

5-1-2004

Alternative Methods of Increasing the Precision of Weighted Repeat Sales House Prices Indices*

Michelle H. Dreiman

Office of Federal Housing Enterprise Oversight

Anthony Pennington-Cross

Marquette University, anthony.pennington-cross@marquette.edu

Accepted version. *Journal of Real Estate Finance and Economics*, Vol. 28, No. 4 (May 2004): 299-317.

DOI. © 2004 Springer Publishing Company. Used with permission.

Anthony Pennington-Cross was affiliated with the Office of Federal Housing Enterprise Oversight at the time of publication.

[Shareable Link](#). Provided by the Springer Nature [SharedIt](#) content-sharing initiative.

Alternative Methods of Increasing the Precision of Weighted Repeat Sales House Prices Indices*

Authors: Michelle H. Dreiman, Office of Federal Housing Enterprise Oversight, Office of Policy & Research, 1700 G Street, NW, Washington, DC 20552, E-mail: sdreiman@ofheo.gov; Anthony Pennington-Cross, Office of Federal Housing Enterprise Oversight, Office of Policy & Research, 1700 G Street, NW, Washington, DC 20552, E-mail: anmpc@yahoo.com

Abstract: *Weighted repeat sales house price indices have become one of the primary indicators used to identify housing market conditions and to estimate the amount of equity homeowners have gained through house price appreciation. The primary reason for the acceptance of this methodology is that it derives a location specific (typically, census division, state or metropolitan area) average change in house prices from repeated observations of individual house prices. It is this repeat attribute that allows repeat sales price indices to claim that it is a preferable index which does a better job of holding quality constant.*

The amount of time between the two observed prices for a single property is determined by when the home transacts. Some homes transact twice in a period of months and others do not transact for decades. It is likely that individual house price appreciation rates vary from the mean appreciation rate, as estimated by the index, in a systematic fashion. In general, the longer the time between transactions the more variance there is in individual house price appreciation.

This paper extends this concept to include new dimensions. For instance, houses that appreciate faster than the mean, as estimated by the index for that location, may experience a different variation structure than homes that appreciate slower. This process can be viewed as an asymmetric treatment of the variance of house price appreciation around the estimated index. In addition, the variance of expensive and affordable homes may also be different and time varying.

This paper finds evidence that adding the dimensions of price tiers and asymmetry to the variance estimate has merit and does affect the estimated index as well as homeowner equity estimates. Homeowner equity estimates are especially sensitive to these added dimensions because they depend on both the revised index and the estimated variances, which are specific to each dimension considered—time between transaction, asymmetry, and price tier.

Introduction

Using a large database of mortgage transactions either purchased or securitized by Fannie Mae and Freddie Mac, this paper estimates a house price index (HPI) using the weighted repeat sales (WRS) methodology developed by Case and Shiller (1987).

WRS attempts to hold quality constant by examining only properties with repeat transactions, each of which provides a measure of the appreciation between two time periods. Using the repeated transactions, the average growth rate of house prices is estimated for a variety of locations. In this paper, we include estimates at the state level. The WRS uses an econometric analysis to form an index that has become the standard way to measure average house price appreciation or depreciation in specific geographic areas over time. The index preserves the intuitively simple interpretation of any index. For example, if the index is 100 in state j in 2000 and increases to 105 in state j in 2001, the average house price in state j increased by 5 percent over the period 2000-2001.

While the large sample size and constant quality nature of WRS indices are major strengths, various outstanding issues and caveats have been documented. For example, any “repeat sales” or WRS index technically cannot be constant quality to the extent that home improvements or physical depreciations occur between transactions. The passage of time itself can help to increase the value (e.g., a small sapling grows into a majestic tree) of some properties and decrease it in others (e.g., a majestic tree is blown over in a thunderstorm). The simple fact that trees on a property may grow or die between transactions and change property value suggest that “constant quality” may be overstated. In addition, the index may exhibit a hidden bias because certain properties may have more repeated observations than others (Clapp and Giaccotto, 1998). This paper is based on data that is limited to collateralized loans that meet the underwriting requirements and conforming loan limits of Fannie Mae and Freddie Mac. Consequently the index does not necessarily reflect value changes in properties without mortgages or properties with nonconforming loans, which may be more prevalent in some regions. Because the indices are revised with the addition of new historical data as new transaction pairs become available, the WRS indices are also subject to revision volatility. Incidentally, prior to the advent of the joint Fannie Mae and Freddie Mac data for the purpose of WRS index estimation, there were few data sets of sufficient size to produce price indices based on this method. As a result, Case and Quigley (1991) introduce a hybrid approach that jointly estimates a combination hedonic and WRS index using a neighborhood data set. The hybrid approach allows one to take advantage of the information contained in transactions which occur

only once.¹ However, the combined Fannie Mae and Freddie Mac data is sufficient to produce a sound WRS index. For a detailed discussion of potential biases associated with WRS indices, see Calhoun (1991).

In this paper, we address the role played by the time between transactions and the variance of the error term during the estimation. We find qualified support for the conventional approach, which views the variance and the resulting standard error of the index as an increasing function of the time between transactions. In addition, we introduce two new dimensions—asymmetry and house price tiers. The asymmetric approach allows the variance to be different for properties that appreciate above and below the estimated mean appreciation rate. The house price tier approach allows the variance to be different depending on whether the property is expensive, moderate, or affordable.

The results provide considerable support for the existence of these additional dimensions to the variance of the HPI error structure. There are many potential explanations for the observed relationships between the variance and the three dimensions (time between transaction, asymmetry, and house price tier) that will be discussed later in the paper. One example is that properties that are purchased at a very low price typically require relatively more owner improvements. Some properties will undergo dramatic renovations while some may be allowed to deteriorate and continue to filter down into lower price tiers. If these activities are more frequent for lower priced properties then the variance should be higher for these types of properties. The opposite may be true for homes that transact at relatively high prices since they may need less renovation work. These properties would thus experience relatively less variation in house price appreciation. In addition, since some expensive homes whose loans conform to the Fannie Mae and Freddie Mac loan limits may appreciate faster than the loan limits over time, it is likely that some of these properties with larger increases in value will move outside of the sample, as subsequent loans on the property may exceed the conforming limits.

The proper weighting schemes for both sets of homes are imperative if the WRS process is to be efficient and appropriately indicative of actual changes in house prices. Homeowner equity can play an important role in determining the perceived riskiness of any mortgage application. This is because borrowers with negative equity (the mortgage is larger than the house value) in a home find it “in the money” to default on the mortgage and move out of the house. Therefore, any revisions to HPI that impact estimates of the equity growth or decline for homeowners can have impacts on the availability of credit as well as mortgage pricing for both the borrower and the investor.

