# Home price risk, local market shocks, and index hedging

DeForest McDuff*
Princeton University

March 24, 2008

**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 for homes from a large suburban county of Washington D.C.. I frame the analysis in terms of home price risk and index hedging, and I attempt to answer two key questions in this regard. The first is to determine at what level home price shocks occur; i.e., how local are housing markets? The second is to estimate how much homeowners would be willing to pay for access to home price index markets various local levels, which would permit hedging the local components of home price risk. I construct and evaluate a variety of local real estate indices that group homes by district, zip code, home type, and price band, and I calculate several "house-specific" indices using locally weighted regressions across these variables. Local indices are advantageous relative to metro-level indices for two reasons: (1) they capture local shocks that do not appear in the broader market indices (so they permit homeowners to hedge a larger portion of their basis risk), and (2) they permit hedging of price changes which are relevant to the large fraction of moves that occur within metropolitan areas. I estimate that local market indices explain 3-7% more of the variation in home price shocks than is explained by a city index, depending on the time interval between sales. A typical homeowner would be willing to pay 5-10% more to hedge with a local index compared to a city index, and homeowners anticipating local moves would be willing to pay as much as 25% more.

*Industrial Relations Section, Firestone Library, Princeton, NJ 08544. E-mail: rmcduff@princeton.edu. I would like to thank Markus Brunnermeier, Fernando Ferreira, David Lee, Burton Malkiel, Ricardo Reis, Jesse Rothstein, Hyun Shin, Albert Siaz, Todd Sinai, and seminar participants at the Industrial Relations Section Graduate Labor Lunch and the Benheim Center for Finance Student Research Workshop for helpful conversations and suggestions. I would additionally like to thank the National Science Foundation and Princeton University for their generous financial support throughout the project.

# 1 Introduction

All real estate markets are local, but how local is local? Using over 20 years of home transaction data, I target two aspects of this question. First, I attempt to determine the local level at which home price shocks occur. I create a variety of local housing indices using the standard repeat-sales methodology and evaluate their predictive power for individual home sale prices. In doing so, I learn not only about the size of local markets but also the dimensions over which they are defined, including the relative importance of geography, home type, and price.

Second, I ask how much a typical homeowner would be willing to pay to hedge his home price risk with an actuarially fair index that recognizes this locality. Although home prices are not typically as volatile as other asset classes, home buyers are often highly leveraged and very long in real estate. Home price risk is therefore important to many households. Until recently, homeowners have had few financial vehicles for reducing home price risk. The recent creation of futures markets for metropolitan-level home price indices is certainly a step in the right direction, but metro-level indices leave much to be desired, as they fail to capture within-metropolitan market differences. In my empirical analysis of home price shocks, I attempt to estimate what portion of home price shocks are local and how much value would be added if local indices were available. The answer depends both on how local markets are defined and on the underlying relationships between price shocks at various market levels.

Home prices are subject to a variety of supply and demand shocks at different levels, ranging from the national all the way down to the purely idiosyncratic. Home price indices can only permit hedging against shocks at or above the level at which they are defined. There is thus a trade-off in the design of such a index. Larger markets contain a more heterogenous mixture of homes but smaller markets do not contain enough transactions for meaningful inference. A local home price index, which consists of both the city and local market components, will explain more of the variation in home prices than the city index alone. But if locality is defined too narrowly, sampling error will produce too much noise.

If one thinks of home price shocks from one period to the next as a random distribution, the market components form the distribution mean and the individual or idiosyncratic component is the distribution variance. Practically speaking, the market components are unobservable and even the home prices themselves are observable typically only one or twice per decade. For this reason, the distinction between the two components is possible only after estimating home price indices for various markets.

The main result in the paper is that local market components account for 3-7% of the variation in home price shocks compared to 50-75% at and above the city market component (much of this is national), depending on the time interval between sales. These two market portions of home price shocks can be hedged away with futures contracts whereas the remaining individual component of home price risk cannot. I find local market information to be quite narrow, with the best geographical indices being created with markets as small as 10 square miles and containing only 3% of the county residences, on average. The majority of this 3-7% increase occurs quite locally and cannot be captured at even the county-level. When translated into dollar terms, a typical homeowner would be willing to pay around 5-10% more to have the local index available compared to the city index alone. For homeowners anticipating local moves, the benefit is as large as 25% in dollar terms. Finally, local indices provide value to hedging local moves that is unavailable with a city index alone.

The main computational challenge in the paper is the calculation of S&P/Case-Shiller (C-S) Home Price Indices [19] for various subsets of homes in the sample. The C-S methodology uses repeat home sales to generate an index that tracks real estate price appreciation over time, free of sample selection bias. I take advantage of the transaction format of the raw data to construct a repeat sales database for this purpose. Once I have indices created I various levels, I compare their predictive power using leave-one-out cross-validation so that all predictions are out-of-sample. Since the home transaction data in this paper come from a single Maryland county (population 900,000), the Washington D.C. metropolitan area is the broadest market studied. For this market, I use the official S&P/Case-Shiller Index that is used for futures contracts on the Chicago Mercantile Exchange (CME). All other home price indices are calculated from the home transactions database.

The local markets analyzed in this paper generally fall into one of two broad categories. The first type divides the sample into partitions so that each home is exclusively in a single submarket. The four partition types are based on district, zip code, home type, and market price. The former two are already coded in the tax database, and the latter two are based on the first principal component of home characteristics and a hedonic price regression, respectively. The second main type of local market is house-specific; that is, the market is centered at the home itself and extends outward to the nearest $N$ homes. I define and construct several types of house-specific indices, including a Nearest Neighbor Index based on geographical distance, a Nearest Type Index based on the first 3 principal components of home characteristics, a Nearest Price Index based on market price, and finally a 3-Dimensional Index based on all three continuous variables. I calculate house-specific indices using locally weighted home price index regressions, weighting transactions based on an epanechnikov kernel over the relevant dimension.[1]

The main advantage of house-specific indices with a continuous distance measure is that the size of the local market is not predetermined. By changing the bandwidth, I can include as many or as few homes as necessary in the index calculation. One of the main empirical contributions of this paper is the determination of the optimal bandwidth that maximizes predictive power. In doing this, I learn about the optimal size of local market information. Too large a bandwidth causes the market to be too heterogenous, whereas too small a bandwidth contains too few transactions to have any predictive power. In the end, I choose the kernel bandwidth that minimizes mean squared error in home price prediction and thus maximizes the index correlation.

[1]Using locally weighted regression with home transaction data in and of itself is not an entirely novel approach - see McMillen [12] for an example and relevant references - but to my knowledge it has not been used to create "house-specific" indices, where the weighting scheme is unique to each home.

Ultimately, homeowners will prefer local market indices to city market indices primarily because they have higher correlations to home price shocks. Local indices add value through two main channels. The first is to reduce the *basis risk* associated with index hedging. Basis risk arises from the presence of home price shocks that are orthogonal to the index and therefore cannot be hedged. More local home price indices have less basis risk since the idiosyncratic component of home price risk is smaller, so long as they are not so local as to introduce substantial measurement error. Not only will the hedge be more effective, but homeowners will hedge a larger fraction of their risk. The second channel through which local indices add value is by allowing a homeowner to hedge *within-city* moves, which is not possible with city indices alone.

In order to estimate the value associated with hedging within-city moves, I analyze a variety of sample moving scenarios faced by homeowners, including a ”representative” agent who moves with some probability according to a poisson process. Table 1 presents some summary statistics of U.S. Census migration rates for Montgomery County, MD, the source of the home transaction data. This table demonstrates the importance of hedging local moves for many homeowners. In 2000, half of all homeowners lived in a different house than in 1995: 22.5% in the same county, 6.2% in a different county in the same state, and 13.8% in a different state. As metro-level indices offer no hedging value for moves within the same county (at least when buying and selling a home of similar value), even a well-functioning market in metro-level indices would leave 58.4% of movers unhedged.

Although home prices rarely suffer a nominal price decline, it is not uncommon for housing prices to decline in real terms. Figure 1 shows the inflation-adjusted price history for the Washington D.C. Index and the Montgomery County Index over the time period spanned by the data. From the high point in 1990 to the low point in 1997, home prices in the D.C. area declined by 25% in real terms. Given the speed and extent to which prices rose in the last few years, it would not be unsurprising to see an equal or greater decline in the years moving forward.

If home prices were perfectly correlated with market indices, homeowners could hedge away all of the uncertainty associated with home price shocks, but the idiosyncratic component of home price risk puts on upper bound on risk reduction. In this paper, I attempt to approach this upper bound by maximizing local home price indices over several dimensions. Index hedging creates a mutually beneficial exchange to the degree that homeowners can transfer their home price risk to individuals, investors perhaps, who want to take on the risk. Homeowners may even buy more units of housing that they otherwise would have in an unhedged position.

Finally, after examining the size and scope of local market shocks empirically, I set up a simple model to define and assess the value of index hedging. The setup fits nicely into a regression framework, whereby a homeowner’s optimal hedging strategy is given by the regression coefficients and the amount of the home price risk reduction is the fraction of home price shocks explained by the index. The framework is general enough to be applied to a variety of homeowner risk profiles, and I consider several specific cases in more detail.

## 2 Home price indices and housing submarkets

Although the Case-Shiller home price indices have only recently gained mainstream media attention with their introduction on the Chicago Mercantile Exchange, the academic history of home price indices dates back to Bailey et al. [1], who first proposed a repeat-sales home price index calculation in a regression framework. The academic popularity of such indices grew substantially in the late 1980s when Case and Shiller [5] and [6] refined the repeat-sales empirical methodology and estimated home price indices with home transaction data for several U.S. cities. Over the years, the Case-Shiller methodology has become more or less the academic gold standard for estimating home price indices. An alternative approach to estimating home price indices is a hedonic regression of prices on home characteristics, which either include time dummies or allow time-varying characteristic coefficients. Research efforts have also been made towards combining the repeat-sales and hedonic approaches into single, hybrid-version indices.[2] In this paper, I focus on the repeat-sales methodology as my main computation of home price indices.

The ability for homeowners and housing-related businesses to hedge home price risk with such home price indices was examined in detail by Shiller and Weiss [18], who proposed a variety of risk reduction financial instruments, including futures, options, and event-triggered derivatives. Indeed, Macromarkets, LLC has recently partnered with the CME to develop some of these instruments in practice. For a variety of reasons, including the low trading volume (as of this writing) in the index futures markets and perhaps a lack of awareness among homeowners who might benefit from such hedges, home price risk reduction with index derivatives remains largely a vision for the future.

There is also some discussion regarding issues with index revisions and contract settlements in repeat-sales indices. Index revisions occur when previously published index estimates are revised based on new data. Clapham et al. [7], Baroni et al. [2], and Deng and Quiqley [10] examine the magnitude of such revision errors, with the former warning that revisions are "not inconsequential for the settlement process". Although the authors propose some practical solutions to the revision process, they ultimately conclude that home price futures markets might better be served by hedonic price indices. While revision biases may be substantial in some contexts, I choose not to directly address them in this paper as they distract from the ultimate aim of examining within-market home price dispersion. One caveat, however, is that revisions may be more important in smaller markets like the ones in this paper since econometric precision is a major factor for price prediction. As a result, I am able to estimate more local markets than might be possible in real time.

Another body of literature relevant for this paper is that which focuses on housing submarkets, or smaller groupings of homes within a broader market, for example, a city within a country or a locality within a metropolitan area. Papers that examine housing submarkets generally take 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,

[2] See Bourassa et al. [4] for a good overview and references to this research area.

or zip code. Goodman and Thibodeau [14] use all of these divisions - including a more flexible hierarchical model - to improve the predictive power of a hedonic regression of home prices on home characteristics. Their approach essentially allows the researcher to choose various sets of submarkets and determine which ones have the best hedonic predictive power.

Several studies extend this methodology by recognizing that housing submarkets can span home characteristics beyond geography. Goodman and Thibodeau [15] 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 paper also provides references to other studies with similar techniques.

The second approach is more flexible and allows the data to statistically define housing submarkets. Dale-Johnson [8], Maclennan and Tu [17], and Bourassa et al. [3] use factor analysis and clustering techniques to group homes together, the basic idea being to estimate hidden factors of home clusters using a broad data set of home characteristics and/or sales transactions and organize homes into subgroups based on how similar they are in price, land area, square footage, location, etc. The latter study demonstrates the improved predictive power of a hedonic model that uses housing submarket information in this setup.

While I draw upon ideas from the literature, I take a novel approach on several dimensions. By creating home price indices with the Case-Shiller methodology, I implicitly focus on price co-movements rather than price levels as the defining characteristic of a market. Two houses in precisely the same market will see their prices move together since they respond to the same supply and demand forces. Home A and Home B need not necessarily be the same price or the same size to be subject to the same market forces. The more important determinant is whether their prices respond to the same supply and demand shocks. Fundamentally, I view each home is a unique product, and I seek to identify which characteristics make homes close substitutes.

