.	0.41	0.41	1.00											
4	District Index	.	0.10	−0.03	−0.01	1.00										
5	Zip Code Index	.	0.10	−0.03	−0.01	0.62	1.00									
6	Home Type Index	.	0.09	−0.03	−0.01	0.30	0.28	1.00								
7	Price Band Index	.	0.10	−0.03	−0.01	0.35	0.32	0.77	1.00							
8	Nearest Neighbor Index	500	0.08	−0.04	−0.03	0.32	0.36	0.21	0.24	1.00						
9	Nearest Neighbor Index	2,500	0.15	−0.02	−0.01	0.62	0.69	0.31	0.36	0.47	1.00					
10	Nearest Type Index	500	0.07	−0.05	−0.02	0.21	0.21	0.30	0.31	0.23	0.25	1.00				
11	Nearest Type Index	6,500	0.12	−0.04	−0.03	0.39	0.38	0.63	0.64	0.29	0.44	0.44	1.00			
12	Nearest Price Index	500	0.04	−0.04	−0.02	0.16	0.16	0.32	0.37	0.16	0.18	0.23	0.29	1.00		
13	Nearest Price Index	15,000	0.09	−0.05	−0.02	0.35	0.32	0.78	0.92	0.25	0.36	0.33	0.69	0.40	1.00	
14	3-Dimensional Index	6,000	0.16	−0.03	−0.03	0.59	0.55	0.51	0.56	0.37	0.63	0.37	0.74	0.27	0.60	1.00

The above table presents the index correlations for the shocks summarized in Table 4. The correlations are equally weighted over all 75,497 homes in the sample. Indices 4–14 subtract the changes in the County Index and the City Index. The County Index subtracts the changes in the City Index. N refers to the number of homes included in the index calculation

Table 6 Explained home price shocks: partition indices

Dependent variable: change in log home price

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
City Index	1.020 (0.005)	0.984 (0.005)	0.987 (0.004)	0.988 (0.004)	0.988 (0.004)	0.989 (0.004)	0.991 (0.004)
County Index		0.938 (0.047)	0.939 (0.040)	0.940 (0.040)	0.938 (0.044)	0.939 (0.043)	0.940 (0.037)
District Index			0.800 (0.037)				0.340 (0.040)
Zip Code Index				0.669 (0.031)			0.328 (0.038)
Home Type Index					0.815 (0.036)		0.243 (0.039)
Price Band Index						0.840 (0.034)	0.422 (0.039)
Constant	0.011 (0.003)	0.012 (0.003)	0.010 (0.002)	0.010 (0.002)	0.009 (0.003)	0.009 (0.002)	0.007 (0.002)
Net City Index	1.020	0.046	0.048	0.048	0.050	0.050	0.052
Net County Index	.	0.938	0.139	0.271	0.123	0.099	−0.393
N	75,947	75,947	75,947	75,947	75,947	75,947	75,947
R^2	0.711	0.715	0.729	0.730	0.728	0.731	0.740
Adjusted R^2	0.711	0.715	0.729	0.730	0.728	0.731	0.740

The above columns regress changes in log home prices on changes in log home price indices for the entire sample of repeat home sales. The City Index is the published S&P/Case-Shiller Index. All remaining indices are constructed from the data. Index calculations exclude the specific home used in the above regression so that the left-hand side and the right-hand side never utilize the same data. The District Index and the Zip Code Index are based on pre-coded partitions in the data. The Home Type Index is based on the first principal component of home characteristics. The Price Band Index is based on fitted values from a hedonic price regression. See "Local Home Price Indices" for details. The County Index subtracts changes in the City Index. All other indices subtract changes in the City Index and County Index. The Net City Index and the Net County Index rows present coefficients when regressing on full indices rather than index residuals. All standard errors are clustered by subdivision

Table 7 Explained home price shocks: house-specific indices

Dependent variable: change in log home price

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
City Index	1.020	0.984	0.989	0.986	0.989	0.991	0.987	0.992	0.989	0.991	0.993
	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)
County Index		0.938	0.952	0.958	0.942	0.959	0.946	0.941	0.981	0.964	0.977
		(0.047)	(0.039)	(0.034)	(0.043)	(0.039)	(0.045)	(0.042)	(0.033)	(0.033)	(0.030)
Nearest Neighbor Index (N=500)			0.302								0.090
			(0.017)								(0.013)
Nearest Neighbor Index (N=2,500)				0.729						0.496	0.260
				(0.028)						(0.031)	(0.035)
Nearest Type Index (N=500)					0.276						0.074
					(0.014)						(0.010)
Nearest Type Index (N=6,500)						0.943				0.527	0.174
						(0.032)				(0.034)	(0.040)
Nearest Price Index (N=500)							0.200				0.029
							(0.014)				(0.009)