Organization of paper

This paper begins by briefly describing the typical WRS method. In the section that follows, we discuss results obtained by allowing the variance of the error term to systematically relate to three separate dimensions (time dependent, asymmetric, and price tier). Upon finding that variance is asymmetric (different for above and below the mean estimated appreciation rates), we test the premise that properties in different price tiers may drive this apparent asymmetry. We find that they do, and we can avoid this problem by specifying variance parameters dependent on price tier. Each observed repeat transaction is weighted depending on the estimated variance to increase the precision of the estimation using generalized least squares (GLS). This process imposes fewer assumptions on the variance during estimation. Finally, we use the estimates to predict the probability that households owe more on their home than it is worth (negative equity).

The WRS method²

The WRS method was first proposed by Bailey et al. (1963) and implemented empirically in the 1980s by Case and Shiller (1987, 1989). Property specific home values are measured only when two consecutive transactions are observed. The WRS uses this information to estimate the average appreciation rate for the location.

The time between consecutive repeat transactions for individual properties can vary dramatically from less than 4 months to over 10 years. To address these time differences we use the following econometric procedures. First, we use time dummies to generate average value changes for cohorts of transactions. We use the errors (ε) from this regression to examine for evidence that the variance (ε^2) is related to the time between transactions. We assume that this variance is a function of time between transactions ($t - s$) and time between transactions squared $(t - s)^2$, where t is the date of the second transaction and s is the date of the first transaction. Finally, we use the fitted residuals from the second regression to create weights in a GLS regression. We use a GLS regression because the variance of the errors cannot be assumed to be constant. With the previous step we have information about how that variance may change. We use this information in the last step as we construct the index. The model is laid out in detail below.

Model design

Following the approach utilized by Case and Shiller (1987) and later modified by Abraham and Schauman (1991). It is assumed that the natural logarithm of price, P_{it} , of an individual house i at time t , can be expressed in terms of a market price index β_t , a Gaussian random walk H_{it} , and white noise N_{it} , such that

$$\ln(P_{it}) = \beta_t + H_{it} + N_{it}. \quad (1)$$

This specification allows us to express the total percentage change in price for house i which transacts in time periods s and t as:

$$\Delta V_i = \ln(P_{it}) - \ln(P_{is}). \quad (2)$$

Substituting (1) into (2) provides:

$$\Delta V_i = \beta_t - \beta_s + H_{it} - H_{is} + N_{it} - N_{is}. \quad (3)$$

We observe $\ln(P_{it}) - \ln(P_{is})$ and we can calculate ΔV_i from (2). The index that we wish to estimate reflects $\beta_t - \beta_s$ but we do not observe either β_t or β_s directly. We can think of $(\beta_t - \beta_s)$ as the average appreciation rate over the time period $(t - s)$. Individual properties may be distributed about these averages, but before we can obtain these average values we must consider the other components of equation (3). We do not observe $H_{it}, H_{is}, N_{it}, N_{is}$. To estimate $\beta_t - \beta_s$, we make the following assumptions:

$$E[H_{it} - H_{is}] = 0, \quad (4)$$

$$E[(H_{it} - H_{is})^2] = A(t - s) + B(t - s)^2, \quad (5)$$

$$E[N_{it}] = 0, \quad (6)$$

$$E[H_{it}N_{js}] = 0, \quad (7)$$

$$E[N_{it}^2] = C = \frac{1}{2}\sigma_N^2, \quad (8)$$

for all i and j , and $t > s$ and A and B are coefficients to be estimated. The difference in the market price index, $\beta_t - \beta_s$, represents the expected rate of appreciation in individual property values in a given market. The Gaussian random walk H_{it} describes how variation in individual house price appreciation rates around the rate of change in the market index causes house prices to vary over time.

The white noise term N_{it} represents cross-sectional dispersion in housing values arising from purely idiosyncratic differences in how individual properties are valued at any given point in time. N_{it} is assumed to be uncorrelated over time and across properties.

In a sample of repeat sales or mortgage transactions, the difference in the natural logarithm of the price of house i can be expressed more generally by:

$$\Delta V_i = \sum_{\tau=0}^T \ln(P_{i\tau}) D_{i\tau}, \quad (9)$$

where $D_{i\tau}$ is a dummy variable that equals 1 if the price of house i was observed for a second time at time τ , -1 if the price of house i was observed for the first time at time τ , and zero otherwise.³ Using equation (1) to substitute for $\ln(P_{i\tau})$ yields:

$$\Delta V_i = \sum_{\tau=0}^T (\beta_\tau + H_{i\tau} + N_{i\tau}) D_{i\tau}. \quad (10)$$

Alternatively, this expression can be written as:

$$\Delta V_i = \sum_{\tau=0}^T \beta_\tau D_{i\tau} + \varepsilon_i \quad \text{where} \quad \varepsilon_i = \sum_{\tau=0}^T (H_{i\tau} + N_{i\tau}) D_{i\tau}. \quad (11)$$

The parameters $\beta_\tau, \tau = 0, 1, 2, \dots, T$ for the market index can be estimated by ordinary least squares (OLS) regression.⁴ An OLS regression of equation (11) on a sample of repeat transactions provides an initial estimate of each β . OLS however is not an efficient estimator because we cannot assume that the variance of the error term ($\sigma_{\varepsilon_i}^2$) is constant. The squared deviations of observed house prices from the market index are given by:

$$\varepsilon_i^2 = [\Delta V_i - \sum_{\tau=0}^T \beta_\tau D_{i\tau}]^2. \quad (12)$$

We assume that the squared deviations of observed house price changes around β_τ will provide us with an estimate for $\sigma_{\varepsilon_i}^2$. Using equations (5) and (8) it can be shown that the expectation of this expression is given by:

$$E[\varepsilon_i^2] = \hat{A}(t-s)_i + \hat{B}(t-s)_i^2 + \hat{C} \cong \sigma_{\varepsilon_i}^2. \quad (13)$$

The estimated variance of the error term in (13) will change for each combination of s and t .