Unfortunately, price co-movements are not directly observable since homes transact so infrequently, so I must first construct the price paths before I evaluate them. In doing so, I do not abstract away from price levels and home characteristics entirely. Instead, I use notions of which factors might be relevant for price co-movement and then allow the data to determine which correspond to substitutability. The house-specific indices allow me to calculate a unique price path for each home and evaluate home substitution accordingly. Price indices which are created from their correct underlying markets will have the greatest predictive power.

# 3 Data

The data used in this paper come from the 2006 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, so it includes a cross-section of home characteristics as well as the 3 previous sales transactions for each home. I focus on Montgomery County, which is the largest county in Maryland by population, home to approximately 900,000 residents and 250,000 single-family homes. Sections of the county near the Washington D.C. border are mostly urban while parts of the county far from the city are quite rural, so there is substantial variation in home prices and types.

Since I use the Case-Shiller repeat-sales methodology for calculating and evaluating home price indices, my final data set consists exclusively of homes which have transacted twice over the sample period. The full county database contains 227,554 single-family homes and ultimately 75,947 repeat sales pairs from 1985-2006. Following the lead of Case and Shiller [5] and Standard & Poor's [19], I take steps to exclude repeat-sales that may not be competitive or take place over too short a time interval.[3] Table 2 displays the summary statistics for the cleaned dataset of repeat-sales used throughout the paper. Because the Maryland Department of Assessments and Taxation lacks latitude and longitude for each home, I assign them using an ArcGIS address-matching algorithm, using which I identify the exact location of 88.1% of the county addresses.

I also use the tax assessment database to divide the sample into four different local market partitions: 'District', 'Zip Code', 'Home Type', and 'Price Band'. The District partition is based on 11 districts coded in the data. These districts often overlap with the county's 19 high school districts, though the high schools themselves are not directly coded.[4] Figure 2 displays the full sample of 200,493 geocoded homes as well as the imputed district boundaries. The Zip Code partition is created from zip code identifiers also in the data. By collapsing smaller zip codes into larger ones based on location, I create a full partition based on 30 zip code submarkets. The Home Type partition is constructed from the first principal components of the following home characteristics: square footage, land area, year built, construction grade, condition, type of structure, number of stories, type of exterior, and maintenance condition. The homes are then ordered based on this principal component and divided into 11 equally sized submarkets.[5] The Price Band partition is based on the fitted values from a hedonic price regression on a similar set of characteristics used

[3] The raw tax assessment database contains the price and date of the previous 3 transactions for each home in the county. After linking together consecutive sales at the same residence, I start with 121,210 repeat-sales pairs with complete price and date information. I consider only those transactions after 1985. First, I drop all sales that are not "arms-length", indicating a non-competitive sale (a sale from one family member to another, for example, would not be recorded as arms-length). Then, 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-sale and drop observations that are outside one and a half standard deviations of the return distribution mean (the dropped tails turn out to include 2.6% of the sample) in order to eliminate unusual sales circumstances as well as homes which are most likely to have changed in quality. Finally, I drop all observations for which I cannot locate the longitude and latitude of the address. The final sample contains 75,947 observations.

Because the database contains a maximum of 3 transactions for each home, there is some concern for missing sales if a home has sold four or more times. A frequency plot of the earliest sale year relative to the construction year (since there are no sales before the home exists) suggests that this is not likely to cause problems; possible "missing" sales only affect 10% of homes built after 1995 and 20% of homes built after 1985. For this reason and the sales transaction density for the early years, I calculate all indices for 1985-2006 only. Although the S&P/Case-Shiller indices are not published for 1985-1986, I impute the index values for these years based on home price appreciation in Montgomery County during that time interval.

[4] In an attempt to assign high school districts to the sample, I used the official maps from the Montgomery County Public Schools web site and calibrated the longitude/latitude coordinates into computer pixel coordinates with Google Maps. The result was a noisy high school district variable assignment which turned out to have less predictive power than the 'District' variable provided by the database, despite their strong overlap. In the end, I use only district variables provided in the database rather than assigning high school districts myself.

[5] I create 11 submarkets for both the Home Type and Price Band partitions to make them comparable to the District partition coded in the data, though experimenting with other sizes suggests that anywhere from 5 to 15 submarkets would yield similar results.

for the principal components.[6] Finally, the database contains subdivision codes which I use for clustering standard errors.

# 4 Local market shocks

## 4.1 Local home price indices

The most basic computation in this paper is a home price index regression, proposed by Bailey et al. [1] and popularized by Case and Shiller [5]. Essentially, the calculation uses the time and price variation of repeat home sales to estimate the path of average home price appreciation for a set of homes. Because the repeat-sales methodology is so fundamental to the analysis, I provide a brief review.

The home price index calculation 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 ($\alpha_i$), a market component ($\beta_{jt}$), and an error term ($\varepsilon_{ijt}$):

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

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

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

Under this formulation, the market component time series ($\beta$) can be estimated with repeat sales pairs of prices and dates. The dependent variable is the change in the log price between sales, and the independent variable is an $N$ by $T$ matrix (number of homes by time periods) that represents the timing of the sales ($t_0$ and $t_1$) for each home:

$$\underbrace{\Delta \log P}_{Nx1} = \underbrace{\beta}_{Tx1} * \underbrace{Z}_{NxT} + \underbrace{\varepsilon}_{Nx1}, \tag{3}$$

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

[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 11 submarkets.

$$Z = \underbrace{\begin{bmatrix} 0 & -1 & 0 & ... & 0 & 1 & 0 & ... & 0 & 0 & 0 \\ -1 & 0 & 0 & ... & 0 & 0 & 0 & ... & 0 & 1 & 0 \\ ... \\ 0 & 0 & 0 & ... & 0 & -1 & 0 & ... & 0 & 0 & 1 \end{bmatrix}}_{NxT}. \tag{4}$$

Thus, $\beta$ is a $Tx1$ vector representing the home price index over the time period covered by the data.

Case and Shiller [5] note that transaction pairs with longer time between sales tend to have larger error terms, on average, which will be true if the error term ($\varepsilon_{ijt}$) has a random walk component. To correct for this heteroskedasticity, they estimate a second-stage regression of the error terms on the time between sales and rerun the first-stage specification using the square roots of the fitted values from the second-stage as weights. Throughout the paper, I adopt this convention to downweight repeat-sales over longer periods of time, though the results are not very sensitive to this adjustment.

The most basic local market indices to construct and estimate are those which divide the sample into separate partitions. I calculate four main partition indices based on district, zip code, home type, and price band using equation (3) with various subsamples of homes. After constructing these local indices, I perform a variance decomposition by regressing the change in log home prices on the changes in log home price indices:

$$\Delta \log P_{ij(t_0,t_1)} = \alpha + k_{city}\Delta\beta_{city,j(t_0,t_1)} + k_{county}\Delta\beta_{county,j(t_0,t_1)} + k_{local}\Delta\beta_{local,j(t_0,t_1)} + \varepsilon_i, \tag{5}$$

where $\beta_{city}$ denotes the Washington D.C. Metro Index, $\beta_{county}$ denotes the County Index, and $\beta_{local}$ denotes the relevant Local Index. The $k$'s are the regression coefficients, and the $j$ and $(t_0, t_1)$ subscripts on the indices indicate the change in market $j$'s index between time $t_0$ and $t_1$. In other words, two homes in the same market will have different right-hand-side values if their sales occurred at different times. Throughout the paper, equation (5) is the main evaluative measure of index predictive power.

One can imagine the two extremes of such a regression. In a world where all home prices move one-for-one (in percentage terms) with the index, the regression will have a perfect fit. Conversely, in a world where movements within the home price distribution are entirely random, the index tell us only about the mean price change and the regression will have no explanatory power. In other words, the explained variation in home price movements represents the broad market risk which can be hedged and the error term represents the idiosyncratic home price risk which cannot. Equation (5) allows us to distinguish between the two risks and determine their magnitudes empirically.

The regression format is convenient for a variance decomposition because the explained portion of home price shocks will differ depending on what indices are included on the right-hand-side. A more detailed analysis in Section

5 indicates that in the context of index hedging, the regression coefficients provide the optimal hedging strategy and the $R^2$ is the fraction of wealth variance reduction from hedging. Finally, the coefficients indicate how to create the best composite index as a weighted sum of the raw indices. Such an index has identical predictive power to having all of the above indices.

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 many times, each instance leaving out a single home - so that indices are never evaluated with the same homes used to create them.[7]

Before discussing the explanatory power of the various home price indices, I first present some descriptive statistics of home price and index shocks. Table 3 presents the full sample and 1-year time series means and standard deviations of log changes in prices and home price indices. The full sample statistics are unweighted averages and the time series statistics are averages over time. The 1-year actual home price shock mean of 0.079 corresponds to roughly an 8% annual increase in home prices, most of which can be attributed to the explosive growth in the Washington D.C. housing market during the last decade.

In order to examine the relative sizes of these shocks, I present index changes as residuals by subtracting the City Index appreciation from the County Index ($\tilde{\beta}_{county} = \beta_{county} - \beta_{city}$) and subtracting both the City Index and the County Index from each Local Index ($\tilde{\beta}_{local} = \beta_{local} - \tilde{\beta}_{county} - \beta_{city}$). In other words, the local indices are reported as residuals of the larger market components. The County Index has a mean greater than zero since Montgomery County, MD has experienced slightly higher average home price appreciation in percentage terms than the rest of Washington D.C. over this time period, but the residual variance of the County Index is quite small relative to the other index shocks. All of the local indices have means close to zero since they are submarkets of the County Index, which is calculated using the same data. The size of the local index shocks tends to be larger for indices that use lower numbers of homes, both because the local variation is greater and the indices are less precise.

Table 4 presents the main regressions that explore the home price shock explanatory power of local partition indices. Each column in the table represents the regression from equation (5) with different combinations of home price indices included on the right-hand-side. For coefficient consistency across columns, 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. Thus, the City Index coefficient represents the total sum of coefficients if each were index included in full. In order to interpret how the explanatory power loads on each individual index, the reader should compare the

[7] In the case where home price movements are regressed on changes in a single index, the regression coefficient must mechanically equal 1 if the index calculation is unweighted (as opposed to the standard Case-Shiller downweighting of sales over longer intervals) and the same home used to evaluate the index is used in the right-hand side index calculation. While this issue is important to recognize, it will not be the case here since neither of these two conditions are satisfied.

local index coefficients to the ”Net City Index” and ”Net County Index” rows at the bottom of the table.

The first thing to note in Table 4 is the increasing $R^2$ when local indices are added. The City Index by itself achieves a pretty good fit ($R^2 = 0.711$), and adding the County Index doesn't improve predictive power substantially ($R^2 = 0.715$). Adding each of the local partition indices in Columns (3)-(6) generate 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 $R^2$ difference between column (1) and the other columns indicates the extra predictive value of indices that capture local price shocks.

The coefficients are also useful in interpreting how the explanatory power is distributed across indices. The low coefficients on the Net City Index and Net County Index indicate that the local indices absorb most of the predictive power. In Section 5, I demonstrate that the coefficients are equivalent to the optimal index hedging strategy. The coefficients indicate that an agent trying to hedge his home price risk would optimize by hedging with a local index in lieu of a broader market index. Finally, the coefficients in the final column demonstrate that each partition captures a different dimension of local market shocks, with the explanatory power loading roughly equal amounts on all four partition indexes.

In later sections of the paper, I will revisit these results in the context of index hedging. For now, I move on to the construction and evaluation of house-specific indices.

### 4.2 House-specific indices

One of the main innovations of this paper is to calculate house-specific home price indices using locally weighted regressions based on geographical distance, home type, and price. I call them ”house-specific” because I use different weights to calculate a unique home price index for each home in the sample. The basic idea of locally weighted regression, pioneered by Fan [11], is to minimize weighted mean-squared error, with weights decreasing over a continuous distance variable based on the kernel selection and bandwidth. I adopt the notation of Deaton [9] in the framework below.

