OFHEO WORKING PAPERS

Working Paper 02-2

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

By
Michelle H. Dreiman
and
Anthony Pennington-Cross

* 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 or other officers, agencies, or instrumentalities of the United States Government. While this paper discusses some implications of a house price index calculated differently than OFHEO's house price index, the results are part of a broader examination of related issues. OFHEO is not considering making any changes to the HPI at this time.

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

by

Michelle H. Dreiman
and
Anthony Pennington-Cross

Office of Federal Housing Enterprise Oversight
Office of Policy & Research
1700 G Street, NW
Washington, DC 20552
(202) 414-3800

April 2002

* 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 or other officers, agencies, or instrumentalities of the United States Government.
While this paper discusses some implications of a house price index calculated differently than OFHEO's house price index, the results are part of a broader examination of related issues. OFHEO is not considering making any changes to the HPI at this time.

Summary

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

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.

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

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 of 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 % over the period 2000 to 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 exactly 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 (for example, a small sapling grows into a majestic tree) of some properties and decrease it in others (for example, 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. Also, the index may exhibit a hidden bias because certain properties may have more repeated observations than others (Clapp, 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 non-conforming 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. For a detailed discussion of these potential biases and others associated with weighted repeat sales 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 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 house price index 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). For example, properties that are purchased at a very low prices 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. In addition, the revised indices will impact estimates of the equity growth or decline for homeowners.

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 Weighted Repeat Sales Method[1]

The WRS method was first proposed by Bailey, Muth, and Nourse (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 $(t\text{-}s)^2$, 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 generalized least squared regression. We use a generalized least squares 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

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

[1] The material in this section borrows heavily from Calhoun (1996).

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

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

Substituting (1) into (2) provides:

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

We observe $\ln(P_{it}) - \ln(P_{is})$ and we can calculate ΔV_i from (2). The index that we wish to estimate reflects β_t-β_s but we do not observe either β_t or β_s directly. We can think of (β_t-β_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 β_t-β_s, we make the following assumptions:

$$E[H_{it} - H_{is}] = 0 \tag{4}$$

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

$$E[N_{it}] = 0 \tag{6}$$

$$E[H_{it} N_{js}] = 0, \tag{7}$$

$$E[N_{it}^2] = C = \tfrac{1}{2}\sigma_N^2 \tag{8}$$

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. Note that the Case and Shiller (1987) approach assumed that the variance of the error term grew proportionally with time between transaction due to the Gaussian random walk. Abraham and Schauman (1991) proposed the use of a quadratic under the assumption that the variance could not grow over time without bounds.

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:

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

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.[2] Using equation (1) to substitute for $\ln(P_{i\tau})$ yields:

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

Alternatively this expression can be written as:

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

The parameters β_τ, $\tau = 0,1,2,\ldots,T$ for the market index can be estimated by ordinary least squares (OLS) regression.[3] 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 ($\sigma^2_{\varepsilon_i}$) is constant. The squared deviations of observed house prices from the market index are given by:

$$\varepsilon_i^2 = \left[\Delta V_i - \sum_{\tau=0}^{T} \beta_\tau D_{i\tau} \right]^2 \tag{12}$$

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

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

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 weights needed to obtain GLS estimates of the β_t parameters in the following regression:

$$\frac{\Delta V_i}{\sqrt{E[\varepsilon_i^2]}} = \sum_{\tau=0}^{T} \beta_\tau \frac{D_{i\tau}}{\sqrt{E[\varepsilon_i^2]}} + \frac{\varepsilon_i}{\sqrt{E[\varepsilon_i^2]}} \tag{14}$$

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

$$I_\tau = 100 \cdot e^{\hat{\hat{\beta}}_\tau} \tag{15}$$

where $\hat{\hat{\beta}}_\tau$ are the GLS parameter estimates.[4]

---

[2] 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.

[3] 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.

[4] If the restriction $\beta_1 = 0$ is imposed in estimation, then $I_1 = 100$.

New Variance 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, Muth, and Nourse (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 $\sigma^2_{\varepsilon_i}$ and its use in their subsequent GLS estimation. They assumed that $\sigma^2_{\varepsilon_i}$ was simply a linear function of time between transactions. Abraham and Schauman 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.

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. One of the primary reasons for this may be that 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. 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 variance ($E[\varepsilon_i^2]$) 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 to the variance term is not relevant then the differences should be zero. If there is more variance above the average, the difference should be positive.

The variance of house price appreciation rates is usually higher for properties that appreciate faster than the mean rate.[5] 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 variance is 1.7, 1.6, and 1.7 percent, after 10 years the difference is 2.9, 3.6, and 3.1 percent, and 15 years the difference is 3.6, 6.2, and 4.2 percent. Texas also has a relatively high difference (3.2 percent after 15 years). The average difference between above and below mean variance across states after 5 years was 1.2 percent, after 10 years was 1.8 percent, and after 15 years was 2.1 percent.