The expected values of the squared deviations, $E[\varepsilon_i^2]$, are used to derive the expected standard error, $E(se_i)$, which is defined as the square root of $E[\varepsilon_i^2]$. The expected errors are then used as the weights needed to obtain GLS estimates of the β_t parameters in the following regression:

$$\frac{\Delta V_i}{E(se_i)} = \sum_{\tau=0}^T \beta_{\tau} \frac{D_{i\tau}}{E(se_i)} + \frac{\varepsilon_i}{E(se_i)}. \quad (14)$$

Equation (14) is estimated to derive WRS HPI. Index numbers for periods $\tau = 1, 2, 3, \dots, T$ are given by:

$$I_{\tau} = 100 \cdot e^{\hat{\beta}_{\tau}}, \quad (15)$$

where $\hat{\beta}_{\tau}$ are the GLS parameter estimates.⁵

New variance and standard error estimates

We focus on the portion of the regression which models the variance of the error term as a function of time and time squared. The original method proposed by Bailey et al. (1963) conducted OLS estimation to produce coefficients on the time dummies in equation (11). The Case and Shiller (1987) method extended this approach to include the estimates of $\sigma_{\varepsilon_i}^2$ and its use in their subsequent GLS estimation. They assumed that $\sigma_{\varepsilon_i}^2$ was simply a linear function of time between transactions. Abraham and Schauman (1991) proposed the use of a quadratic (inclusion of a time-squared term), to take into account the fact that there is an upper bound to the amount of variance of properties about the index or mean. This method was tested extensively against the conventional linear form and is used by both publicly available repeat sales HPI, the Office of Federal Housing and Enterprise house price index (OFHEO HPI) and Freddie Mac conforming mortgage home price index (CMHPI). Consistent with these indices and Abraham and Schauman (1991), the quadratic approach is also used in this paper. While OFHEO and this paper utilizes a geometric weighting scheme, it is also worthwhile to note that Geotzmann (1992) proposed the use of arithmetic weighting in repeat sales indices.

First, we hypothesize that the time dependant nature of the variance may not be the same for properties that disperse above versus below the mean estimated appreciation rate. The mean appreciation rate is the one estimated by the initial OLS regression results (the Bailey, Muth, and Nourse estimates). Intuition suggests that the variance is smaller in magnitude on the

negative tail (the depreciation side) than the positive tail (the appreciation side) of the distribution. There are a number of reasons this could be the case. First, this is consistent with the lognormal distribution observed in house prices and assumed in this procedure. Individual properties may diffuse far above the mean, but obviously cannot decline below zero. In fact, the distribution of values in any given quarter obtained from our database is consistent with wide right tails. Second, the index does not control for home renovations. Therefore, some individual property appreciation rates on the positive tail may look “artificially high” as they are contaminated by noise induced by home improvements. In other words, quality is not being held constant for a group of properties. Another possible reason is that homes appreciating at much less than average rates are less likely to be observed as the owners are less likely to sell under this scenario. Genesove and Mayer (2001) document this finding using a model of loss aversion.

We test for evidence of this phenomenon by applying equation (13) separately for properties that appreciate above and below the mean, as identified by the first stage estimate. This procedure provides two sets of estimates of the parameters A, B, and C, which are used to calculate the expected variance ($E[\varepsilon_i^2]$).

Second, we consider whether the variance is constant across homes in different price categories. A larger variance for lower priced homes is consistent with the hypothesis that lower priced properties are more likely to experience substantial home improvements. They may then be more likely to show large positive deviations from the typical appreciation rate. Another possible reason lower priced properties may experience larger variations in appreciation is that they were undervalued at the time of the first transaction. The reverse could be true for properties with higher house prices. That is, they may experience fewer than the average number of home improvements and they may come into the sample relatively overvalued. Both phenomena would cause some of the higher priced properties to appreciate at a much lower rate, leading to larger variations in the negative tails of the distribution. Therefore, we next include dummies for the highest and lowest 10 percentile tiers interacted with time and time squared. The top and bottom 10 percentiles are identified by the price of the property at the first transaction and are relative to all other observed transactions in the time period. We can then calculate the variance process for properties that are in the highest and lowest tiers separately from the rest of the properties while simultaneously considering the asymmetric dimension added earlier.

We estimate both the asymmetric and the asymmetric price tiered approach for 10 different states (California, Florida, Georgia, Illinois, North Carolina, Nevada, New York, Kansas,

Texas, and Washington states).

Results of state variance estimation

Equation (13) is estimated separately for the group of properties in which the $\varepsilon_i > 0$ and $\varepsilon_i < 0$. Therefore, we obtain separate estimates of A, B, and C for those properties that appreciate above versus below the estimated mean, and thus compute separate volatility or variance estimates for each of these groups. Table 1 displays the difference between the estimated standard error, $E(se_i)$, which is defined as the square root of the estimated variance for the above and below mean appreciation rate properties for each state after 1 year, 5 years, 10 years, and 15 years. If the newly introduced asymmetric dimension is not relevant then the differences should be zero. F-tests were utilized to test the difference in the A, B, and C parameters for the above and below mean groups. The parameters are always statistically different at above a 99 percent confidence level.

The standard error of house price appreciation rates is usually higher for properties that appreciate faster than the mean rate.⁶ Florida, Georgia, and North Carolina consistently exhibit the largest difference between the above and below mean cohorts. For instance, after 5 years since the last transaction the difference in the estimated standard error is 2.49, 2.68, and 2.98 percent, after 10 years the difference is 5.54, 7.68, and 6.62 percent, and 15 years the difference is 5.57, 10.50, and 7.65 percent. Texas also has a relatively high difference (4.59 percent after 15 years). The average difference between above and below mean variance across states after 5 years was 3.53 percent, after 10 years was 3.84 percent, and after 15 years was 4.06 percent.

This phenomenon is illustrated graphically in Figure 1. The growth of the standard error over time is depicted for each state using the asymmetric and symmetric estimation results. The expected standard error is depicted on the y -axis and time between transactions on the x -axis. Expected standard error above the mean is depicted as a positive number. To depict the expected standard error below the mean a negative number is presented in the figures. This is done to aid the visual presentation and is not meant to indicate that a negative expected standard error was estimated. Because at time period 0 the actual house price is known $E(se_i) = 0$. As time passes $E(se_i)$ is increasingly greater than 0 both above and below the mean. For the symmetric estimates the expected standard error is the same above and below the mean in each time period. The asymmetric approach allows the estimated expected standard error above and below the index to be different.

From Figure 1 it is easy to see that Georgia, for example, exhibits large differences

between above average and below average appreciating properties by the 15th year but Washington has the largest difference after 1 year. Visually the figures reinforce the consistency of the results.

The data allows us to set up a test to determine whether differences in price tier may be responsible for the asymmetry observed in the first round of tests. For all states, the asymmetry appears to be induced by the appreciation behavior of properties in the lowest and highest 10 percentile tiers. As mentioned in the previous section, we introduce a new specification for equation (13), which keeps the same original variables, but also includes interactions for time and time squared with dummy variables for the highest and lowest 10 percentile tiers. This model is estimated separately for properties that appreciate above and below the mean. *F*-tests indicate that the interaction dummy parameters are statistically different than the A and B parameters for the middle 10-90 percentile tier, both for below and above mean groups at higher than a 99 percent confidence level.