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

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

The distance function, $\left\|x_0^d - x_i^d\right\|$, provides a measure of similarity between home 0 and home $i$ based on measures like geography and home characteristics. For example, when I weight homes exclusively based on price differential, the distance function in this single dimension is given by $\left\|x_0^d - x_i^d\right\| = |\hat{p}_0 - \hat{p}_i|$, with $\hat{p}_i$ being the fitted value of home

$i$ from a hedonic price regression. I write $h(N)$ as such since I use "nearest-neighbor" bandwidths that adjust in size to include exactly $N$ homes. This feature is especially relevant given the variation in home density across localities. The basic choice is whether to have bandwidths defined relative to fixed measures, say distance in miles, or whether to allow the kernel to expand in less dense areas so that I include the same number of homes in each calculation. In the end, I choose the latter in order to have econometric consistency when evaluating various indices. Using $\theta_i(x_0^d, N)$ as weights, the house-specific home price index for home 0 is a weighted version of the Case-Shiller index calculation from equation (3):

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

All locally weighted regressions presented in this paper use epanechnikov kernels, for which weights decline smoothly in distance, that extend outward to cover the nearest $N$ homes. Indicator kernels were also tested and considered but were ultimately omitted as they perform less well overall. Finally, note that the full county index and the partition indices from Section 4.1 can be viewed as a special case of equation (7), where the kernel is an indicator function for whether home 0 and home $i$ are in the same partition.

I calculate four main types of local indices using locally weighted regressions: a Nearest Neighbor Index based on geographical location ($x_N$), a Nearest Type Index based on home type characteristics ($x_T$), a Nearest Price Index based on price ($x_P$), and 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 Type Index, I calculate the first three principal components based on the following home characteristics: square footage, land area, year built, construction grade, condition, type of structure, number of stories, type of exterior, and maintenance condition, and the distance function for $x_T$ is just the spherical distance in the first three principal components (normalized to have the same variance). The Nearest Price Index is based on a fitted hedonic price model (described in Section 3), such that the distance function is simply absolute price differential: $\|x_{P0} - x_{Pi}\| = |\hat{p_0} - \hat{p_i}|$.

Finally, the 3-Dimensional Index calculates the distance between homes as a 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 can be described as an ellipse where $A_T$ and $A_P$ determine the relative skewness 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 a down-weighting for homes in different districts by fraction $\delta$ to simulate the effect of a home being "down the street but in the other district". The main advantage of this kernel is that it can capture local shocks that occur across several dimensions: i.e. the nearest $N$ homes that are the similar in distance, type, and price simultaneously. Whereas the 1-dimensional kernels have a single degree of freedom (the kernel bandwidth: $h(N)$), the 3-Dimensional

Index is maximized over 4 degrees of freedom: $h(N)$, $\delta$, $A_T$, and $A_P$. More details on this maximization process are provided in the following section.

## 4.3 Kernel bandwidth maximization

The local price index, $\beta_{local}$, depends on the selected bandwidth and on the space over which distance is measured. The optimal local index is the one that maximizes the predictive power of regression (5). Thus, I select a particular distance measure and optimize the bandwidth relative to that measure. Including too many homes (wide bandwidth) makes "local" shock not so local anymore as the local market becomes less homogenous. Including too few homes (narrow bandwidth) introduces too much noise to the index. I investigate this tradeoff in depth by calculating local indices for the entire bandwidth space and ultimately choosing the one that maximizes the $R^2$.

In order to reduce the curse of dimensionality, I parameterize the distance measure ($\left\|x_0^d - x_i^d\right\|$) and typically maximize over one home characteristic at a time. In each instance, I select a specific distance measure and maximize the explained variance of regression (5) with a grid search over the number of homes to include in the index ($N$). For example, I choose $\|x_{P0} - x_{Pi}\| = |\hat{p_0} - \hat{p_i}|$ for the Nearest Price Index and then find the $N$ that maximizes the predictive power of $\hat{\beta}(x_P, \theta(x_P, N))$. The optimal bandwidth will differ depending on whether the broad market indices are also available. If they are, large scale shocks are absorbed by the broad indices and the local index can be more sensitive to local shocks. For the 3-Dimensional Index, I perform a gradient parameter search over the 4 parameters simultaneously: $N$ (number of homes in the index), $\delta$ (out-of-district penalty), and $A_T$ and $A_P$ (relative importance of home type and price, in standard deviations). In all cases, maximizing over $N$ is just short-hand for maximizing over $h(N)$, meaning that the bandwidth is selected so that exactly $N$ homes have non-zero weighting.

Figure 3 displays how the kernel bandwidths are maximized with respect to the locally weighted regressions. The y-axis plots the $R^2$ of changes in log home prices on changes in the log home price indices as in equation (5), with the City Index, County Index, and either the Nearest Neighbor, Nearest Type, or Nearest Price Index on the right-hand-side. The x-axis plots number of homes used to construct a weighted index, as in equation (7). As previously discussed, the bandwidths grow or shrink with home density, so that their sizes are chosen to include exactly $N$ homes with positive weights. Recall that the Nearest Neighbor Index weights homes based on geographical distance, the Nearest Type Index weights homes based on the first three principal components of home characteristics, and the Nearest Price Index weights homes based on the difference in price.

A typical calculation for Figure 3 proceeds as follows. First, I choose the desired index type and the number of homes to include in the index construction, say the Nearest Neighbor Index with $N = 2000$ homes. One-by-one, I create a unique home price index for each of the 75,497 homes in the sample. Each index requires determining the kernel bandwidth so that exactly 2000 homes get positive weights and then performing the weighted index calculation in equation (7). Next, I perform the main regression in equation (5) with the full sample, including $\beta_{city}$ and $\beta_{county}$ as right-hand-side regressors. In this case, the $R^2 = 0.740$, so I plot the point $(N = 2000, R^2 = 0.740)$ in Figure 3. The process is repeated for a grid search over $N$ for each type of index.

Figure 3 illustrates the trade-off between econometric precision and index locality. When $N$ is small to the far left of the graphs, the local index contains mostly noise, so the explanatory power essentially relies on $\beta_{city}$ and $\beta_{county}$. As $N$ approaches 0, the local index contains no information and the graphs approach $R^2 = 0.715$, which is the explanatory power of $\beta_{city}$ and $\beta_{county}$ alone. Moving up the graph as $N$ gets bigger, explanatory power increases as more homes are added and the estimates of the local market shock improve. At some point, the local index hits a maximum and adding additional homes decreases predictive power as the market becomes less and less local. The maximization occurs when the cost of making the market larger exactly offsets the improvement in predictive power from a more precise local index. Asymptotically as $N$ goes to $\infty$, the predictive power reverts back to $R^2 = 0.715$ as the local index no longer contains any local information.

The location of the maximization and the shapes of the plots in Figure 3 provide information about the relative sizes and characteristics of local market shocks over each dimension. For example, Nearest Neighbor Index maximization indicates that local markets defined over geography are quite homogenous relative to the other two measures. This is illustrated by the sharp improvement in predictive power for small values of $N$. Even the smallest local markets in this space add significant improvement. The best Nearest Neighbor Index occurs using the nearest $N = 2500$ repeat-sales pairs $(R^2 = 0.741)$. This local market corresponds to roughly 10 square miles (2% of the county land area) and 7500 single-family homes (3% of the total number of residences) for the median home.[8]

The other two indices achieve maximums at larger market sizes. The Nearest Type Index achieves roughly the same maximum $(R^2 = 0.741)$ but uses a local market roughly 2.5 times as large $(N = 6500)$ as the Nearest Neighbor Index. The Nearest Price Index does not achieve as high a maximum value $(R^2 = 0.733)$ as the other two, indicating that searching exclusively over price is a less effective strategy than searching over 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 $(N = 15000)$ for the average home price index.

Table 5 performs the same bandwidth maximization exercise for a variety of index combinations. The maximizations from Figure 3 are in Row 2. The top row shows the maximization with a local index exclusively, dropping $\beta_{city}$ and $\beta_{county}$ as regressors, resulting in slightly larger markets and less predictive power. The subsequent rows represent different maximums that depend on what other indices 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

[8]One of the disadvantages of using nearest neighbor kernels is that I must report the local market size as a median or mean rather than as a fixed value for all homes. In addition, one might be more interested in the total number of homes in a market rather than the number of repeat-sales (though they will be proportional to the extent that sales volume is evenly distributed). Still, I think these drawbacks are outweighed by the econometric consistency gained by using the same number of observations in each index calculation.

($R^2 = 0.746$). The best index combinations are those which capture both spatial and home type information.

Finally, I attempt to capture local market shocks by maximizing the 3-dimensional kernel over geography, home type, and price simultaneously. The maximized parameters are: $N = 6000$, $\delta = 0.575$, $A_T = 0.9$, and $A_P = 1.2$, and such an index achieves a fit of $R^2 = 0.750$ when $\beta_{city}$ and $\beta_{county}$ are included. In this case, the optimal $N$ is larger than the Nearest Neighbor Index but it spans 3-dimensions. Indeed, this index does not consider the "mansion down the street" or the "identical home on the other side of the county". The $\delta$ value of 0.575 indicates that homes in different districts are about half as important as homes in the same district. Interestingly, the variation in price is given the most weight in the distance calculation even though this variable performed the worst using single-variable kernels. In other words, the market is narrow with respect to price as long as the search scope is limited in other dimensions.

Revisiting Table 3, I display some summary statistics of the local market shocks measured by the house-specific indices. As expected, the local shocks have a higher variance than the broad market indices. The higher variance is driven both by higher local market volatility and measurement error for small local markets. I include indices where $N = 500$ to illustrate the lack of econometric precision when an index is constructed using too few homes.

Table 6 presents the raw correlations of these shocks which ultimately determine the explanatory power of the indices. The City Index has a 0.84 correlation with the actual home price shocks as well as a correlation of 0.41 with the residual County Index. Columns 2 and 3 contain almost exclusively zero correlations since the local indices are displayed as residuals of the County Index. As expected, the correlations are high enough to suggest that the local indices contain some of the same information but low enough so as to contain orthogonal information as well.

Table 7 presents the main regressions of changes in log home prices on changes in log home price indices for the house-specific indices. As before, each column in the table represents a different combination of home price indices included on the right-hand-side of the regression, and the local indices are calculated as residuals of the County Index. On the whole, the 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 ($R^2 = 0.756$).

Recall also that the coefficients provide information about which indices absorb most of the predictive power. In columns (3)-(9), one can compare the coefficient on the local index with those in the "Net City Index" and "Net County Index" rows. Note that the coefficients on the $N = 500$ indices are substantially lower than the optimized indices since they contain more estimation noise. In the final column, the 3-Dimensional Index absorbs by far the most predictive power, indicating that it is perhaps the best measure of a local market.

### 4.4 Variance decomposition of home price shocks

Up until this point, home price shocks have been considered without reference to the fact that shocks vary substantially by the time interval between sales. Although it may be obvious that price shock variance increases with time, it is not obvious before looking at the data how the proportion of each component varies in the time dimension. This is especially important in an index hedging context because it determines which time intervals can be most effectively hedged by broad market indices.

Figure 4 presents the variance decomposition of home price risk over the time interval between sales into city, local, and idiosyncratic components. The city risk component is the portion of the graph below the dotted red line and represents the home price shock variation explained by $\beta_{city}$. The local risk component is the portion of the graph above the dotted red line but below the dashed blue line and represents the additional explanatory power of $\beta_{county}$ and $\beta_{local}$. The remaining portion above the dashed blue line is the idiosyncratic home price risk that cannot be hedged by any index. Explanatory power achieves a maximum at 4.5 years, when the city and local components explain 82% of home price shock variance. Over the entire range of time intervals, home price shocks attribute as little as 3% and as much as 7% to local markets. The inverted-U shape indicates that home prices most strongly correlated with their markets over a 4-5 year time interval. Shorter time intervals are perhaps subject to more short-term sales variance, whereas long-term correlation is more likely influenced by individual maintenance and upkeep relevant for property values relative to the market.

Figure 5 presents the risk decomposition unadjusted for the increasing shock variance with time. The risk share is the same as Figure 4 but the risk is presented in absolute rather than relative terms. As before, the city risk lies below the dotted red line, the local risk lies between the dotted red line and the dashed blue line, and the idiosyncratic home price risk lies above the dashed blue line. In the context of index hedging, the idiosyncratic risk is the value ”left on the table” that cannot be captured by a home price index. Although the proportion of explained home price shocks is maximized at 4-5 years, the overall hedging benefit is still larger for longer time intervals.

Thus far, I have emphasized the size of local market shocks and the extent to which indices can explain the variance in home prices. I now turn the focus towards estimating the utility and dollar value of index hedging from the point of view of a representative homeowner.

## 5 Index hedging

The S&P/Case-Shiller metropolitan home price index futures markets began trading on the Chicago Mercantile Exchange (CME) in May, 2006. One benefit of these markets is to allow individuals and business to hedge home price exposure with futures and options and transfer this risk to a broader set of investors. As of January 2008, futures contracts trade on the CME for a US composite index and 10 metropolitan areas for time intervals up to five years ahead. From the perspective of an individual homeowner, the availability of such indices should be welcomed as an opportunity to shed some unnecessary home price risk and turn out to be especially important in the downward housing market that seems to have begun in the United States since 2005.