This phenomenon is illustrated graphically in Figures 1 through 10. The growth of the variance over time is depicted for each state using the asymmetric and symmetric estimation results. The expected variance of the error term is depicted on the y-axis and time between transactions on the x-axis. Expected variance above the mean is depicted as a positive number. To depict the expected variance 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 variance was estimated. Because at time period 0 the actual house price is known $E[\varepsilon_i^2]=0$. As time passes $E[\varepsilon_i^2]$ is increasingly greater than 0 both above and below the mean. For the

[5] 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.

symmetric estimates the expected variance is the same above and below the mean in each time period. The asymmetric approach allows the estimated expected variance above and below the index to be different

From Figures 1-10 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 one year.

Table 1: Absolute Difference in $E(\varepsilon^2_i)$

	Years Since Last Transaction			
State	*1*	*5*	*10*	*15*
California	0.26%	0.47%	0.29%	0.38%
Florida	0.48%	1.71%	2.86%	3.57%
Georgia	0.47%	1.63%	3.61%	6.19%
Illinois	0.46%	1.51%	1.49%	0.02%
North Carolina	0.47%	1.70%	3.08%	4.29%
Nevada	0.27%	0.38%	0.52%	0.64%
New York	0.20%	1.40%	1.34%	0.99%
Kansas	0.62%	0.84%	1.38%	2.20%
Texas	0.42%	1.32%	2.34%	3.24%
Washington	0.90%	0.75%	0.80%	1.11%
Average	0.46%	1.17%	1.77%	2.06%

The absolute difference in $E(\varepsilon^2_i)$ is defined as the absolute value of the percent difference between the expected variance for properties that appreciated faster than the mean relative to the expected variance for properties that appreciated slower than the mean. For instance, after one year since the last transaction in California there is a 0.26 percent difference in the expected variance (above mean versus below mean). This difference grows to 0.38 percent by 15 years.

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.

For all states, the variance is virtually symmetric for properties whose levels fall between the 10 and 90 percentile tiers (see Figures 11-20). A few states still exhibit a small degree of asymmetry after controlling for the top and bottom tiers. Also notable is that among properties that appreciate below average, the top 10 percentile experiences the largest negative or downward variance. However, the top 10-percentile estimated variance is still generally less than the bottom 10 percentile for most states (with exception of Nevada and Illinois).[6]

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 house price index 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

Figures 21-30 provide 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

[6] 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 size.

time period. As Figures 21-30 and Table 2 show, all ten 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.

When 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.

Table 2: Average Differences in WRS Price Indexes

State	Symmetric Index Less Asymmetric Index	Symmetric Index Less Asymmetric with Price Tiers Index
California	0.86%	0.64%
Florida	4.08%	3.50%
Georgia	5.54%	4.70%
Illinois	4.10%	2.22%
Kansas	1.99%	0.64%
North Carolina	5.40%	4.97%
Nevada	2.52%	2.47%
New York	1.81%	0.52%
Texas	3.36%	2.67%
Washington	1.41%	0.84%

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

Home Owner Negative Equity

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

$$\pi_{\tau,t-s} = \Theta\left(\frac{\log(upb_{t-s}) - \log(P_\tau)}{\sqrt{E[\varepsilon_{t-s}^2]}}\right) \tag{17}$$

where $\pi_{\tau,t\text{-}s}$ 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*), $upb_{t\text{-}s}$ 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, $E[\varepsilon_{t-s}^2]$ is the expected value or estimated variance of the error term as defined in equation (13), and Θ is the cumulative normal density function (see Deng, 1997, and Deng et al., 1994).[7] 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 house price index for the state of residence.

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%

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, thirty year tem, 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.

[7] The expected variance is time varying as defined by the parameter estimates of A, B, C and the time between transactions (t-s).

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 ten 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% to 90%). 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 house price index, the expected variance of error term over time and the price tier of the house at origination.

Conclusion

The introduction of new dimensions to the variance of the error term while estimating the WRS house price index 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 finding also have implications for estimates of the households' equity position. For the ten 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 house price index, the variance of prices around the house price index 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 could have implications on the returns of home ownership for different segments of the population and 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.

References

Abraham, J.M. and Schauman, W.S. (1991). "New Evidence on Home Prices from Freddie Mac Repeat Sales," *AREUEA Journal*. 19(3), 333-352.