For all states, the expected standard error is virtually symmetric for properties whose levels fall between the 10 and 90 percentile tiers (see Figure 2). A few states still exhibit a small degree of asymmetry after controlling for the top and bottom tiers. The figures consistently show that among properties that appreciate below average, the top 10 percentile experiences the largest negative or downward variance. In contrast, among properties that appreciate above average the bottom 10-percentile estimated standard error is still the largest.⁷

Implications of asymmetry

This section explores two implications of the proposed estimation technique. First, different specifications of the expected time dependent nature of the variance are likely to lead to different WRS HPI estimates. Second, different specifications are also likely to affect estimates of the equity that borrowers have in a house. Specifically, this section will address how the different specifications affect the probability of the borrower having negative equity. See Foster and Van Order (1984) for an example of using the A, B, and C parameter estimates to calculate the probability of negative equity (pneq).

Index comparisons

Figure 3, provides a visual presentation of the WRS price indices using symmetric, asymmetric, and asymmetric with price tiers specifications from the first quarter of 1980 through the first quarter of 2001. Table 2 provides comparative numbers for all 10 states over the same time period. As Figure 3 and Table 2 show, all 10 states experienced a downward revision in the

index when an asymmetric process is introduced. The two states associated with large metropolitan areas (New York and California) experienced some of the smallest revisions on average. The average revision ranges from a 5.54 percent in Georgia using the symmetric approach to a 0.52 in New York using the asymmetric price tier approach. Ninety percent confidence bands were utilized to determine the statistical differences the indexes. The second column of Table 2 indicates that for most states, statistically different results are obtained between the symmetric and asymmetric indexes in most quarters since 1990. Fannie Mae and Freddie Mac volumes are significantly higher in the 1990s than in the 1970s and 1980s. For this reason, the differences are often insignificant in earlier years, resulting in a misleading calculation when earlier years are included. States with a smaller percentage of significantly different quarters since 1990 (such as Kansas, Nevada, and Washington), become statistically different almost 100 percent of the time if the calculation is restricted to data from 1995 forward. This is because these states contain fewer transactions than the other states in the sample.

When price tiers are included the revision of the index is slightly muted for states. This is seen most clearly in Florida and Georgia. In addition, as shown in the figures, the revisions tend to grow with time. Ninety percent confidence bands again indicate that the tiered index is most often different than the standard index in the quarters since 1990. As with the confidence tests for the asymmetric versus symmetric, smaller states have fewer quarters with statistically valid differences. Again, the index becomes statistically different in the most recent years when the sample is large enough.

New York is the only state in which significance tests are much different for the tiered index. While most states with fairly large numbers of transactions indicate the tiered and standard indexes are statistically different most of the time, the two indexes are never statistically different in New York. This is most likely because the conforming loan limit restricts the true “upper tier” from appearing in our sample. The “upper tier” for New York using Fannie Mae and Freddie Mac data most likely belongs in the “average” bucket if the true tiers were known. That is, results would likely differ substantially if the transactions included jumbo loans for this state.

It is also worthwhile to note that for 8 of the 10 states, the asymmetric indexes are never statistically different than the tiered indexes in the last 12 years (not shown on Table 2). New York and Illinois exhibit differences a minority of the time. This would support our notion that the asymmetric diffusion is likely driven by different appreciation patterns among tiers. In fact, results show that for the middle 10-90 percent tiers, there is little asymmetry in diffusion.

The significantly different results emerging from the price tiered diffusion estimation leads

to the question of whether indexes themselves should be delineated by price tier. We created price tiered indexes for a couple of states for which there were enough transactions. These results are discussed with graphical presentations in Appendix 1. This paper, however, focuses primarily on the specification of the variance.

Home owner negative equity

The specification of the expected variance and the resultant expected standard error can also have affects on estimates of the probability of a mortgage going into a negative equity position. For an individual property (i), the pneq can be calculated as follows:

$$\pi_{\tau,t-s} = \Theta \left(\frac{\log(\text{upb}_{t-s}) - \log(P_{\tau})}{E(\text{se}_{t-s})} \right) \quad (16)$$

where $\pi_{\tau,t-s}$ is the probability that the property is worth less than the mortgage and depends on the τ , the current time period, as well as how long it has been since the last transaction ($t - s$), upb_{t-s} is the unpaid balance on the mortgage and depends on how long the borrower has been paying the mortgage, P_{τ} is the value or price of the home, $E(\text{se}_i)$ is the expected value or estimated standard error as defined by the square root of the expected variance as shown in equation (13), and Θ is the cumulative normal density function (see Deng, 1997; Deng et al., 1994).⁸ Assume that the mortgage interest rate is fixed at 8 percent for the life of the loan, the term is fixed at 30 years, the home initial value is 100 dollars, and a 10 dollar down payment was made. In addition, the borrower is assumed to make all payments on time so that the unpaid balance is reduced on schedule through the 30 years. Lastly, the values or price of the home is updated in each period by the WRS HPI for the state of residence.

Table 3 provides the summary comparative information for the hypothetical mortgage as described above that is originated in the first quarter of 1985 for each of the 10 states. The average probability of negative equity is calculated from the date of origination through the end of 2000. The probability of negative equity is always lowest for houses that are in the middle tier (10-90 percent). This is true for states with very high probabilities of negative equity, such as Texas, and states with low probabilities of negative equity, such as California. In addition, houses in the highest tier always have the highest probability of negative equity. The variety of probabilities ranging from 6.1 to 37.7 percent indicate that location and price tier can make a large difference in the equity position of borrowers with identical mortgage products. This difference can be traced back to the growth in the state HPI, the expected variance of error term over time and the price tier of the house at origination.

From the perspective of an investor homeowner equity, or the lack of it, has two related impacts. First, homeowners with negative equity are more likely to default on a mortgage (Deng et al., 1994; Pennington-Cross, 2003). Second, once a loan is defaulted one of the primary determinants of the severity of the loss is how much equity is in the house (Quigley and Van Order, 1995; Crawford and Rosenblatt, 1995). Therefore, as homeowner equity becomes negative both the probability of default increases thus stopping any future income stream and the credit loss also rises. As a result, Table 3 indicates that high price homes, at least within the conventional conforming market, have on average higher probabilities of negative equity than other price tiers and as a result should also have higher default rates and higher loss rates given a default.