Consider the hedging strategy of a homeowner or prospective homeowner planning to buy or sell a home at some date in the future. He can either remain unhedged to this home price risk by simply participating in the market at that future time, or he can hedge his home price risk by entering into a real estate futures contract based on a home price index for that location. Mechanically, the hedging strategy works as follows. A homeowner who wants to sell a home can enter into a futures contract that pays off if a home price index falls below a market-determined level. If prices unexpectedly rise, he makes money by selling his home at a higher price but loses money on the futures contract. Conversely, falling prices means that he loses money on his home sale but makes money on the hedge. In either case, the homeowner has reduced his exposure to changes in the value of his house, either up or down. The same logic and strategy can be applied to a prospective homeowner buying into a market or to any individual planning to change his or her home price exposure (trading up for a more expensive house, for example).

The effectiveness of such a hedging strategy depends on the correlation between home price shocks and the index used to hedge. With a perfectly correlated index, homeowner can fully hedge and remove home price risk entirely. But because home values do not move one-for-one with the index hedge, a hedge homeowner reduces his *home price risk* but incurs *basis risk* in the process. A higher correlated hedge will make him better off for two reasons. First, basis risk is reduced for any given amount of hedging, and second, homeowners will purchase larger hedges than they would have with a lower correlated index.

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 for the home price index hedge. In the base case, the agent owns a home and plans to exit the market at some future data. He needs only to hedge his exposure to his current home's price. Thus, he owns 1 unit of housing and hedges with $k$ units of an index futures contract. In the more general setup, the agent anticipates a move from his current home to one of many potential future homes. After choosing an index portfolio, the agent's assets incur one-time price shocks: home $i$ receives shock $\varepsilon_i$ and index $j$ receives shock $\nu_j$. I assume that all shocks occur in logs and are normally distributed with standard deviations equal to $\sigma_{\varepsilon_i}$ and $\sigma_{\nu_j}$, respectively. The correlation structure between the shocks is non-zero and known by the agent.

In order to arrive at a closed-form solution, I assume that the agent has a 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 wealth shocks are log-normally distributed, so that the agent effectively maximizes the mean of his log wealth distribution minus one half of the variance times the risk aversion parameter: $\max\left[\mu_{\log w} - \frac{(\lambda-1)\sigma^2_{\log w}}{2}\right]$. With actuarially fair indices, an agent's portfolio does not change the distribution mean and thus agents seek merely to minimize the variance of wealth. Under this formulation, the certainty equivalent (CE) of a future wealth distribution is just the term inside the brackets, and the value of index hedging is the certainty equivalent of the hedged distribution minus the certainty equivalent of the unhedged distribution. If hedging with an particular index reduces an agent's wealth variance by fraction $f$, the value of that hedge is given by:

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

According to equation (8), the value of a perfect hedge $(f = 1)$ equals $\frac{(\lambda-1)\sigma^2_{\log w}}{2}$; the agent would be willing to pay this amount for access to a fair market in such hedges, as this would permit him to eliminate his risk entirely.

I focus on futures contracts exclusively rather than consider alternative financial instruments, most notably put options as suggested by Shiller and Weiss [18], for several reasons. First, basis risk is still relevant for homeowners who hedge with put options, who would expect the puts to pay off in precisely the same scenarios as when their home value declines. Second, futures contracts are the most natural hedge for a homeowner that wants to remove home price speculation entirely; homeowners still take on home price exposure when hedging with put options. Finally, one would expect the main qualitative results are to carry over with other derivative contracts.

## 5.1 Hedging a single home

I begin with the simplest case of index hedging, where an agent owns 1 unit of housing and hedges with $k$ units of a real estate index. After the shocks are realized, the agent's wealth process is given by $\log(w) = \log(w_0) + \varepsilon + k\nu$, with the correlation between the two assets given by $corr(\varepsilon, \nu) = \rho$. The mean of $\varepsilon$ represents the expected home price appreciation and may be non-zero whereas the mean of $\nu$ equals zero since the hedge is actuarially fair. So $\log(w) \sim N(\log(w_0) + \mu_\varepsilon, \sigma^2_\varepsilon + k^2\sigma^2_\nu + 2k\sigma_\varepsilon\sigma_\nu\rho)$. Note that the agent has no control over his distribution mean, but only the distribution variance, which he seeks to minimize. The crucial parameter in this model is the correlation between the home price shock and the index shock, which will depend upon which index is used.

The optimal hedging strategy and minimization of wealth variance fit nicely into a regression framework. The variance of $\log(w)$ is minimized when $k^* = -\rho\frac{\sigma_\varepsilon}{\sigma_\nu} = -\frac{cov(\varepsilon,\nu)}{var(\varepsilon)}$. Note that $k^*$ is the coefficient from a regression of $\varepsilon$ on $\nu$, and that $k^*$ is negative when $\rho$ is positive since the agent wants the value of his hedge to increase when the value of his home decreases, and vice versa. This corresponds to the notion that an agent long on housing will in general short the housing index. Although $k^*$ decreases with $\sigma_\nu$, the unhedged fraction of his home value is $var\left(\log(w) \mid k^*\right) = \sigma^2_\varepsilon\left(1-\rho^2\right)$; the homeowner merely adjusts the size of his hedge to compensate for the variance differential between the shocks. The reduction in wealth variance is just one minus the $R^2$ from the regression used to compute $k^*$. The quadratic correlation reduction can be interpreted as the combination of two multiplicative effects: (1) the hedge performs better for any amount purchased and (2) the agent buys more of it. With perfect correlation

($\rho = 1$), the variance of wealth equals zero, and with complete independence ($\rho = 0$), the agent buys no index hedge and incurs a shock equal to the unhedged wealth variance. Remember from equation (8) that the value of index hedging can be written as a fraction of the perfect hedge. In words, if an agent is willing to pay $10,000 to remove all of his home price risk, he should be willing to pay $10,000*$\rho^2$ to hedge with an index that has correlation $\rho$.

The same analysis carries over to multiple indices, where the optimal hedge falls directly out of a variance decomposition. A home price shock can be decomposed into a city market component ($C$), a local market component ($L$), and an idiosyncratic component ($D$). If a homeowner have two available indices, the city index ($\nu_C$) and the local index ($\nu_L$), a homeowner's optimal hedging strategy is given by the negative coefficients of a regression of $\varepsilon$ on $\nu_C$ and $\nu_L$. As before, the certainty equivalent value of the portfolio is increased via hedging by an amount equal to the $R^2$ times the certainty equivalent of the perfect hedge. Partial hedging with each index is equivalent to hedging with a composite index, which weights the two indices according to their regression coefficients.

This setup also useful in interpreting the regression coefficients presented in Table 4 and Table 7. For example, Column (2) in Table 7 suggests that the County Index nearly dominates the City Index since the optimal strategy is to go short 0.938 units of the County Index and only 0.048 units of the City Index. Remember that the agent actually wants to be short the home price index if he is long his own home. Just as the County Index dominated the City Index, certain local indices dominate both the County Index and the City Index in terms of hedging strategy. An example of this is Column (9), where the agents shorts 0.911 units of the 3-Dimensional Index, 0.070 units of the County Index, and only 0.008 units of the City Index. When the agent can hedge with multiple local indices, he sometimes hedges in the opposite direction with the County Index, as in Column (10) when he shorts all of the local indices but goes long the County Index by 0.306 units. The final regression in Column (11) indicates that the primary hedge will come from the 3-Dimensional Index, as indicated by the largest coefficient.

In assessing the dollar value of index hedging, the two main considerations are the amount of risk there is to begin with and the extent to which index hedging reduces this risk. According to equation (8), the willingness to pay for index hedging can be estimated with three parameters: $\rho$ (the correlation between one's home price and the index), $\sigma_\varepsilon$ (the standard deviation of the log home price shock), and $\lambda$ (the coefficient of relative risk aversion). Because the first two can be estimated from the data, the only free parameter is $\lambda$. I calibrate the risk aversion to be $\lambda = 2.377$ by assuming that the median homeowner has a $10,000 valuation for removing his home price risk entirely over a 5-year period. The standard deviation of home price shocks over this time period is around 18% of one's home value, corresponding to roughly $95,000 for a home worth $500,000. This risk aversion parameter as such is not meant to be a dogmatic assumption, but rather to place a reasonable dollar value - as determined by this author - on index hedging.

Table 8 presents the value of hedging one's home price risk as a function of which indices are available. Row 1 represents an unhedged homeowner, who by definition is exposed to the price risk on his own home. Row 15 represents a homeowner who can hedge his home price risk using a perfectly correlated index. All rows in between represent various local hedges that have been studied throughout this paper. The value of the perfect hedge is determined based on the estimate of $\sigma_\varepsilon$ for each time interval. The estimates used in this table are $\hat{\sigma_\varepsilon} = 0.113$ for 2-years, $\hat{\sigma_\varepsilon} = 0.166$ for 5-years, and $\hat{\sigma_\varepsilon} = 0.223$ for 8-years. The only other parameter to determine the value is $\rho$, which will vary depending on the index used.

Each row in Table 8 contains an estimate for $\rho$, the hedge value in dollar terms, and the hedge value as a percentage of the perfect hedge. Readers should observe that the 5-year value in Row 15 equals $10,000 by assumption. The City Index in Row 1 has the highest correlation in the 5-year range, capturing 78% of the value of the perfect hedge compared to 58% and 68% for the 2-year and 8-year intervals. The 3-Dimensional Index in Row 11 is the best performing local index, improving the value of index hedging relative to the City Index by five 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, driven by the increasing variance of home price shocks over time. Indeed, the dollar values for the 8-year interval are almost six times as large as those for the 2-year interval. Including multiple local indices in the regression increases the value by another percentage point or so, but the gains are not substantial relative to the 3-Dimensional Index.[9]

On the whole, the share of home price shocks attributed to local markets is not huge but not insignificant either. The relative sizes of the city shock versus the local shock versus the idiosyncratic home price shock make it such that adding local indices increases the value of index hedging by 5-10%. In the following section, I consider a more general setup where homeowners care not only about the prices of their own homes, but also about the prices of homes to which they might move.

## 5.2 Hedging multiple homes

In the more general setup, I consider homeowners who anticipate continued participation in the real estate market. An agent who plans to retain his house forever has no effective exposure to price risk. One who plans to move to another house is exposed only to the differential shocks to the two houses (plus common shock multiples of the difference in current values). Fundamentally, homeowners are ”long” their own homes and ”short” the homes they want to move into. If a homeowner is planning to move from city A to city B in 1 year, he hopes that prices rise in city A and fall in city B between now and then, but will be hurt if city B prices rise relative to city A. By going short a city A index and going long a city B index, he can reduce his overall price risk by locking in the expected price appreciation differential today rather than waiting to see how prices unfold in 1 year. But without available local indices, he has no option if he wants to hedge the possibility that he might move to another location within his current city.

As before, I utilize a regression framework to estimate the value of index hedging when multiple homes and

[9]Since home price shocks vary substantially over a typical real estate cycle, there was some concern that results would be driven by oversampling in recent years (due to more homes and better data). But performing similar calculations after weighting sales by the inverse of their frequency over time results in pretty much the same outcome.

indices are involved. Under this formulation, the agent's wealth process is a weighted sum of home price shocks and index shocks. The agent's wealth is given by $w = w_0 + \sum_{i=0}^{N} a_i\varepsilon_i + \sum_{i=0}^{N} k_i\nu_i$, where the homes and home price indices are number $0, 1, 2, ..., N$, and $a_i$ represents the agent's exposure to house $i$ (positive for currently owned homes and negative for possible future homes). The optimal hedging strategy is given by the coefficients of a regression of $\sum_{i=0}^{N} a_i\varepsilon_i$ on $\nu_0$, $\nu_1$, ... , $\nu_N$, and the fraction of overall wealth reduction equals one minus the $R^2$.