Nearest Price Index (N=15,000)								0.987 (0.038)		0.247 (0.036)	0.114 (0.033)
3-Dimensional Index (N=6,000)									0.911 (0.029)		0.457 (0.039)
Constant	0.011 (0.003)	0.012 (0.003)	0.012 (0.002)	0.010 (0.002)	0.012 (0.003)	0.010 (0.002)	0.012 (0.003)	0.010 (0.002)	0.011 (0.002)	0.009 (0.002)	0.010 (0.002)
Net City Index	1.020	0.046	0.037	0.028	0.047	0.033	0.041	0.051	0.008	0.028	0.016
Net County Index	.	0.938	0.650	0.229	0.666	0.016	0.746	−0.046	0.070	−0.306	−0.219
N	75,947	75,947	75,947	75,947	75,947	75,947	75,947	75,947	75,947	75,947	75,947
R^2	0.711	0.715	0.729	0.741	0.726	0.741	0.720	0.733	0.750	0.751	0.756
Adjusted R^2	0.711	0.715	0.729	0.741	0.726	0.741	0.720	0.733	0.750	0.751	0.756

The above columns regress changes in log home prices on changes in log home price indices for the entire sample of repeat home sales. The City Index is the published S&P/Case-Shiller Index. All remaining indices are constructed from the data. All index calculations exclude the specific home used in the above regression so that the left-hand side and the right-hand side never utilize the same data. The Nearest Neighbor, Nearest Type, and Nearest Price Indices are created with an epanechnikov kernel over the nearest N homes based on distance (Neighbor), principal components of home characteristics (Type), and a fitted price regression (Price). The 3-Dimensional Index utilizes a 3-dimensional epanechnikov kernel over distance, type, and price. See "House-specific Indices" for calculation details. The County Index subtracts changes in the City Index. All other indices subtract changes in the City Index and County Index. The Net City Index and the Net County Index rows present coefficients when regressing on full indices rather than index residuals. All standard errors are clustered by subdivision

Variance Decomposition of Home Price Shocks

Up until this point, home price shocks have been considered without reference to the time interval between sales. Although price shock variance clearly increases with time, the relative sizes of the city, local, and idiosyncratic components are unclear before looking at the data. The variance decomposition is especially important in an index hedging context because it indicates the time intervals over which home prices can most effectively be hedged.

Figure 4 presents the variance decomposition of home price risk over time. The portion below the dotted red line represents the variation explained by the city index (β^{city}). The portion of the graph above the dotted red line but below the dashed blue line represents the additional explanatory power of the county index (β^{county}) and the local index (β^{local}). Finally, the remaining portion represents the idiosyncratic home price risk unexplained by the indices.

Market components achieve a maximum at four-and-a-half years, at which point the city and local components explain 82% of the variance in home price shocks. The inverted-U shape likely occurs since shorter intervals have more short-term sales variance whereas longer intervals have more variance due to maintenance and upkeep. Over the range of time intervals, local markets account for as little as 3% and as much as 7% of home price shocks, compared to 50–75% for the city component.

Figure 5 presents the risk decomposition unadjusted for the increasing shock variance with time. The risk share is the same as Fig. 4 but is presented in absolute