Conclusion

The introduction of new dimensions to the variance of the error term while estimating the WRS HPI can lead to large and small changes in the estimated indices. These new dimensions allow asymmetry and price tier effects to interact with the time between transaction. Typically the index revises downward when the estimated expected variance is made more general by allowing it to contain an asymmetric and house price tier component. More detailed examination of the expected variance indicates that the most expensive homes exhibited a larger time dependent component on the negative side (down side, relative to the index) and the least expensive homes exhibited a larger time dependent component on the positive side (upside, relative to the index). The rest of the homes exhibited roughly a symmetric variance. This implies that while the variance seems to be asymmetric overall, this asymmetry is being driven by the least and most expensive homes. These findings also have implications for estimates of the households' equity position. For the 10 states included in this study, the various specifications show middle tier houses have the lowest probability of negative equity and that the most expensive houses have the highest probability of negative equity. In addition, there is substantial variation in the estimated equity positions of borrowers in different locations and price tiers even if they use identical mortgages (10 percent down payment, 8 percent fixed rate, 30-year term, originated in the first quarter of 1985). These differences, which can exceed 30 percentage points, can be traced back to the state's growth in the HPI, the variance of prices around the HPI over time, and the price tier of the house at origination.

While this paper has focused on the role of the time between transactions and the importance of alternative specifications for the variance of the error term, it is also natural to

consider whether the underlying house price process is different for alternative segments of the housing market. These segments could include a more detailed geographic definition of location such as the metropolitan area or even the neighborhood. In addition, different price segments of the housing market may experience different patterns of house price appreciation within each of these geographical definitions. These issues may have implications on the returns of home ownership for different segments of the population as well as the benefits of home ownership as a mechanism to increase wealth for low-income households. These and many other econometric and public policy issues are beyond the scope of this paper and are potentially fruitful avenues for future research.

Notes

1. The hybrid model was further extended to relax the random walk assumption in both the works of Hill et al. (1997) and Englund et al. (1998), who uses a rich Swedish data set.
2. This paper draws heavily from the Technical Appendix to the OFHEO HPI, which was written by Charles Calhoun in 1996. The approach used here is also laid out in the documentation of Freddie Mac's CMHPI, which was published by Stephens et al. (1995).
3. Note that the time period τ , which indicates the time period the index is estimated for, is different from t , which was used previously to denote the time period of the second transaction.
4. It is necessary to restrict one of the market index parameters to avoid perfect collinearity among the explanatory variables. It is convenient to use $\beta_r = 0$, where r is the base period of the reported index.
5. If the restriction $\beta_1 = 0$ is imposed in estimation, then $I_1 = 100$.
6. The quadratic usually results in turning points after a number of years, which cause this result to reverse as in the case of Illinois and New York. However, few properties in the sample experience 15 years between transactions.
7. We also tested the functional form utilized for the variance estimates using a simple nonparametric approach. We generated frequencies by grouping properties by the number of years between transactions and calculated the average variance. Our tests confirmed that the functional form chosen is appropriate. There was very little difference between the estimated parametrical and nonparametric results. The smaller states exhibit larger differences, but that is due to noise resulting from small sample sizes.
8. The expected variance is time varying as defined by the parameter estimates of A, B, C

and the time between transactions ($t - s$).

9. We limit the data to 1987 forward, because the previous years have too few transactions to generate a reasonable index for the top and bottom tiers.

10. However, the MSAs contain too few observations to estimate valid indices by tier.

*The views expressed in this research are those of the authors and do not represent policies or positions of the Office of Federal Housing Enterprise Oversight (OFHEO) or other officers, agencies, or instrumentalities of the United States Government. While this paper discusses some implications of a house price index (HPI) calculated differently than OFHEO's HPI, the results are part of a broader examination of related issues. OFHEO is not considering making any changes to the HPI at this time.

References

- Abraham, J. M., and W. S. Schauman. (1991). "New Evidence on Home Prices from Freddie Mac Repeat Sales," *AREUEA Journal* 19(3), 333-352.
- Bailey, M. J., R. F. Muth, and H. O. Nourse. (1963). "A Regression Method for Real Estate Price Index Construction," *Journal of the American Statistical Association* 58, 933-942.
- Calhoun, C. A. (1991). "Estimating Changes in Housing Values from Repeat Transactions," presented at the Western Economic Association International Meetings, Seattle, Washington, July 1991.
- Calhoun, C. A. (1996). "OFHEO House Price Indexes: HPI Technical Description," Washington, D.C.: Office of Federal Housing Enterprise Oversight, March 1996.
- Calhoun, C. A., S. H. Dreiman, and M. Vandergoot. (1999). "Appraisals, Repeat Mortgage Transactions, and House Price Indices," Presented at the Western Economic Association Meetings, San Diego, CA, July 1999.
- Case, B., and E. Szymanoski. (1995). "Precision in House Price Indices: Findings of a Comparative Study of House Price Index Methods," *Journal of Housing Research* 6(3), 483-496.
- Case, B., and J. M. Quigley. (1991). "The Dynamics of Real Estate Prices," *The Review of Economics and Statistics* 73, 50-58.
- Case, K. E., and R. J. Shiller. (1987). "Prices of Single Family Real Estate," *New England Economic Review* 45-56.
- Case, K. E., and R. J. Shiller. (1989). "The Efficiency of the Market for Single-Family Homes," *American Economic Review* 79(1), 125-137.

- Clapp, J., and C. Giaccotto. (1998). "Revisions in Repeat Sales Price Indices: Here Today, Gone Tomorrow," *Real Estate Economics* 27(1), 79-104.
- Crawford, G. W., and E. Rosenblatt. (1995). "Efficient Mortgage Default Option Exercise: Evidence from Loss Severity," *The Journal of Real Estate Research* 10(5), 543-555.
- Deng, Y. (1977). "Mortgage Termination: An Empirical Hazard Model with Stochastic Term Structure," *Journal of Real Estate Finance and Economics* 14(3), 309-329.
- Deng, Y., J. Quigley, and R. Van Order. (1994). "Household Income, Equity, and Mortgage Default Risks," Working Paper, University of California-Berkeley.
- Englund, P. (1998). "Improved Price Indexes for Real Estate: Measuring the Course of Swedish Housing Prices," *The Journal of Urban Economics* 44, 171-196.
- Foster, C., and R. Van Order. (1984). "An Option-Based Model of Mortgage Default," *Housing Finance Review* 3(4), 351-377.
- Genesove, D., and C. Mayer. (2001). "Loss Aversion and Seller Behavior: Evidence from the Housing Market," *Quarterly Journal of Economics* 116(4), 1233-1260.
- Geotzmann, W. (1992). "The Accuracy of Real Estate Indices: Repeat Sales Estimators," *Journal of Real Estate Finance and Economics* 5, 5-53.
- Hill, R. C., J. R. Knight, and C. F. Sirmans. (1997). "Estimating Capital Asset Prices," *Review of Economics and Statistics* 79, 226-233.
- Pennington-Cross, A. (2003). "Credit History and the Performance of Prime and Nonprime Mortgages," Presented at the 2003 Allied Social Science Association Conference in the American Real Estate and Urban Economics Association sessions.
- Quigley, J. M., and R. Van Order. (1995). "Explicit Test of Contingent Claims Models of Mortgage Default," *Journal of Real Estate Finance and Economics* 11, 99-117.
- Stephens, W., Y. Li, V. Lekkas, J. Abraham, C. Calhoun, and T. Kimner. (1995). "Conventional Mortgage Home Price Index," *Journal of Housing Research* 6(3), 389-418.