In order to simplify the analysis, I restrict the wealth process to 3 types of homes: $\varepsilon_0$ for a homeowner's current residence, $\varepsilon_1$ for a potential house in a new city, and $\varepsilon_2$ for a potential house in the same city. I define home price indices similarly, so that $\nu_i$ is the most local index relevant for home price shock $\varepsilon_i$. Thus, a variety of homeowner moving scenarios can be analyzed with the setup. Some examples include homeowners not planning to move ($a_0 = 1$, $a_1 = 0$, and $a_2 = 0$), homeowners moving to another city ($a_0 = 1$, $a_1 = -1$, and $a_2 = 0$), homeowners moving locally ($a_0 = 1$, $a_1 = 0$, and $a_2 = -1$), homeowners upgrading their homes ($a_0 = 1$, $a_1 = -2$, and $a_2 = 0$), and homeowners moving with uncertainty ($a_0 = 1$, $a_1 = -\alpha$, and $a_2 = -\pi$), where $\alpha$ is the probability of an *intercity* move and $\pi$ is the probability of a *within-city* move.[10]

As before, I model home price shocks as the sum of independent city, local, and idiosyncratic components: $\varepsilon_i = C_i + L_i + D_i$. In order to estimate the model, I assume that idiosyncratic shocks and local market shocks are uncorrelated across homes and localities and that all homes symmetrically receive component shocks with the same variances. Specifically, the formulation is:

$$\begin{pmatrix} \varepsilon_0 \\ \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} = \begin{pmatrix} C_0 + L_0 + D_0 \\ C_1 + L_1 + D_2 \\ C_0 + L_2 + D_2 \end{pmatrix},$$

$$\begin{pmatrix} D_0 \\ D_1 \\ D_2 \end{pmatrix} = N\left(0, \sigma_D^2 I\right), \quad \begin{pmatrix} L_0 \\ L_1 \\ L_2 \end{pmatrix} = N\left(0, \sigma_L^2 I\right), \quad \begin{pmatrix} C_0 \\ C_1 \end{pmatrix} = N\left(0, \sigma_C^2 \begin{bmatrix} 1 & \gamma \\ \gamma & 1 \end{bmatrix}\right).$$

Note that $\varepsilon_0$ and $\varepsilon_2$ have the same city component, $C_0$, which has a non-zero correlation ($\gamma$) with the other city component, $C_1$. The above formulation has 4 degrees of freedom ($\sigma_D^2, \sigma_L^2, \sigma_C^2, \gamma$) which completely define the system.[11] With this formulation, I also define $\sigma_\varepsilon^2 = var(C_i + L_i + D_i)$ to be the total variance of a home price shock, and $\rho_C = corr(C_i + L_i + D_i, C_i)$ and $\rho_L = corr(C_i + L_i + D_i, C_i + L_i)$ are the correlations between home price shocks and the city and local indices, respectively.

Under this setup, the optimal hedging strategy and wealth variance reduction are additive and separable. As long

[10]It would have been just as easy to restrict $\pi = 1 - \alpha$, but I choose to keep two separate parameters to more closely reflect the values in Table 1.

[11]This model makes sense insofar as local markets are "large enough" so that a local price index has a sufficient signal to noise ratio. The more homes are included in the index construction, the smaller the estimation error and the stronger is this assumption. The regression coefficients in Table 7 suggest that this is probably reasonable for most of the local indices used in the value calculations.

as the homes are in different local or city markets, index $\nu_j$ provides no additional predictive power for $\varepsilon_i$ above and beyond that already included in $\nu_i$. The optimal hedging strategy in this case is $k_i^* = -\, a_i\ \rho_i\ \frac{\sigma_{\varepsilon_i}}{\sigma_{\nu_i}} = -\, a_i\ \frac{cov(\varepsilon_i)}{var(\nu_i)}$ and the wealth variance at the optimum is $var\left(\log w \mid k_i = k_i^*\right) = \sum_{i=0}^{N} a_i^2 \left(1-\rho_i^2\right) \sigma_{\varepsilon_i}^2$. This is identical to the previous solution except for a multiplicative adjustment for $a_i$. Importantly, the optimal hedging strategy and the wealth variance do *not* depend on the correlations between the home prices. The agent uses the same hedging strategy regardless of whether his future homes are highly correlated or not.

Although the home price correlation structure has no impact on the hedged wealth variance, it does affect the unhedged wealth variance and thus the agent's willingness to pay in order to hedge. The intuition here should be obvious: a homeowner moving to more highly correlated markets has less to gain from index hedging than a homeowner moving to more independent markets. Mathematically, a homeowner's unhedged wealth variance is given by $var(\sum_{i=0}^{N} a_i\varepsilon_i)$, which is larger if the shocks are more independent since some of the coefficients are negative. The fraction of variance reduction is given by $f = \frac{var(\sum_{i=0}^{N} k_i^*\nu_i)}{var(\sum_{i=0}^{N} a_i\varepsilon_i)}$, which is just the $R^2$ of $\sum_{i=0}^{N} a_i\varepsilon_i$ regressed on $\nu_0, \nu_1, ..., \nu_N$. Hence, we get the standard result that perfect hedging instruments ($corr\left(\varepsilon_i, \nu_i\right) = 1$) reduce the residual wealth variance to 0 and completely uncorrelated hedging instruments ($corr\left(\varepsilon_i, \nu_i\right) = 0$) do not reduce the wealth variance at all.

In order to value index hedging, I denote $f^C$ and $f^L$ to be the fraction of wealth variance reduced by city and local index hedging, respectively, and reconsider the case of 3 home price shocks. The latter is given by:

$$f^L = \frac{var\left(k_0^*\nu_0 + k_1^*\nu_1 + k_2^*\nu_2\right)}{var\left(a_0\varepsilon_0 + a_1\varepsilon_1 + a_2\varepsilon_2\right)} = \frac{\rho_L^2\psi + \zeta}{\psi + \zeta}, \tag{9}$$

where $\psi = a_0^2 + a_1^2 + a_2^2$ (the variance terms) and $\zeta = 2a_0a_1\rho_C^2\gamma + 2a_1a_2\rho_C^2\gamma + 2a_0a_2\rho_C^2$ (the covariance terms). The fraction of wealth reduction is increasing in $\rho_L$, and a homeowner can remove his wealth variance entirely if $\rho_L = 1$. The case of $\rho_L = 0$ would mean that homes have zero correlation with their local index as well as their city index ($\rho_C = 0$), so the optimal strategy would be to remain unhedged. Similarly, I calculate the value of index hedging using city indices only:

$$f^C = \frac{var\left(k_0^*\nu_0 + k_1^*\nu_1\right)}{var\left(a_0\varepsilon_0 + a_1\varepsilon_1 + a_2\varepsilon_2\right)} = \frac{\rho_C^2(a_1^2 + (a_0 + a_2)^2 + 2a_1(a_0 + a_2)\gamma)}{\psi + \zeta}. \tag{10}$$

In this scenario, the homeowner is hedging three homes with two indices and cannot hedge his intercity move. Instead, he hedges the *net shock* $(a_0\varepsilon_0 + a_2\varepsilon_2)$ that arises from selling and buying into the same city market. In the case of a purely intercity move ($a_0 = 1, a_1 = 0, a_2 = -1$), the agent gets no benefit at all from city hedges. Thus, the overall improvement of local indices is the proportion of home price shocks represented by the difference: $f^L - f^C = \frac{(\rho_L^2 - \rho_C^2)\psi}{\psi + \zeta}$. However, the full benefit of local indices can only be captured if a homeowner knows the specific locality to which he is moving, otherwise he will have to his potential future homes with city indices. In this circumstance, the numerator of equation (9) changes to $(\rho_L a_0^2 + \rho_C(a_1^2 + a_2^2) + \zeta)$ and improvement of local indices reduces to: $f^L - f^C = \frac{(\rho_L^2 - \rho_C^2)a_0^2}{\psi + \zeta}$.

Ultimately, equations (9) and (10) are the main evaluative measures of the total benefit of localized indices. The risk profile of a homeowner hedging multiple homes can be substantially different than the risks faced on his own home exclusively. The important difference is that he now cares about the correlations between the home values. With this key idea in mind, I estimate the dollar value of index hedging for a variety of moving scenarios.

Before presenting the main results, it is worth reviewing the estimation procedure for the various parameters in equations (9) and (10). Table 9 presents the descriptions and estimation procedures for all relevant parameters necessary to determine the value of index hedging. The five parameters at the top of the table represent static values that are fixed throughout. The risk aversion parameter ($\lambda = 2.377$) is selected to calibrate the model, and the estimates for the others are the probability of moving in each quarter ($\hat{p} = 3.5\%$), the probability of an *intercity* move ($\hat{\alpha} = 28.1\%$), the probability of a *within-city* move ($\hat{\pi} = 58.4\%$), and the median home price ($\hat{P} = \$525{,}000$). The moving probabilities are relevant for the "representative" agent who moves with the probabilities equal to the city averages from Table 1. As mentioned, $\lambda$ is chosen to calibrate the model and $P$ is taken as the median home price in the sample.

The parameters estimates in the lower half of Table 9 vary with the time interval between sales, so multiple estimates of each are used and reported. Although the table displays only the 2-year, 5-year, and 8-year values, the full estimation graphs can be found in the Appendix. The estimates are for the standard deviation of the log home price shock ($\hat{\sigma_\varepsilon} = 0.113\text{-}0.223$), the home price correlation of the city and local indices ($\hat{\rho_C} = 0.760\text{-}0.883$ and $\hat{\rho_L} = 0.801 - 0.905$), and the average pairwise correlation of other city indices with the Washington D.C. Index ($\hat{\gamma} = 0.541\text{-}0.576$).[12] In all of the index valuations, the parameters from the appropriate time interval are used. The full plot of parameter estimates over varying time intervals can be found in the Appendix.

I estimate the value of index hedging for 6 different moving scenarios, which are listed at the top of Table 10. They are hedging one's current residence only ($a_0 = 1, a_1 = 0, a_2 = 0$), moving to a new city ($a_0 = 1, a_1 = -1, a_2 = 0$), moving locally ($a_0 = 1, a_1 = 0, a_2 = -1$), moving to an unknown location ($a_0 = 1, a_1 = -\alpha, a_2 = -\pi$), and moving with probability less than 1 ($a_0 = 1, a_1 = -p\alpha, a_2 = -p\pi$, or $a_0 = p, a_1 = -p\alpha, a_2 = -p\pi$). The reason for two listings in the last case is to provide flexibility in the interpretation, the difference being whether the agent absorbs his own home's price shock with certainty ($a_0 = 1$) or with his probability of moving ($a_0 = p$). Either interpretation could make sense in the right circumstances, though I prefer the former. The probability of moving for each time interval is calculated using a poisson process with a constant quarterly hazard rate of $\hat{p} = 3.5\%$. In all cases, I use the 3-Dimensional Index as the local hedge.

[12] I estimate $\gamma$ using the average pairwise correlations of the S&P/Case-Shiller city indices with the Washington D.C. Index as published by the Chicago Mercantile Exchange. For the estimation, I take the 13 cities besides Washington, D.C. that have published indices all the way back to 1987 (Boston, Charlotte, Chicago, Cleveland, Denver, Las Vegas, Los Angeles, Miami, New York, Portland, San Diego, San Francisco, and Tampa), and calculate log price changes over various time intervals. I then calculate the across-time correlations with the Washington, D.C. Index and average over the 13 cities. Although the risk profile depends on the specific city to which a homeowner is moving, I do it this way to represent an equal chance of moving to each. Intuitively, the lower the between city correlation, the more value there is to hedging with a city index.

Whereas the values in Table 8 reflect only the improved correlations of local indices, Table 10 additionally includes the value-added of hedging *within-city* moves. Readers should recognize Column (1) as the same information as the 3-Dimensional Index from Table 8, where the 2-year value of the city index is \$2,679, the local index is \$2,978 (\$2679 + \$299), and the perfect hedge is \$4,650 (\$2679 + \$299 + \$1672). The subsequent columns represent the corresponding values for the other moving scenarios.

Column (2) presents the values for hedging an intercity move. By adding low-correlated price risk in the form of a potential future home, the total risk over 2-years as measured by the value of the perfect hedge increases from \$4,650 in Column (2) to \$6,405. The additional risk comes from roughly doubling the local and idiosyncratic components. The value of hedging with city indices actually decreases due to the positive correlation between the city indices. In total, adding local indices increases the dollar value of hedging an intercity move by around 25% for a 2-year time interval and 12% for the 5-year and 8-year intervals. Column (3) presents the values for hedging a local move. The total risk in this scenario is less than the is less than in Columns (1) and (2) since the agent is trading one asset for another highly correlated asset. Even the specific row corresponding to the local component is lower than Column (2) since the local markets themselves are more correlated. However, the local index allows homeowners to hedge 10-20% of their home price risk compared to none at all with city indices. In both of these two columns, only half of the benefit of local indices is captured if a homeowner does not know the specific locality to which he is moving. In this case, he would hedge his current home with a local index (since he certainly knows its location!) and the future home with a city index to represent a random probability of moving to each house within the city.

The remaining columns present variations of a "representative" agent who moves with probabilities equal to the county averages in Table 1. Recall that $p$ is the probability of moving (which increases over longer time intervals) and that $\alpha = 28.1\%$ and $\pi = 58.4\%$ are the probabilities of intercity and within-city moves, respectively, conditional on moving. Column (4) is a homeowner who moves with certainty but does not know whether his location will be within-city or intercity. Column (5) and (6) represent a homeowner who moves according to a poisson process every period. The value of index hedging is typically lower in these cases since the probability of experiencing a non-correlated home price shock is lower. In sum, the real risk in these cases arises from the homeowner with exposure to more independent housing markets or with differential exposure across homes.