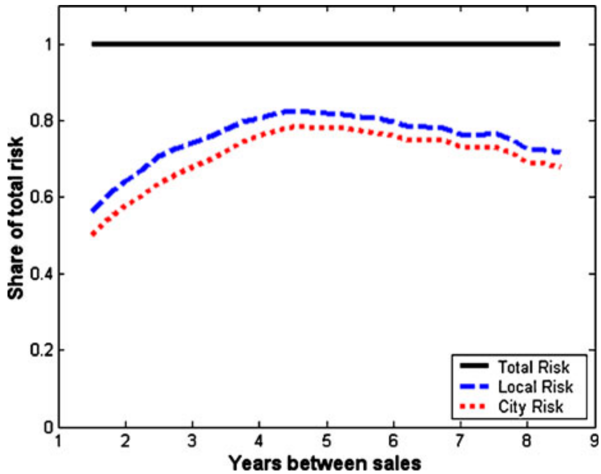


Fig. 4 Variance decomposition of home price risk. The above figure plots the variance decomposition of home price risk as a function of the years between sales. The share of city risk lies below the dotted red line. The share of local risk lies between the dashed blue line and the dotted red line. The share of idiosyncratic risk lies above the dashed blue line. The city share is determined by the explanatory power of the S&P/Case-Shiller Washington D.C. Metropolitan Index on home price shocks. The local share refers to the additional explanatory power provided by the 3-Dimensional Index, which maximizes predictability over distance, home type, and price. All estimates are smoothed over a 5-quarter interval, so that the point corresponding to X years represents all repeat home sales occurring over a $X-2$ to $X+2$ quarter time interval

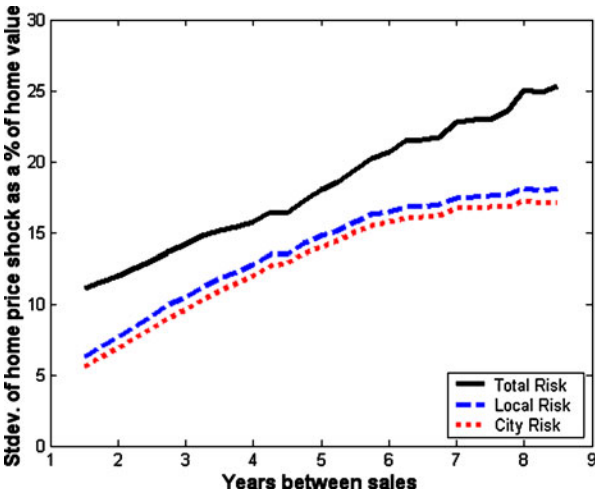


Fig. 5 Value decomposition of home price risk. The above figure plots the total home price variance decomposition as a function of the years between sales. The total risk is given by the standard deviation of home price shocks around the mean for each time interval. The share of city risk lies below the dotted red line. The share of local risk lies between the dashed blue line and the dotted red line. The share of idiosyncratic risk lies above the dashed blue line. The city share is determined by the explanatory power of the S&P/Case-Shiller Washington D.C. Metropolitan Index on home price shocks. The local share refers to the additional explanatory power provided by the 3-Dimensional Index, which maximizes predictability over distance, home type, and price. All estimates are smoothed over a 5-quarter interval, so that the point corresponding to X years represents all repeat home sales occurring over a $X-2$ to $X+2$ quarter time interval

rather than relative terms. Although the *proportion* of explained home price shocks is maximized near four-and-a-half years, the overall hedging benefit is still larger for longer time intervals.

Index Hedging

The value of index hedging depends on the initial size of home price risk and the extent to which hedging reduces that risk. Equation 8 indicates that the willingness to pay for index hedging can be estimated by three parameters: (1) ρ—the correlation between one's home price and the index, (2) σ_ε—the standard deviation of the log home price shock, and (3) λ—the coefficient of relative risk aversion. The first two are estimated from the data. For the coefficient of relative risk aversion, I calibrate the model by assuming that the median homeowner has a $10,000 valuation for removing his home price risk entirely over a 5-year period. This results from a risk aversion parameter of $\lambda=2.460$.[14] Estimates for σ_ε vary depending on the time between sales. From the data, I estimate $\widehat{\sigma}_\varepsilon = 0.113$ for 2-years, $\widehat{\sigma}_\varepsilon = 0.166$ for 5-years, and $\widehat{\sigma}_\varepsilon = 0.223$ for 8-years.

[14] A $10,000 valuation of removing home price risk seems reasonable given that the standard deviation of home price shocks over a 5-year period is around 18% of home values, corresponding to roughly $95,000 for a home worth $500,000. Still, the risk aversion parameter is not meant to be a dogmatic assumption. The reader can easily adjust this value to his or her own liking. However, I choose it to provide some intuition for the magnitude of the value of hedging.