Bailey, Martin J., Richard F. Muth, and Hugh O. Nourse (1963). “A Regression Method for Real Estate Price Index Construction,” *Journal of the American Statistical Association* 58, 1963, 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., Dreiman, S.H., and Vandergoot, M. (1999). “Appraisals, Repeat Mortgage Transactions, and House Price Indices,” presented at the Western Economic Association Meetings, San Diego, CA, July 1999.

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*, 1989.

Clapp, J. and C. Giaccotto (1998). "Revisions in Repeat Sales Price Indices: Here Today, Gone Tomorrow,” *Real Estate Economics*, 27 (1) 79-104.

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.

Foster, C. and R. Van Order, “An Option-Based Model of Mortgage Default,” *Housing Finance Review*, 3(4):351-77, October 1984.

Figure 1: Expected Variance, By Time Between Transaction *

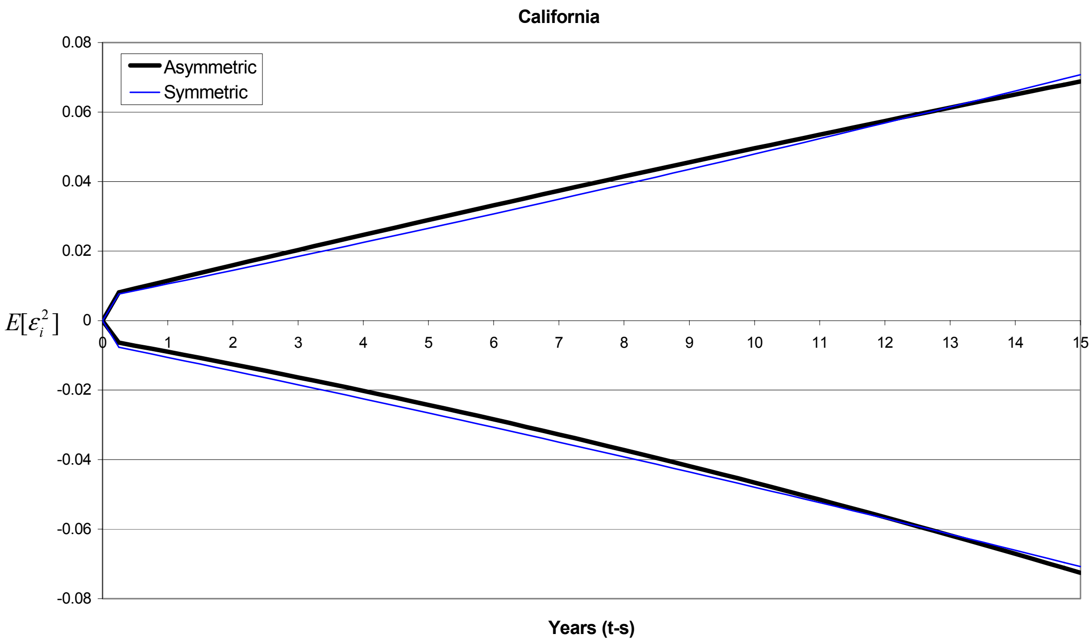
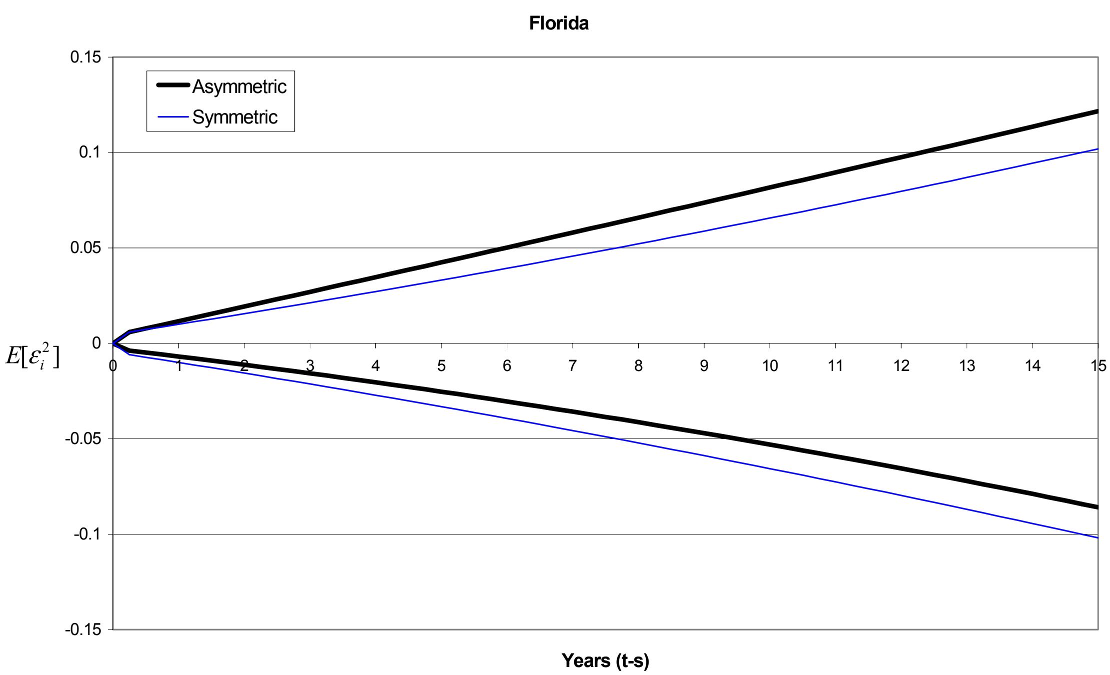