Fundamentally, the value of index hedging depends on the way one thinks about home price risk and also on what measuring stick is used. If homeowners care about price shocks only to the extent that *relative* prices between homes change, then the risk only matters when a transaction occurs. On the other hand, homeowners should care about home prices relative to all other goods and assets since their wealth measured in dollars is changing. I could have explored these interpretations in more detail by perhaps considering all home price shocks adjusted for inflation or home price shocks relative to other asset classes. In the end, there are a multitude of ways to tweak the model to more accurately represent home price risk for a particular homeowner. Ultimately, I make my best effort to capture what seems to be the risk profile of a typical homeowner and to determine the value of index hedging accordingly.

## 6 Conclusion

In this paper, I define and evaluate a variety of local markets within a broader metropolitan area in an attempt to learn about the size and scope of local real estate markets. In an effort to optimally capture the local market component of home prices, I use locally weighted regression techniques that maximize home price explanatory power over continuous variables such as location, home type, and price. One would never expect even the best local indices to explain home price shocks completely since there will always be some house-specific risk component that can never be hedged. Overall, I would expect that the local indices in this paper approach this upper bound of explanatory power because of their flexible maximization over several dimensions.

Ultimately, I demonstrate that there is a fair amount of local variation that can be captured with local market indices. Depending on the time interval in question, as little as 3% or as much as 7% of home price shock variation can be attributed to local markets. Overall, I find markets to be quite local, with the best market information coming from local markets defined over roughly 10 square miles and 3% of the county residences. The findings in this paper suggest that we at least need to be thinking of local markets at least as small as 7,500 homes.

The two main benefits of adding local market indices is that homeowners will hedge more of their home price risk and the hedges will perform better. I estimate that local market indices would increase the value of index hedging by 5-10% for a typical homeowner and as much as 25% for homeowners facing various moving scenarios. Because local market indices allow homeowners to hedge local moves, they additionally provide value where city indices have none.

Although the estimates in this paper come from a single county, one can easily imagine extending the analysis to a broader set of local and metropolitan markets. Since over one quarter of all local moves occur across county lines (see Table 1), local market moves occur in practice over less correlated markets and would indicate a higher valuation of local market indices. In addition, Montgomery County residents have a particularly high value for the metro-level index since its housing market is so closely tied to Washington D.C.. Thus, I would expect the potential gains from local market indices to be even greater in practice for counties further away from the city center.

The introduction and standardization of index futures contracts on the Chicago Mercantile Exchange offers new and exciting prospects for home price risk management using financial derivatives. Especially given the recent downward pressure on home prices in the United States, home price movements can have large wealth effects for millions of Americans. For example, in Montgomery County, MD, the median home value is roughly five times the median household income, so a 10% change in home prices is on the order of magnitude of six months of wages. Businesses whose earnings depend on a healthy housing market may be even more exposed to home price declines than individual homeowners. To the extent that financial derivatives can transfer risk from those who don't want it to those who do, further development in these markets should certainly be viewed as a step in the right direction.

There are many benefits to homeownership that may not be captured if individuals are too sensitive to home price risk. By reducing this risk, hedged homeowners can purchase more housing than they otherwise would in an unhedged position. Over time, I would expect active management of home price risk to become more commonplace as hedging opportunities expand into the retail markets. All else being equal, this is a good thing. This paper shows that city indices have a lot of value already and that local indices can add even more.

# References


[1] Martin J. Bailey, Richard F. Muth, and Hugh O. Nourse. A regression method for real estate price index construction. *Journal of the American Statistical Association*, 58:933–942, 1963.

[2] Michel Baroni, Fabrice Barthlmy, and Mahdi Mokrane. Is it possible to construct derivatives for the paris residential market? *Working Paper*, 2007.

[3] Steven C. Bourassa, Martin Hoesli, and Vincent S. Peng. Do housing submarkets really matter? *Journal of Housing Economics*, 12(1):12–28, 2003.

[4] Steven C. Bourassa, Martin Hoesli, and Jian Sun. A simple alternative house price index method. *Journal of Housing Economics*, 15(1):80–97, 2006.

[5] Karl E. Case and Robert J. Shiller. Prices of single-family homes since 1970: New indexes for four cities. *New England Economic Review*, pages 45–56, 1987.

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

[7] Eric Clapham, Peter Englund, John M. Quigley, and Christian L. Redfearn. Revisiting the past and settling the score: Index revision for house price derivatives. *Real Estate Economics*, 34(2):275–302, 2005.

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

[9] Angus Deaton. *The Analysis of Household Surveys*. The Johns Hopkins University Press, Baltimore, MD, 1997.

[10] Yongheng Deng and John M. Quigley. Index revision, house price risk, and the market for home price derivatives. *Fisher Center for Real Estate and Urban Economics. Working Paper No. W07-003*, 2007.

[11] Jianqing Fan. Design-adaptive nonparametric regression. *Journal of the American Statistical Association*, 87(420):998–1004, 1992.

[12] Arthur Getis, Jesus Mur, and Henry G. Zoller. *Spatial Econometrics and Spatial Statistics*, chapter 12: Locally Weighted Regresion and Time-Varying Distance Gradients, by Daniel P. McMillen, pages 232–249. Palgrave MacMillan, Great Britain, 2004.

[13] Allen C. Goodman and Thomas G. Thibodeau. Housing market segmentation. *Journal of Housing Economics*, 7(2):121–143, 1998.

[14] Allen C. Goodman and Thomas G. Thibodeau. Housing market segmentation and hedonic prediction accuracy. *Journal of Housing Economics*, 12(3):181–201, 2003.

[15] Allen C. Goodman and Thomas G. Thibodeau. The spatial proximity of metropolitan area housing submarkets. *Real Estate Economics*, 35(2):209–232, 2007.

[16] CME Group. *S&P/Case-Shiller Home Price Indices Futures and Options, Introductory Guide*. Chicago, IL, 2007. http://www.cme.com/housing.

[17] Duncan Maclennan and Yong Tu. Economic perspectives on the structure of local housing systems. *Housing Studies*, 11(3):387–406, 1996.

[18] Robert J. Shiller and Allan N. Weiss. Home equity insurance. *Journal of Real Estate Finance and Economics*, 19(1):21 – 47, 1999.

[19] Standard and Poor's. *S&P/Case-Shiller Metro Area Home Price Indices*. The McGraw Hill Companies, Inc., New York, NY, 2006.

**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% |

Notes: The above table displays 5-year migration rates for Montgomery County, MD and the State of Maryland according to the 2000 United States Census. For the purposes of this paper, a "within-city" or "local" move is defined to be any move within the same state (22.48% + 6.15% for Montgomery County). Parameter estimates used throughout the paper are taken from Montgomery County column, although the above table indicates that the state averages are quite similar.

Table 2: Summary statistics for the repeat homes sales

| Variable | Units | N | Mean | SD | Min | Max |
|---|---|---|---|---|---|---|
| **Land value** | ($1,000s) | 75947 | 274.2 | 144.5 | 15.0 | 2440.0 |
| **Building value** | ($1,000s) | 75947 | 245.2 | 174.4 | 23.3 | 4416.5 |
| **Total assessment** | ($1,000s) | 75947 | 519.4 | 286.7 | 87.2 | 6076.0 |
| **Square footage** | (1,000s of sq ft) | 75947 | 1.80 | 0.86 | 0.42 | 17.65 |
| **Land area** | (acres) | 75944 | 0.26 | 0.55 | 0.00002 | 30.45 |
| **Year built** | (year) | 75947 | 1974 | 18.4 | 1900 | 2006 |
| **Construction grade** | (1=worst, 9=best) | 75929 | 4.39 | 0.75 | 2 | 9 |
| **Condition** | (1=worst, 6=best) | 75841 | 3.07 | 0.26 | 1 | 5 |
| **Longitude** | (degrees) | 75947 | -77.14 | 0.10 | -77.47 | -76.90 |
| **Latitude** | (degrees) | 75947 | 39.10 | 0.07 | 38.94 | 39.33 |
| **Price, 1st Sale** | ($1,000s) | 75947 | 223.5 | 165.8 | 2.0 | 5000.0 |
| **Price, 2nd Sale** | ($1,000s) | 75947 | 356.0 | 257.5 | 14.8 | 5325.0 |
| **Quarter, 1st Sale** | (1=1985-Q1) | 75947 | 37.5 | 21.3 | 1 | 84 |
| **Quarter, 2nd Sale** | (1=1985-Q1) | 75947 | 67.4 | 14.3 | 22 | 88 |

| Partitions | N | Mean # | SD # | Min # | Max # |
|---|---|---|---|---|---|
| **District** | 11 | 6904 | 5143 | 2064 | 16987 |
| **Zip Code** | 28 | 2712 | 1706 | 1100 | 8413 |
| **Home Type** | 11 | 6904 | 1 | 6903 | 6906 |
| **Price Band** | 11 | 6904 | 6 | 6891 | 6920 |

Notes: 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 roughly corresponding to one or 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 reason for having 11 submarkets for the Home Type and Price Band partitions is so that each will match the econometric power of the District partition given by the data.

## Table 3: Summary of home price shock variables

| | Variable | N | Full Sample Mean | Full Sample SD | Time Series, 1-year Mean | Time Series, 1-year 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** | **2500** | 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** | **6500** | -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** | **16000** | -0.001 | 0.054 | -0.001 | 0.029 |
| 14 | **3-Dimensional Index** | **6000** | -0.002 | 0.081 | 0.002 | 0.053 |

Notes: The above table summarizes the home price shocks for all 75,947 homes used in the main regressions of the paper. The first two columns make no adjustment for time of sale or time between sales. The second two columns are for 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 the rest of the indices are computed from the data. The County Index subtracts the appreciation from the City Index, and all other indices subtract the appreciation from both the City Index and the County Index. Indices 4-7 are equally weighted indices for the partitions described in Table 2, and 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. Empirical details on how these indices are constructed can be found in Section 4.

Table 4: Explained home price shocks: partition indices

| Dependent Variable: Change in log home price | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| **City Index** | 1.020 | 0.984 | 0.987 | 0.988 | 0.988 | 0.989 | 0.991 |
| | (0.005) | (0.005) | (0.004) | (0.004) | (0.004) | (0.004) | (0.004) |
| **County Index** | | 0.938 | 0.939 | 0.940 | 0.938 | 0.939 | 0.940 |
| | | (0.047) | (0.040) | (0.040) | (0.044) | (0.043) | (0.037) |
| **District Index** | | | 0.800 | | | | 0.340 |
| | | | (0.037) | | | | (0.040) |
| **Zip Code Index** | | | | 0.669 | | | 0.328 |
| | | | | (0.031) | | | (0.038) |
| **Home Type Index** | | | | | 0.815 | | 0.243 |
| | | | | | (0.036) | | (0.039) |
| **Price Band Index** | | | | | | 0.840 | 0.422 |
| | | | | | | (0.034) | (0.039) |
| **Constant** | 0.011 | 0.012 | 0.010 | 0.010 | 0.009 | 0.009 | 0.007 |
| | (0.003) | (0.003) | (0.002) | (0.002) | (0.003) | (0.002) | (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** | 75947 | 75947 | 75947 | 75947 | 75947 | 75947 | 75947 |
| **$R^2$** | 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 |

Notes: The above columns regress the changes in the log home prices on the changes in the log home price indexes for the entire sample of repeat home sales. The City Index is the published S&P/Case-Shiller Index and 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 contain the same data. The District Index and the Zip Code Index are based on partitions pre-coded in the data, the Home Type Index is based on the first principal component of home characteristics, and the Price Band Index is based on fitted values from a hedonic price regression. See Section 4.1for calculation details. In order to observe residual effects, the County Index subtracts the City Index appreciation, and all other indices subtract both the City Index and County Index appreciation. The Net City Index and the Net County Index rows present the coefficients when regressing on full indices rather than index residuals. All standard errors are clustered by subdivision.

## Table 5: Optimal bandwidths for the continuous kernel indices

| | Partition Indices | | | | | | | Continuous Indices | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | Nearest Neighbor | | Nearest Type | | Nearest Price | |
| | City | County | District | Zip | Type | Price | $R^2$ | N | $R^2$ | N | $R^2$ | N | $R^2$ |
| 1 | . | . | . | . | . | . | . | 4000 | 0.739 | 7500 | 0.741 | 17000 | 0.733 |
| 2 | y | y | . | . | . | . | 0.715 | 2500 | 0.741 | 6500 | 0.741 | 16000 | 0.733 |
| 3 | y | y | y | . | . | . | 0.729 | 2250 | 0.741 | 4250 | 0.745 | 15000 | 0.739 |
| 4 | y | y | . | y | . | . | 0.730 | 2500 | 0.741 | 4250 | 0.745 | 15000 | 0.740 |
| 5 | y | y | . | . | y | . | 0.728 | 2500 | 0.746 | 5000 | 0.741 | 15000 | 0.733 |
| 6 | y | y | . | . | . | y | 0.731 | 2250 | 0.746 | 4000 | 0.742 | 16000 | 0.733 |
| 7 | y | y | y | y | y | y | 0.740 | 2250 | 0.747 | 3250 | 0.747 | 14000 | 0.742 |

Notes: 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 Figure 3.