Appendix
Table 1
Absolute difference in $E(se_i)$

State	Years Since Last Transaction (%)			
	1	5	10	15
California	1.26	1.45	0.67	0.71
Florida	2.49	4.68	5.54	5.57
Georgia	2.68	4.93	7.68	10.50
Illinois	3.11	4.52	2.93	0.03
North Carolina	2.98	5.06	6.62	7.65
Nevada	3.22	2.88	3.02	3.09
New York	0.98	3.66	2.45	1.42
Kansas	3.64	2.70	3.28	4.39
Texas	2.08	3.48	4.59	5.39
Washington	4.74	1.97	1.58	1.88
Average	2.72	3.53	3.84	4.06

Notes. The absolute difference in $E(se_i)$ is defined as the absolute value of the percent difference between the expected standard error for properties that appreciated faster than the mean relative to the expected standard error for properties that appreciated slower than the mean. For instance, after one year since the last transaction in California there is a 1.26 percent difference in the expected standard error (above mean versus below mean). This difference decreases to 0.78 percent by 15 years.

Table 2
Average quarterly differences in WRS price indexes*

State	Symmetric Index Less Asymmetric Index (%)	Percent of Quarters Asymmetric Statistically Different than Symmetric† (%)	Symmetric Index Less Asymmetric with Price Tiers Index (%)	Percent of Quarters Tiered Statistically Different than Symmetric† (%)
California	0.86	100	0.64	98
Florida	4.08	100	3.50	100
Georgia	5.54	100	4.70	100
Illinois	4.10	100	2.22	88
Kansas	1.99	66	0.64	24
North Carolina	5.40	100	4.97	100
Nevada	2.52	44	2.47	54
New York	1.81	98	0.52	0
Texas	3.36	100	2.67	100
Washington	1.41	30	0.84	23

*The average difference is calculated as the average percent difference over the time period 1980 through the first quarter of 2001.

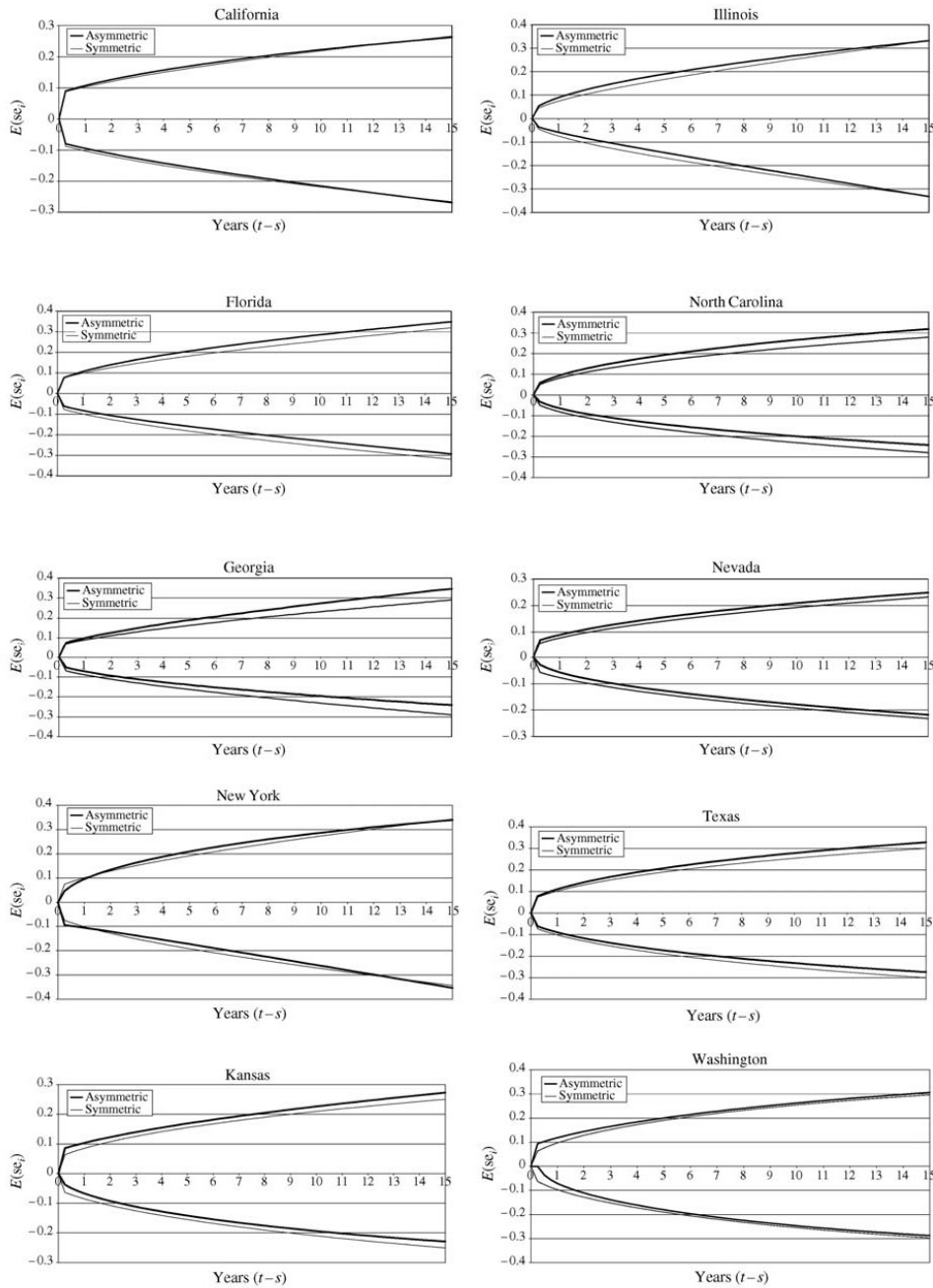
†Only quarters since 1990 were used for these calculations. The earlier years are more likely to exhibit insignificant differences in the indexes due to fewer transactions contributing to the index.