Figure 2: Expected Variance, By Time Between Transaction



* The expected variance of the error term is depicted on the y-axis and time between transactions on the x-axis. Expected variance above the mean is depicted as a positive number. Expected variance 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. Because at time period 0 the actual house price is known, $E(\varepsilon^2)=0$. In general, as time passes $E(\varepsilon^2)$ is increasingly greater than 0 both above and below the mean. For the symmetric estimates the expected variance is the same above and below the mean in each time period. The asymmetric approach allows the estimated expected variance above and below the index to be different.

Figure 3: Expected Variance, By Time Between Transaction

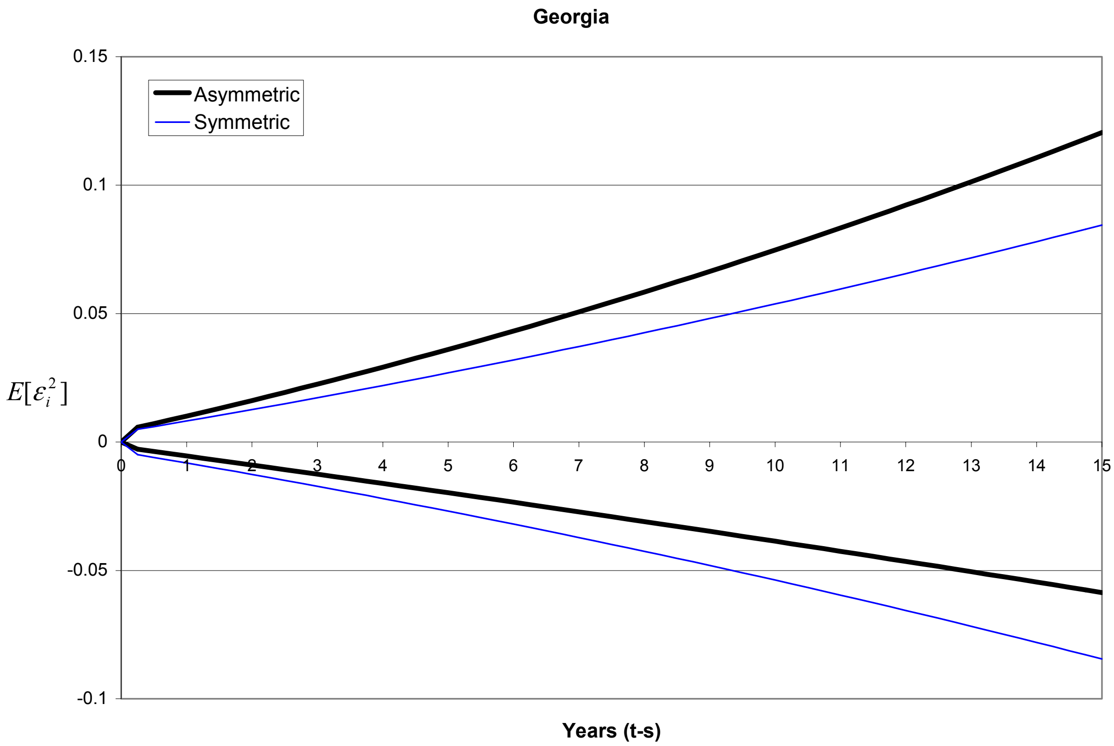


Figure 4: Expected Variance, By Time Between Transaction

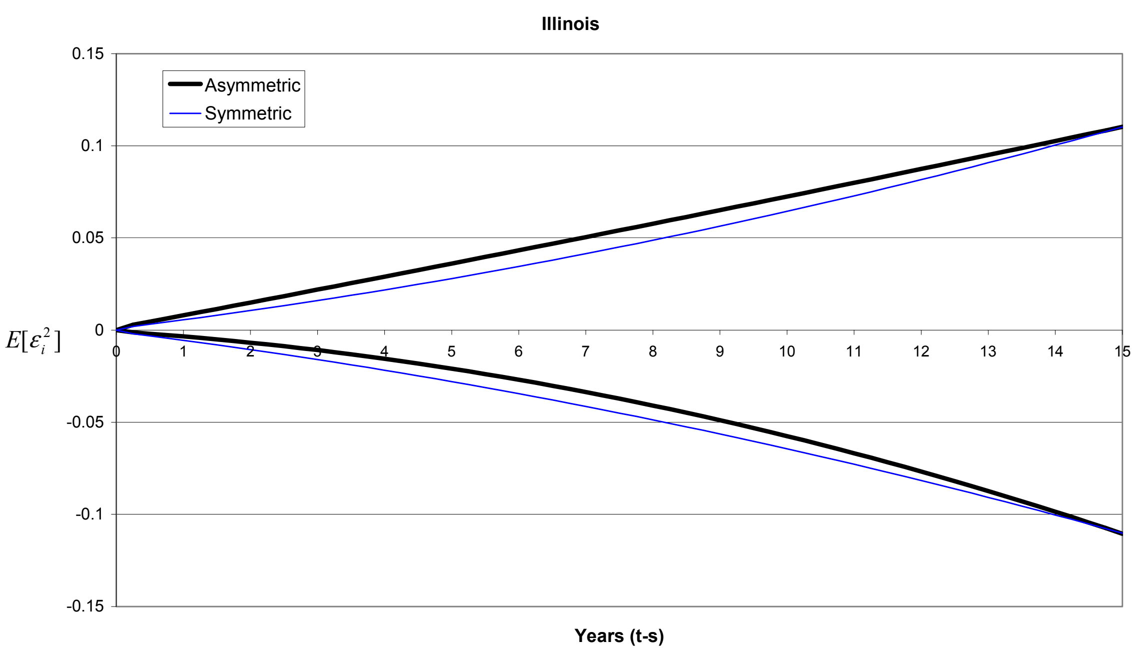


Figure 5: Expected Variance, By Time Between Transaction