Table 8 presents the value of hedging home price risk as a function of which indices are available. Row 1 represents an unhedged homeowner. Row 15 represents a homeowner who hedges with a perfectly correlated index. All rows in between represent indices studied in this paper.

Each row in Table 8 contains an estimate for ρ, the hedge value in dollars terms, and the hedge value as a percentage of the perfect hedge. The dollar values are calibrated on the 5-year perfect hedge value of $10,000 in Row 15, but the percentages are independent of this calibration.

Comparing the hedging value of the various local indices versus the City Index provides a sense for how local indices might benefit homeowners in practice. The City Index in Row 2 has the highest correlation with home prices over a 5-year time interval, capturing 78% of the value of the perfect hedge versus just 58% and 68% for the 2-year and 8-year intervals, respectively.

The best performing local index is the 3-Dimensional Index in Row 11. The 3-Dimensional Index improves the value of index hedging relative to the City Index by four percentage points and 5–10% in dollar value. The benefits of local indices are the greatest in percentage terms for the 2-year interval, but the greatest in dollar terms for the 8-year interval. Including multiple local indices in the regression

Table 8 The value of index hedging

	Index hedge	2-years			5-years			8-years		
		ρ	Hedge value		ρ	Hedge value		ρ	Hedge value	
1	Unhedged	0.000	$ -	0%	0.000	$ -	0%	0.000	$ -	0%
2	City Index	0.760	$ 2,697	58%	0.883	$ 7,797	78%	0.830	$ 12,528	69%
3	County Index	0.774	$ 2,797	60%	0.885	$ 7,832	78%	0.833	$ 12,622	69%
4	District Index	0.791	$ 2,920	63%	0.895	$ 8,008	80%	0.841	$ 12,872	71%
5	Zip Code Index	0.788	$ 2,902	62%	0.893	$ 7,976	80%	0.842	$ 12,897	71%
6	Home Type Index	0.795	$ 2,952	63%	0.897	$ 8,039	80%	0.841	$ 12,866	71%
7	Price Band Index	0.798	$ 2,973	64%	0.898	$ 8,062	81%	0.842	$ 12,909	71%
8	Nearest Neighbor Index	0.791	$ 2,920	63%	0.899	$ 8,077	81%	0.844	$ 12,976	71%
9	Nearest Type Index	0.799	$ 2,985	64%	0.901	$ 8,120	81%	0.848	$ 13,081	72%
10	Nearest Price Index	0.799	$ 2,982	64%	0.899	$ 8,075	81%	0.843	$ 12,943	71%
11	3-Dimensional Index	0.801	$ 2,997	64%	0.905	$ 8,190	82%	0.851	$ 13,192	72%
12	Partition Indices, 4–7	0.806	$ 3,033	65%	0.902	$ 8,141	81%	0.848	$ 13,081	72%
13	Continuous Indices, 8–10	0.807	$ 3,044	65%	0.906	$ 8,203	82%	0.852	$ 13,199	73%
14	All Indices, 2–11	0.810	$ 3,062	66%	0.907	$ 8,230	82%	0.854	$ 13,258	73%
15	Perfect Hedge	1.000	$ 4,672	100%	1.000	$ 10,000	100%	1.000	$ 18,199	100%

The above table displays the value of index hedging for a representative agent trying to minimize the variance of his home price shock with an index hedge. ρ is the correlation between log home price shocks and log home price index shocks. The middle column is an agent's willingness to pay for an index hedge according to the formulation in "Index Hedging. The final column is the value as a percentage of the Perfect Hedge. The dollar value is calculated as the certainty equivalent of an agent who has CRRA utility over log home price shocks with a risk aversion parameter of λ=2.460. The risk aversion parameter is calibrated based on a 5-year value of the Perfect Hedge equal to $10,000 increases the value of local indices slightly, but the gains are small relative to the 3-Dimensional Index alone.[15]

Conclusion

This paper defines and evaluates a variety of local market indices in one metropolitan area in an effort to learn about the size and scope of local real estate markets. I use locally weighted regression techniques to construct "house-specific" indices that maximize home price explanatory power over continuous variables such as location, home type, and price. I present the results an index hedging framework to estimate the value-added of the ability to hedge with various local market indices.