**Table 6: 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** | **2500** | 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** | **6500** | 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** | **15000** | 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** | **6000** | 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 |

Notes: The above table presents the index correlations for the shocks summarized in Table 3. The correlations are equally weighted over all 75,497 homes in the sample, so no adjustments are made for time of sale or time between sales. Indices 4-14 subtract the appreciation from the County Index and the City Index, which is why Columns 2 and 3 contain correlations close to zero. Likewise, the County Index subtracts the appreciation from the City Index; the positive correlation here arises since Montgomery County has tended to have a high market beta to the overall Washington D.C. market. As in Table 3, $N$ refers to the number of homes included in the index calculation.

## 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=2500** | | | | 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=6500** | | | | | | 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=15000** | | | | | | | | 0.987 | | 0.247 | 0.114 |
| | | | | | | | | (0.038) | | (0.036) | (0.033) |
| **3-Dimensional Index N=6000** | | | | | | | | | 0.911 | | 0.457 |
| | | | | | | | | | (0.029) | | (0.039) |
| **Constant** | 0.011 | 0.012 | 0.012 | 0.010 | 0.012 | 0.010 | 0.012 | 0.010 | 0.011 | 0.009 | 0.010 |
| | (0.003) | (0.003) | (0.002) | (0.002) | (0.003) | (0.002) | (0.003) | (0.002) | (0.002) | (0.002) | (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** | 75947 | 75947 | 75947 | 75947 | 75947 | 75947 | 75947 | 75947 | 75947 | 75947 | 75947 |
| **$R^2$** | 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 |

Notes: The above columns regress the changes in the log home prices on the changes in the log home price indexes for the entire sample of repeat home sales. The City Index is the published S&P/Case-Shiller Index and 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 contain 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), respectively. The 3-Dimensional Index utilizes a 3-dimensional epanechnikov kernel over distance, type, and price. See Section 4.2 for calculation details. In order to observe residual effects, the County Index subtracts the City Index appreciation, and all other indices subtract both the City Index and County Index appreciation. The Net City Index and the Net County Index rows present the coefficients when regressing on full indices rather than index residuals. All standard errors are clustered by subdivision.

## Table 8: The value of index hedging

| Index Hedge | 2-years ρ | 2-years Hedge Value | | 5-years ρ | 5-years Hedge Value | | 8-years ρ | 8-years Hedge Value | |
|---|---|---|---|---|---|---|---|---|---|
| 1 Unhedged | 0.000 | $ - | 0% | 0.000 | $ - | 0% | 0.000 | $ - | 0% |
| 2 City Index | 0.760 | $ 2,679 | 58% | 0.883 | $ 7,782 | 78% | 0.830 | $ 12,560 | 68% |
| 3 County Index | 0.774 | $ 2,778 | 60% | 0.885 | $ 7,817 | 78% | 0.833 | $ 12,655 | 69% |
| 4 District Index | 0.791 | $ 2,901 | 62% | 0.895 | $ 7,995 | 80% | 0.841 | $ 12,909 | 70% |
| 5 Zip Code Index | 0.788 | $ 2,883 | 62% | 0.893 | $ 7,962 | 80% | 0.842 | $ 12,934 | 71% |
| 6 Home Type Index | 0.795 | $ 2,932 | 63% | 0.897 | $ 8,025 | 80% | 0.841 | $ 12,903 | 70% |
| 7 Price Band Index | 0.798 | $ 2,954 | 64% | 0.898 | $ 8,049 | 80% | 0.842 | $ 12,946 | 71% |
| 8 Nearest Neighbor Index | 0.791 | $ 2,901 | 62% | 0.899 | $ 8,063 | 81% | 0.844 | $ 13,015 | 71% |
| 9 Nearest Type Index | 0.799 | $ 2,966 | 64% | 0.901 | $ 8,107 | 81% | 0.848 | $ 13,121 | 72% |
| 10 Nearest Price Index | 0.799 | $ 2,963 | 64% | 0.899 | $ 8,061 | 81% | 0.843 | $ 12,981 | 71% |
| 11 3-Dimensional Index | 0.801 | $ 2,978 | 64% | 0.905 | $ 8,177 | 82% | 0.851 | $ 13,234 | 72% |
| 12 Partition Indices, 4-7 | 0.806 | $ 3,014 | 65% | 0.902 | $ 8,128 | 81% | 0.848 | $ 13,121 | 72% |
| 13 Continuous Indices, 8-10 | 0.807 | $ 3,024 | 65% | 0.906 | $ 8,190 | 82% | 0.852 | $ 13,240 | 72% |
| 14 All Indices, 2-11 | 0.810 | $ 3,042 | 65% | 0.907 | $ 8,218 | 82% | 0.854 | $ 13,300 | 73% |
| 15 Perfect Hedge | 1.000 | $ 4,650 | 100% | 1.000 | $ 10,001 | 100% | 1.000 | $ 18,344 | 100% |

Notes: The above table displays the dollar value of index hedging for an individual trying to minimize the variance of his home price shock with an index hedge. $\rho$ is the correlation between log home price shocks and the log home price index shocks. The middle column is an agent's willingness to pay for an index hedge, according to the formulation in Section 5.1. 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 $\lambda = 2.377$, which is calibrated based on the value of a 5-year Perfect Hedge equalling $10,000.

**Table 9: Parameter estimation**

| Parameter | Description | Estimation | Value |
|---|---|---|---|
| $\lambda$ | Risk aversion, U(w)=-w^(1-λ)/(1-λ) | Chosen by assumption to produce a 5-year perfect hedge value of $10,000 | 2.377 |
| $p$ | Prob[moving per quarter] | Constant rate to produce the census 5-year moving rate, from Table 1 | 3.50% |
| $\alpha$ | Prob[intercity move \| moving] | Fraction of movers who moved to a different state, from Table 1 | 28.07% |
| $\pi$ | Prob[within-city move \| moving] | Fraction of movers who moved within the state, from Table 1 | 58.37% |
| $P$ | Median home price | Median sales price in 2006, full data | $525.3k |

| Parameter | Description | Estimation | 2-years | 5-years | 8-years |
|---|---|---|---|---|---|
| $\sigma_\varepsilon$ | Standard deviation of home price shock | Average sd(ΔlnP) over various points in the real estate cycle, full data | 0.113 | 0.166 | 0.223 |
| $\rho_C$ | Home price and city index correlation | Regression of ΔlnP on Δln(city index), full data | 0.760 | 0.883 | 0.830 |
| $\rho_L$ | Home price and local index correlation | Regression of ΔlnP on Δln(local index), full data | 0.801 | 0.905 | 0.851 |
| $\gamma$ | Correlation between 2 city indices | Average pairwise city index correlation with the DC Metro Index, CME data | 0.541 | 0.545 | 0.576 |

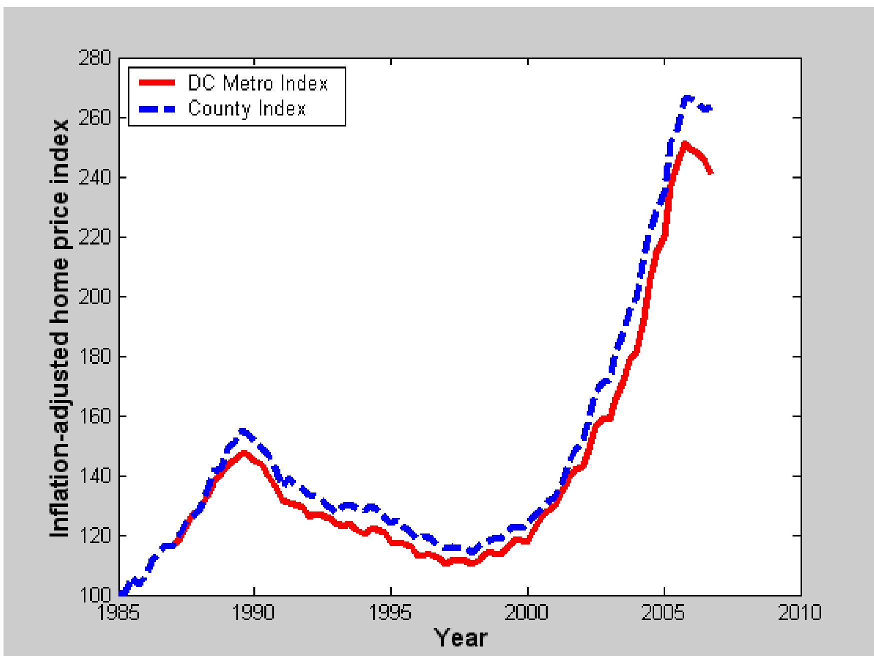
Notes: The above table displays the parameter estimates performed for the primary value calculation in this paper. The top five parameters are estimated as singular values, whereas the bottom five parameters depend on the time interval between sales. The 2-year, 5-year, and 8-year values are displayed for the latter although in practice these parameters are estimated for each quarter from 1.5 years to 8.5 years. See Section 5.2 for more details on how these parameters are calculated and utilized.

**Table 10: The value of localized home price indices for various moving scenarios**

| | (1) Hedge current home only | (2) Move to a new city | (3) Move locally | (4) Move to unknown loc. | (5) Move with prob. < 1 | (6) Move with prob. < 1 |
|---|---|---|---|---|---|---|
| Current home: $a_0$ | 1 | 1 | 1 | 1 | 1 | p |
| Intercity move: $a_1$ | 0 | -1 | 0 | $-\alpha$ | $-p\alpha$ | $-p\alpha$ |
| Within-city move: $a_2$ | 0 | 0 | -1 | $-\pi$ | $-p\pi$ | $-p\pi$ |
| **2-YEARS** | | | | | | |
| City component | $ 2,679 | $ 2,459 | $ - | $ 336 | $ 1,799 | $ 21 |
| Local component | $ 299 | $ 597 | $ 594 | $ 422 | $ 306 | $ 26 |
| Idiosyncratic component | $ 1,672 | $ 3,349 | $ 3,334 | $ 2,364 | $ 1,712 | $ 144 |
| Perfect hedge | $ 4,650 | $ 6,405 | $ 3,928 | $ 3,122 | $ 3,817 | $ 191 |
| **5-YEARS** | | | | | | |
| City component | $ 7,781 | $ 7,070 | $ - | $ 963 | $ 3,134 | $ 250 |
| Local component | $ 396 | $ 791 | $ 780 | $ 555 | $ 436 | $ 144 |
| Idiosyncratic component | $ 1,824 | $ 3,652 | $ 3,605 | $ 2,559 | $ 2,005 | $ 661 |
| Perfect hedge | $ 10,000 | $ 11,513 | $ 4,385 | $ 4,077 | $ 5,575 | $ 1,055 |
| **8-YEARS** | | | | | | |
| City component | $ 12,560 | $ 10,634 | $ - | $ 1,460 | $ 3,332 | $ 675 |
| Local component | $ 673 | $ 1,341 | $ 1,314 | $ 935 | $ 789 | $ 431 |
| Idiosyncratic component | $ 5,111 | $ 10,246 | $ 10,044 | $ 7,124 | $ 6,004 | $ 3,274 |
| Perfect hedge | $ 18,344 | $ 22,221 | $ 11,358 | $ 9,519 | $ 10,125 | $ 4,380 |

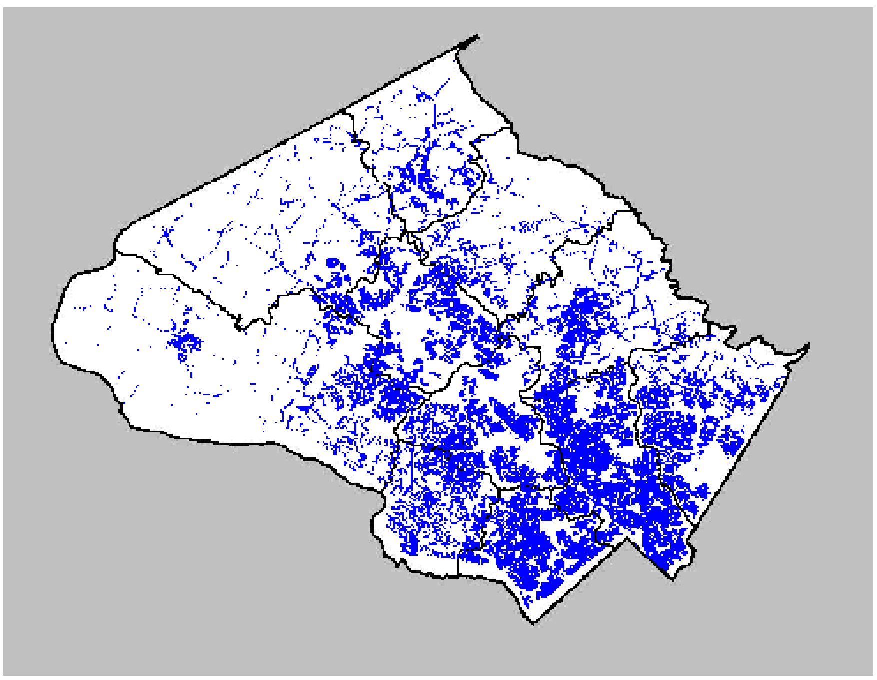
Notes: The above table displays the value of index hedging for various moving scenarios, for which the wealth process has the indicated coefficients. The valuations are assigned based on three components which sum to the value of the perfect hedge. The city index component indicates the amount of value provided by hedging the city-level shock. The local index component indicates the additional value of being able to hedge local shocks. The unhedged risk component represents the idiosyncratic risk that was not able to be hedged by the indices. See Section 5.2 for more details on the calculations used to reach the numbers above.