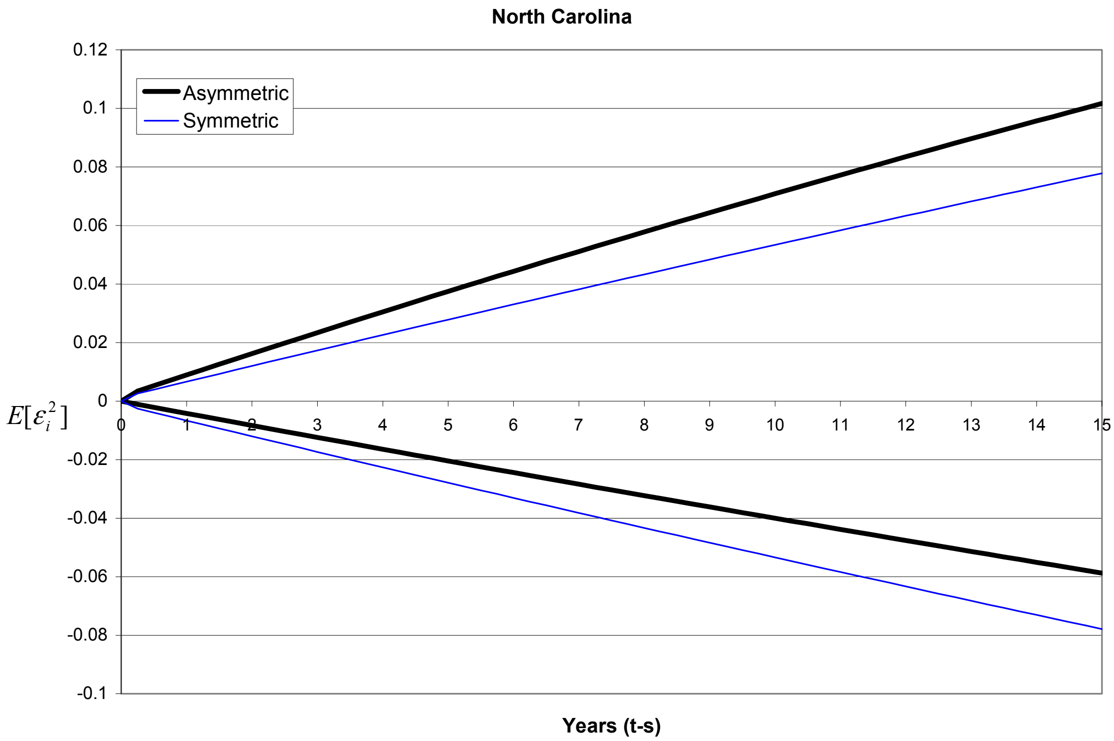


Figure 6: Expected Variance, By Time Between Transaction

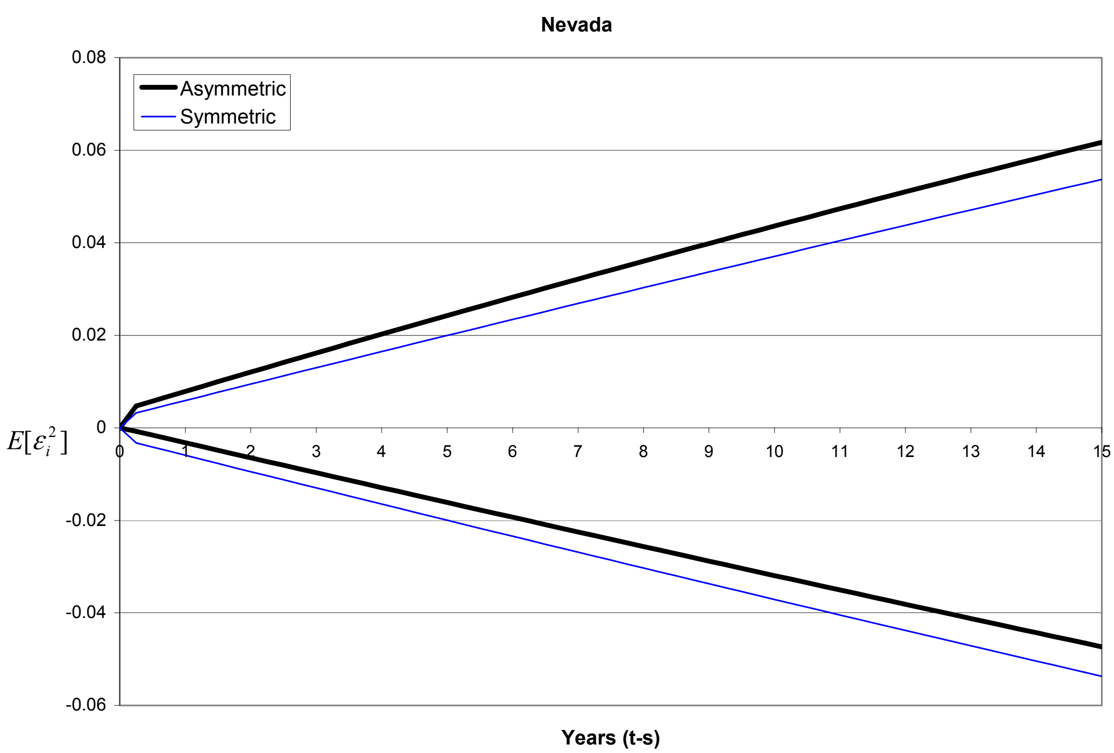


Figure 7: Expected Variance, By Time Between Transaction

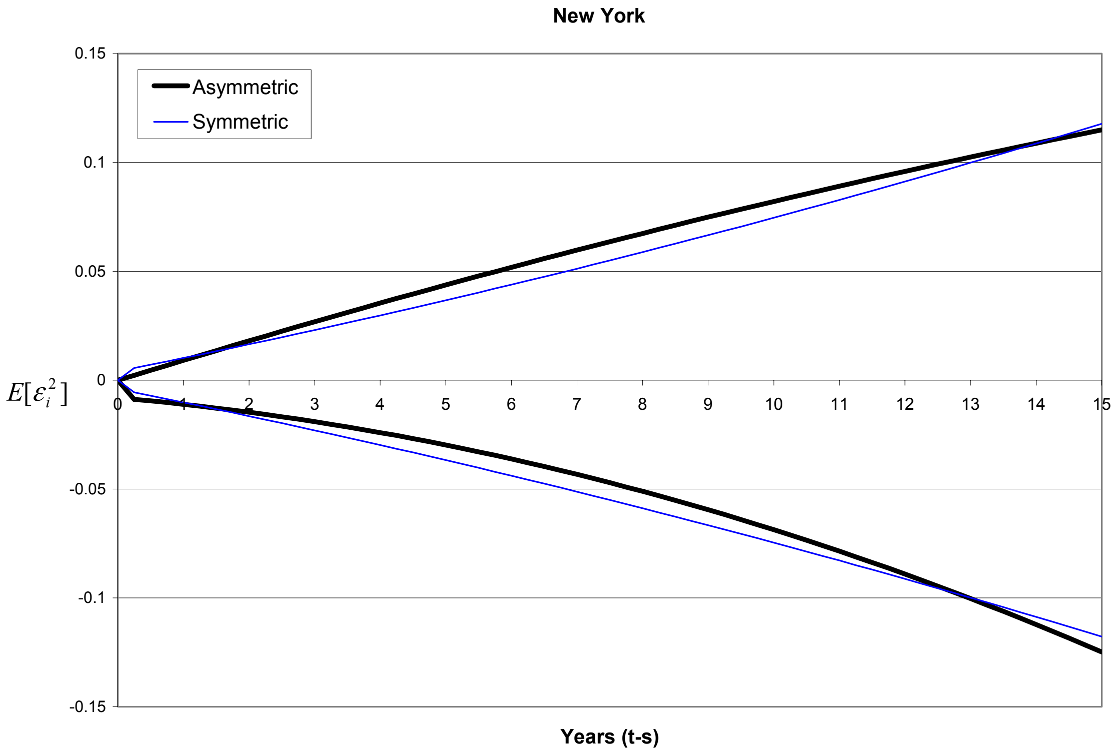


Figure 8: Expected Variance, By Time Between Transaction