Results suggest that local indices add a moderate amount of explanatory power relative to metropolitan indices. In my sample, the metropolitan index explains 50–75% of the variation in home price shocks already, and local indices add 3–7% more. Overall, my sample suggests quite locally defined housing markets. The best predictive power is achieved while searching local markets over roughly 10 square miles, or just 3% of county residences. The findings suggest that local housing markets may contain as few as 7,500 homes.

The index hedging framework suggests that homeowners will pay 5–10% more to hedge with a local index versus a city index. The two main benefits of local indices are that homeowners hedge more risk and that the hedges perform better. Because local market indices allow homeowners to hedge local moves, they provide additional value where city indices have none.

A significant limitation of these findings is that data come from a single county. Though the methods could be applied elsewhere, the sample is certainly limited in scope. Since nearly one-quarter of all moves are local in nature, studying the statistical properties of home prices in other locations is a worthwhile pursuit.

The introduction of housing futures markets on the Chicago Mercantile Exchange offers new and exciting prospects for home price risk management using financial derivatives. To the extent that these markets transfer risk between mutually agreeable parties, they are a step in the right direction. This paper adds to this endeavor by studying the statistical properties of home prices for a particular real estate sample. The more researchers understand about how home prices interact across time and space, the better equipped they will be to design effective risk management markets in housing.

Acknowledgments I thank Markus Brunnermeier, Fernando Ferreira, David Lee, Burton Malkiel, Chris Mayer, John Quigley, Ricardo Reis, Jesse Rothstein, Hyun Shin, Albert Saiz, Todd Sinai, an anonymous referee, and seminar participants at the Industrial Relations Section, Bendheim Center for Finance, and the NBER Summer Institute for Real Estate & Local Public Finance for helpful conversations and suggestions. I would additionally like to thank the National Science Foundation and Princeton University for generous financial support throughout this project.

[15] Since home price shocks vary over a typical real estate cycle, I tested whether these results were driven by any non-random sampling over time. I repeated the calculations while weighting observations by the inverse of their frequency by year. The results were not sensitive to this robustness check.

References

Bailey, M., Muth, R., & Nourse, H. (1963). A regression model for real estate price index construction. *Journal of the American Statistical Association, 58*, 933–942.

Baroni, M., Barthelemy, F., & Mokrane, M. (2008). Is it possible to construct derivatives for the Paris residential market? *Journal of Real Estate Finance and Economics, 37*(3), 233–264.

Bourassa, S., Hoesli, M., & Peng, V. (2003). Do housing submarkets really matter? *Journal of Housing Economics, 12*(1), 12–28.

Bourassa, S., Hoesli, M., & Sun, J. (2006). A simple alternative house price index method. *Journal of Housing Economics, 15*(1), 80–97.

Case, K., & Shiller, R. (1987). Prices of single-family homes since 1970: new indexes for four cities. *New England Economic Review, 5*, 45–56.

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

Clapham, E., Englund, P., Quigley, J., & Redfearn, C. (2005). Revisiting the past and settling the score: index revision for house price derivatives. *Real Estate Economics, 34*(2), 275–302.

Dale-Johnson, D. (1982). An alternative approach to housing market segmentation using hedonic price data. *Journal of Urban Economics, 11*(3), 311–332.

Deaton, A. (1997). *The analysis of household surveys*. Baltimore: The Johns Hopkins University Press.

Deng, Y., & Quigley, J. (2007). Index revision, house price risk, and the market for home price derivatives. *Springer, 37*(3), 191–209.

Fan, J. (1992). Design-adaptive nonparameteric regression. *Journal of the American Statistical Association, 87*(420), 998–1004.

Goodman, A., & Thibodeau, T. (2003). Housing market segmentation and hedonic prediction accuracy. *Journal of Housing Economics, 12*(3), 181–201.

Goodman, A., & Thibodeau, T. (2007). The spatial proximity of metropolitan area housing submarkets. *Real Estate Economics, 35*(2), 209–232.

Maclennan, D., & Tu, Y. (1996). Economic perspectives on the structure of local housing systems. *Housing Studies, 11*(3), 387–406.

McMillen, D. (2004). Locally weighted regression and time-varying distance gradients. In A. Getis, J. Mur, & H. Zoller (Eds.), *Spatial econometrics and spatial statistics* (pp. 232–249). New York: Palgrave Macmillan.

Shiller, R., & Weiss, A. (1999). Home equity insurance. *Journal of Real Estate Finance and Economics, 19*(1), 21–47.