**Figure 1: Inflation-adjusted home price indices for the full time period, 1985-2006**



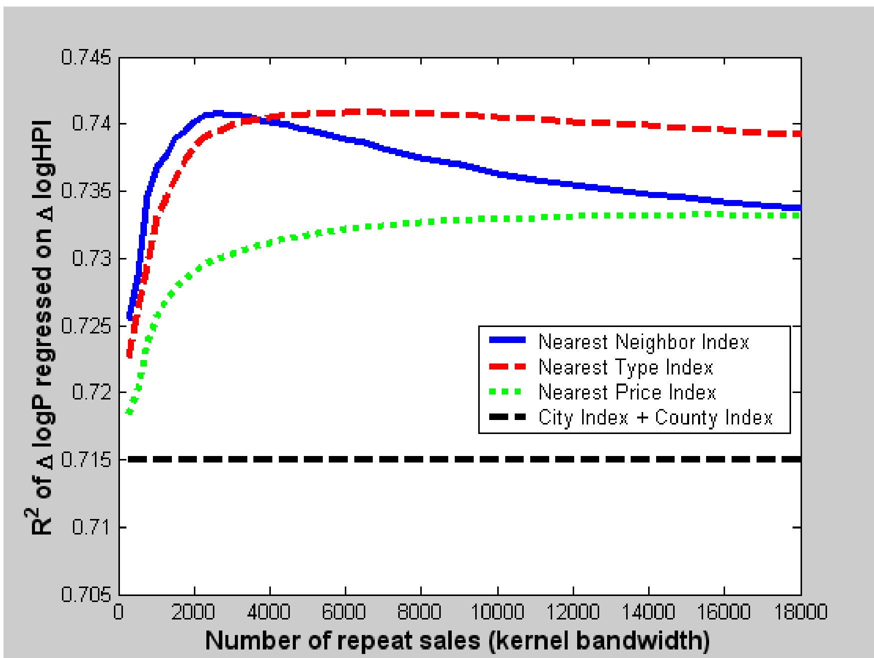

Notes: The above figure plots inflation-adjusted home price indices for the Washington D.C. Metropolitan Area and Montgomery County, MD. The former is published by Standard & Poor's and the latter is calculated using the full data set of 75,497 repeat home sales. Both are adjusted by the Consumer Price 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.

**Figure 2: Geocoded homes in Montgomery County, MD**



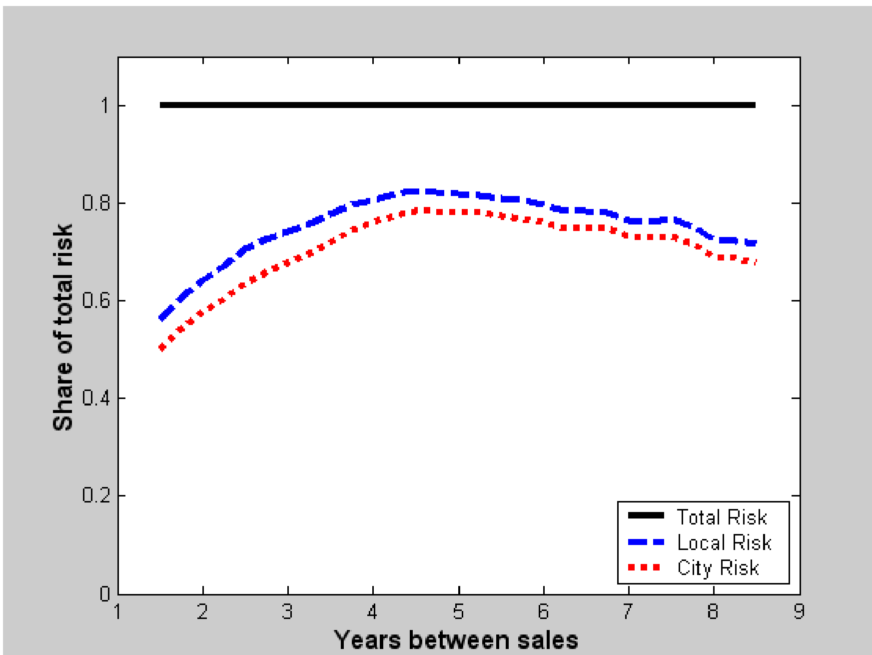
Notes: The above figure plots the approximate locations of the 200,493 single-family homes in Montgomery County, MD for which the geocoding program was able to locate the addresses. The 'District' boundaries are drawn in black. Washington D.C. is located directly to the southeast of the county.

**Figure 3: Maximizing the home price index correlation over the kernel bandwidth**



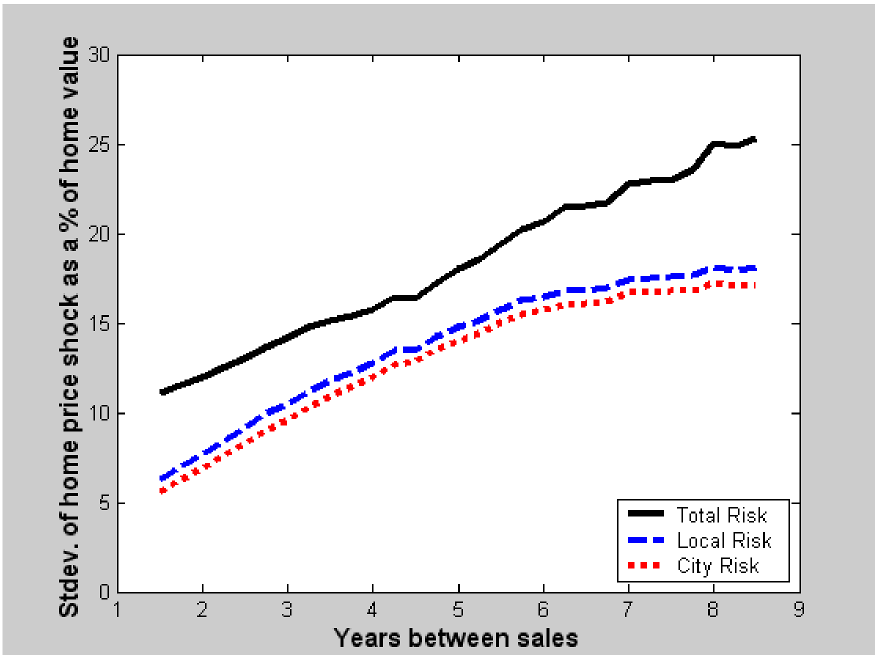

Notes: The above graph plots the $R^2$ when regressing changes in log home prices on changes in the log home price indices as a function of the number of repeat sales included in the index calculation. The indices are calculated using an epanechnikov kernel over the nearest N homes, so that homes just inside the cutoff have a tiny 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.

**Figure 4: Variance decomposition of home price risk**



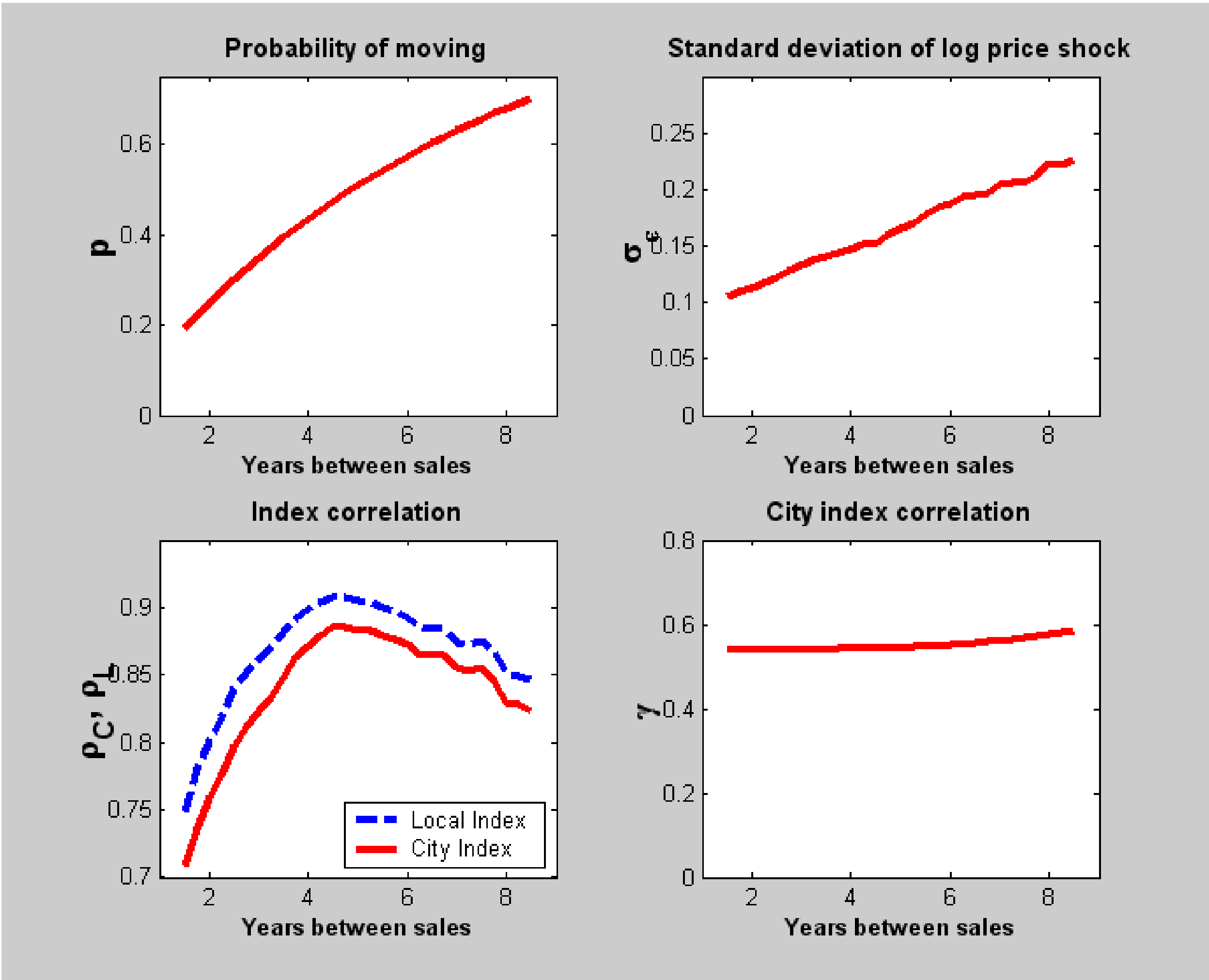

Notes: The above figure plots the variance decomposition of home price risk as a function of the years between sales. The share of citywide risk lies below the dotted red line, the share of local risk lies between the dashed blue line and the dotted red line, and the share of idiosyncratic risk lies 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 from Tables 2 and 6, 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.

**Figure 5: Value decomposition of home price risk in dollar terms**




Notes: 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 citywide risk lies below the dotted red line, the share of local risk lies between the dashed blue line and the dotted red line, and 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 from Tables 2 and 6, 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.

**Figure A–1: Parameter estimates over the time between sales**




Notes: The above figure plots the parameters from Table 9 as they vary over the years between repeated home sales. The City Index refers to the S&P/Case-Shiller Washington D.C. Metropolitan Index, and the Local Index refers to the 3-Dimensional Index from in Tables 2 and 6, which maximized price 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.