Table 3
Average probability of negative equity

State	Price Tier (%)		
	Low 10%	10 to 90%	High 90%
California	11.0	6.1	13.4
Florida	20.7	18.5	26.8
Georgia	14.5	12.2	22.6
Illinois	13.7	8.5	16.2
Kansas	10.5	9.8	19.5
North Carolina	12.2	8.6	21.0
Nevada	10.7	8.9	16.9
New York	21.0	15.2	21.5
Texas	35.8	32.9	37.7
Washington	10.5	9.3	18.5

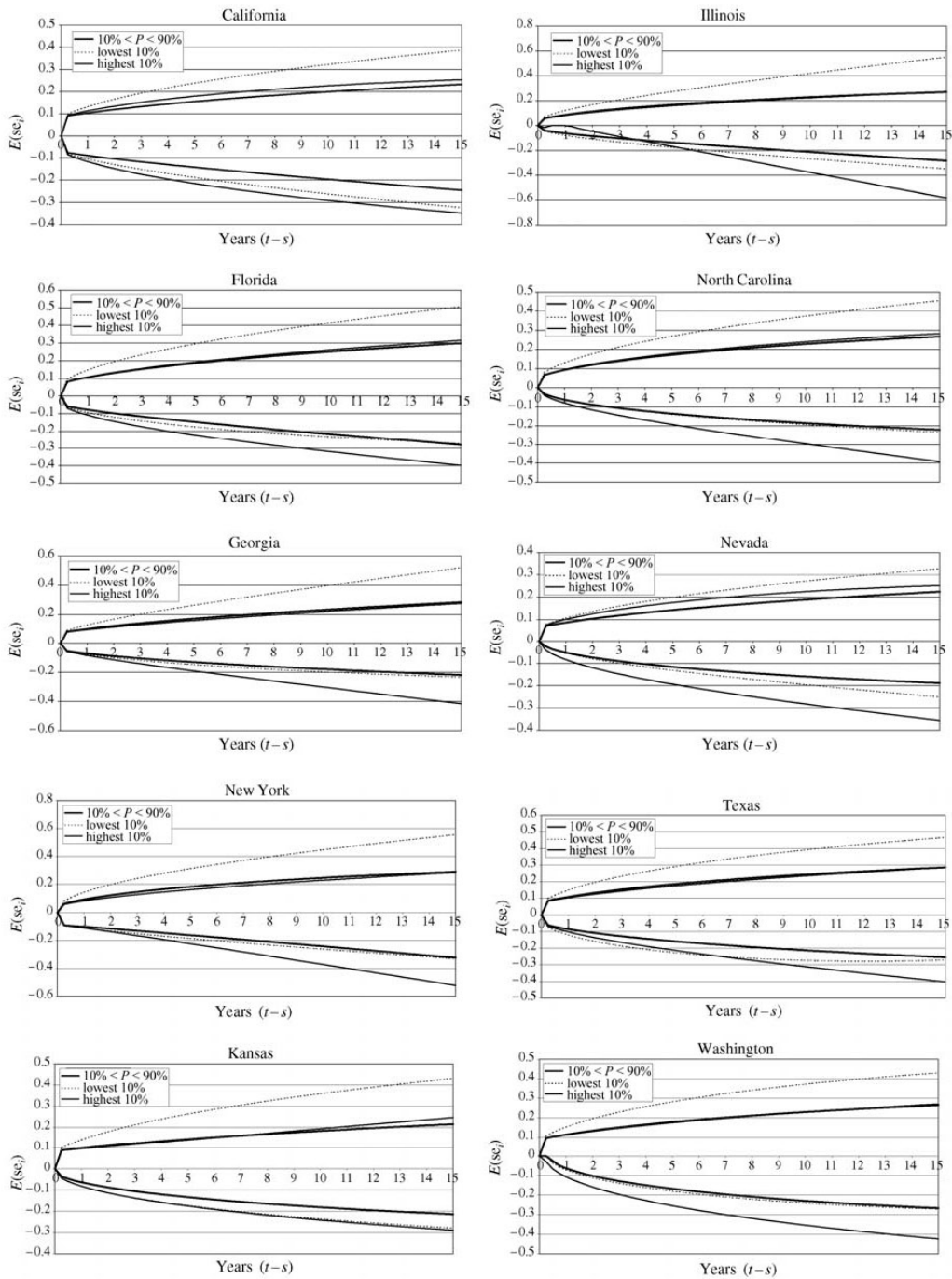
Notes. These estimates use the revised expected variance parameters and the revised indices to calculate the average probability of negative equity for a hypothetical mortgage (originated in the first quarter of 1985, 8 percent fixed interest rate, 30 year term, 100 dollar initial value of the home, and a 10 dollar down payment) over the first 15 years of the mortgage. Averages are calculated from quarterly estimates.

Figure 1
Expected standard error, by time between transaction



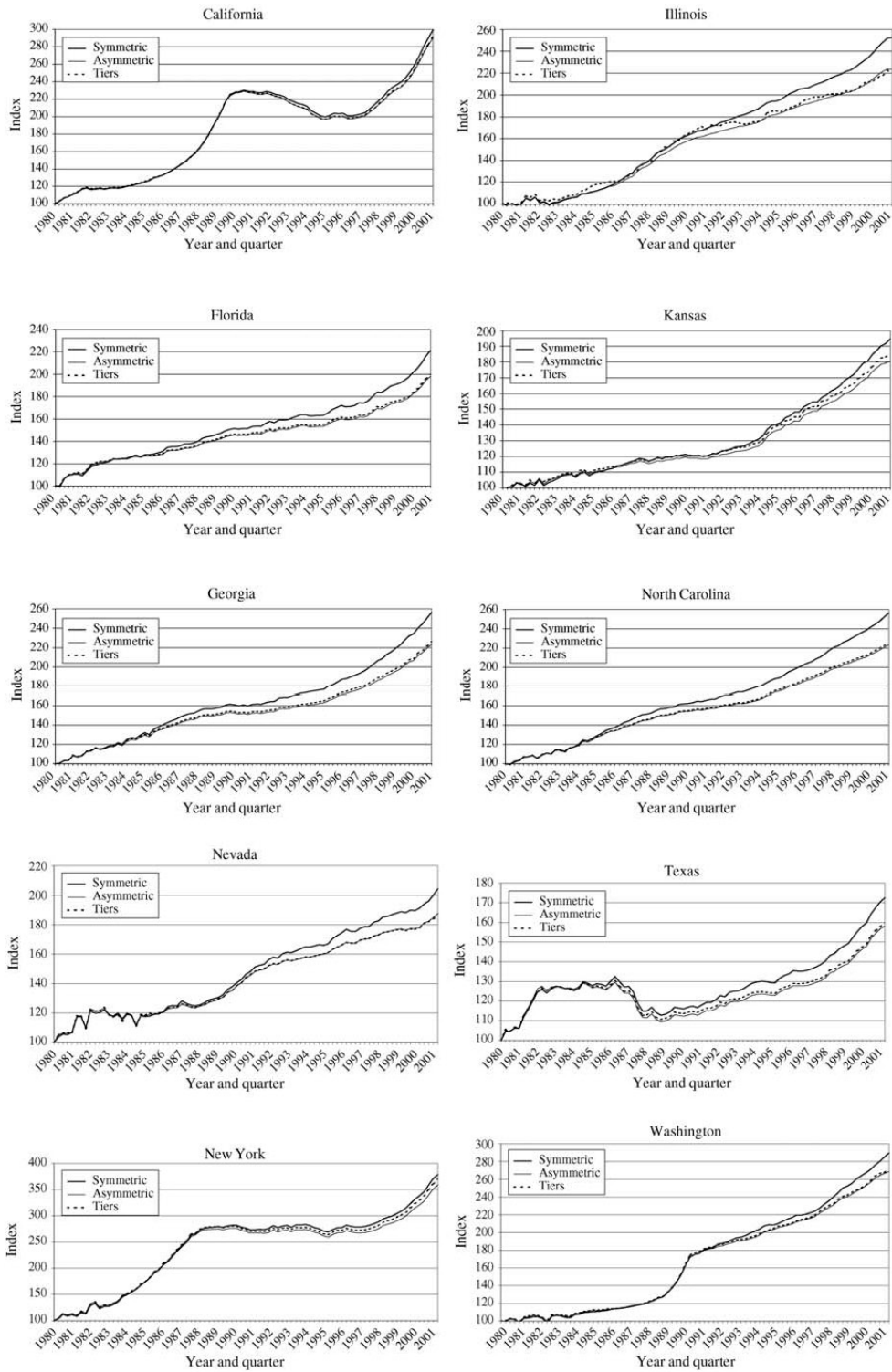
The expected standard error, $E(se_i)$, is depicted on the y-axis and time between transactions on the x-axis. To aid the visual presentation, $E(se_i)$ above the mean is depicted as a positive number and below the mean as a negative number even though it can by definition only be positive. Because at time period 0 the actual house price is known $E(se_i) = 0$. In general, as time passes $E(se_i)$ is increases. For the symmetric estimates $E(se_i)$ is the same above and below the index. The asymmetric approach allows $E(se_i)$ above and below the index to be different.

Figure 2
Expected standard error, by time between transaction and price tier



See the note accompanying Figure 1. In addition, the expected standard error, $E(se_i)$, is derived from different parameter estimates that allows $E(se_i)$ to be different above and below the mean, across time, and across different price tiers. $E(se_i)$ above the mean is depicted as a positive number and $E(se_i)$ below the mean is depicted as a negative number even though it can, by definition, only be positive. This is done to aid the visual presentation. At time period 0 the actual house price is known so that $E(se_i) = 0$.

Figure 3
Repeat sale price index comparison



Appendix 1: Price tiered indexes

In order to better understand the implications of price tiers, we also estimate the index separately by price tier. The price tiers depend on whether the house price is in the top 10 percentile, bottom 10 percentile, or between the 10th and 90th percentiles. A few states had enough transactions to generate a reasonable result, and we focus on those states to shed some light on this question. Specifically, those states are California, Illinois, Florida, New York, and Washington. California, however, has significantly more transactions than any other state.

The results for California are pictured in Figure A-1. We graph the annual average percent change for each year from March 1987 to March 2001.⁹ This smoothes the index so that results are less influenced by quarterly noise. In California during housing booms (late 1980s, late 1990s, and early 2000s) the more expensive properties tend to appreciate more rapidly, and during housing busts, the cheaper properties appreciate more rapidly. However, the absolute differential seems to be greater during housing busts. That is, during busts, the bottom 10 percent experience substantially higher appreciation than the middle and upper tiers. Additionally, declines in the lower end properties appear to lag declines in the middle and upper tiers. This is consistent with conventional wisdom that the more expensive properties in a housing market are generally the first to experience the impact of a downturn.

Appreciation patterns for the three tiers in the state of Washington are quite similar to California, potentially due to proximity. Housing prices in New York and Illinois cycle somewhat differently than the two West Coast states. However, the data also roughly indicate more rapid declines of higher priced properties during busts and more rapid appreciation of these properties during booms. A high degree of volatility in the Florida index makes it difficult to analyze, perhaps due to heterogeneity across metropolitan areas. However, it is notable that the lower end properties appreciate much more rapidly during the most recent period than the higher priced homes. This is not consistent with the results of the other four states.

Using 90 percent confidence bands, we test the statistical differences between the indexes for the states mentioned. Table A-1 indicates that the three indexes are statistically different in most quarters since 1990 for California. Florida and Illinois also indicate the bottom 10 percentile is usually significantly different than the middle or upper percentiles. For New York and Washington none of the three indexes are statistically different.

As illustrated in Table A-1, a disadvantage of producing price tiered indices is the loss of degrees of freedom that comes from limiting samples to 10 percentile buckets. This influences the results of most states to the degree that the data is unreliable. Also, metropolitan area heterogeneity makes it difficult to properly assign rationale to tiered state index results.¹⁰ Diffusion parameters take advantage of a broader sample because they are not estimated separately for each year. Therefore, we can get solid estimates for these parameters even after breaking down the data into tiers. Using this technique of accounting for tiers, we get a more accurate index estimate without incurring the cost in terms of lost degrees of freedom.

Table A-1
Percent of quarters price indexes are statistically different*

State	Bottom/Top (%)	Bottom/Middle (%)	Top/Middle (%)
California	90.24	87.80	97.56
Florida	65.85	90.24	17.07
Illinois	65.85	85.37	4.88
New York	0.00	0.00	0.00
Washington	0.00	0.00	0.00

*Since the first quarter of 1991.

Figure A-1
Annual average percentage change in price index by price tier