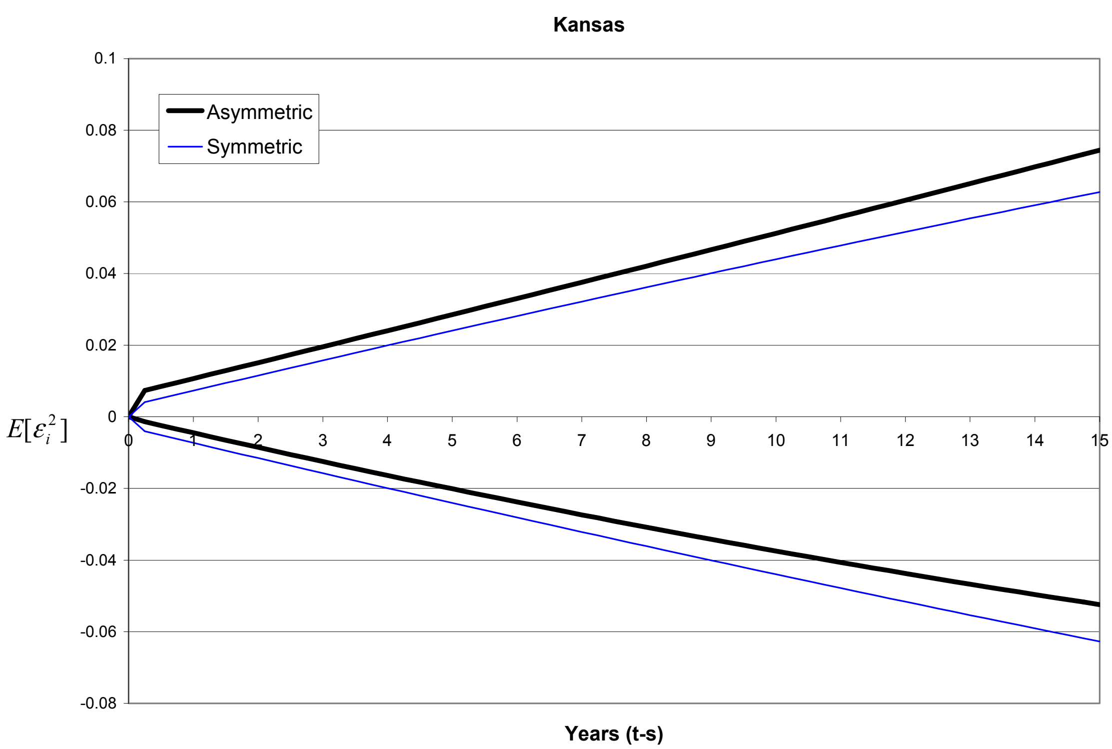


Figure 9: Expected Variance, By Time Between Transaction

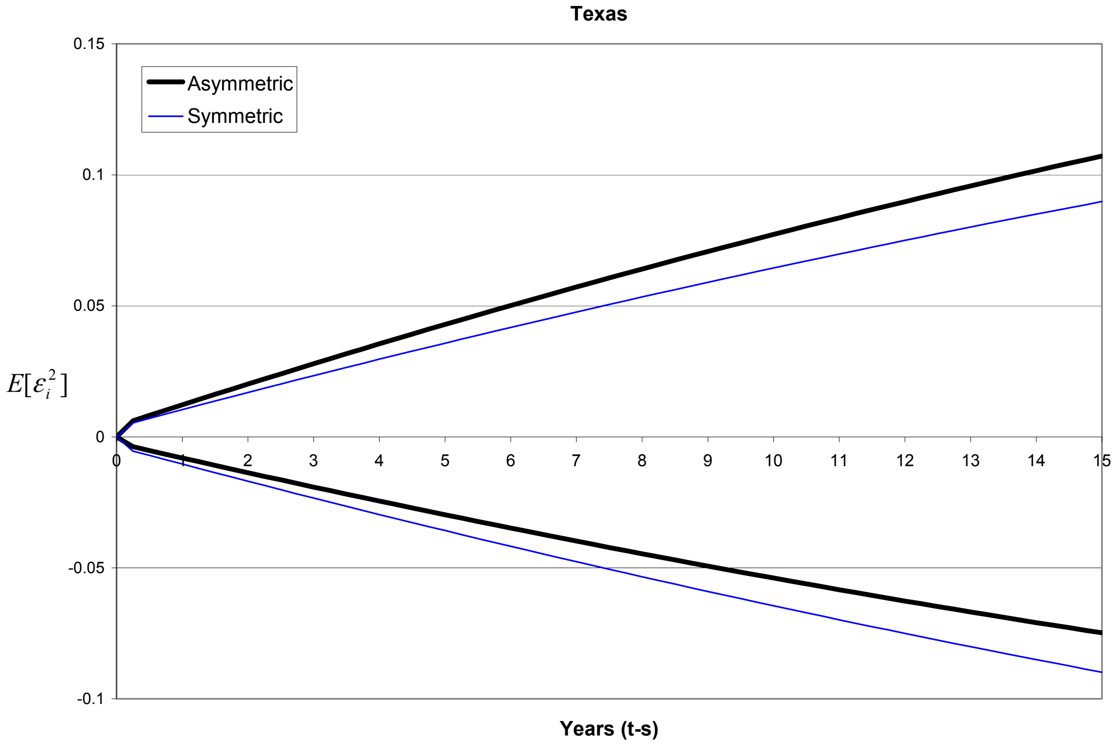


Figure 10: Expected Variance, By Time Between Transaction

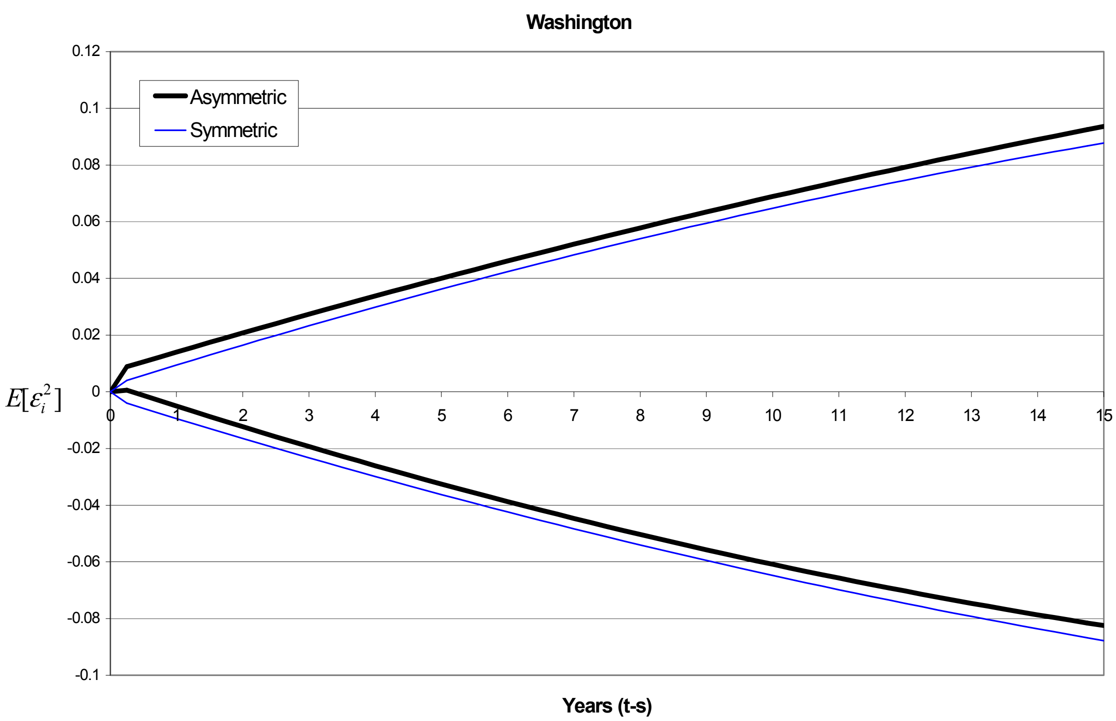


Figure 11: Expected Variance, By Time Between Transaction and Price Tier*

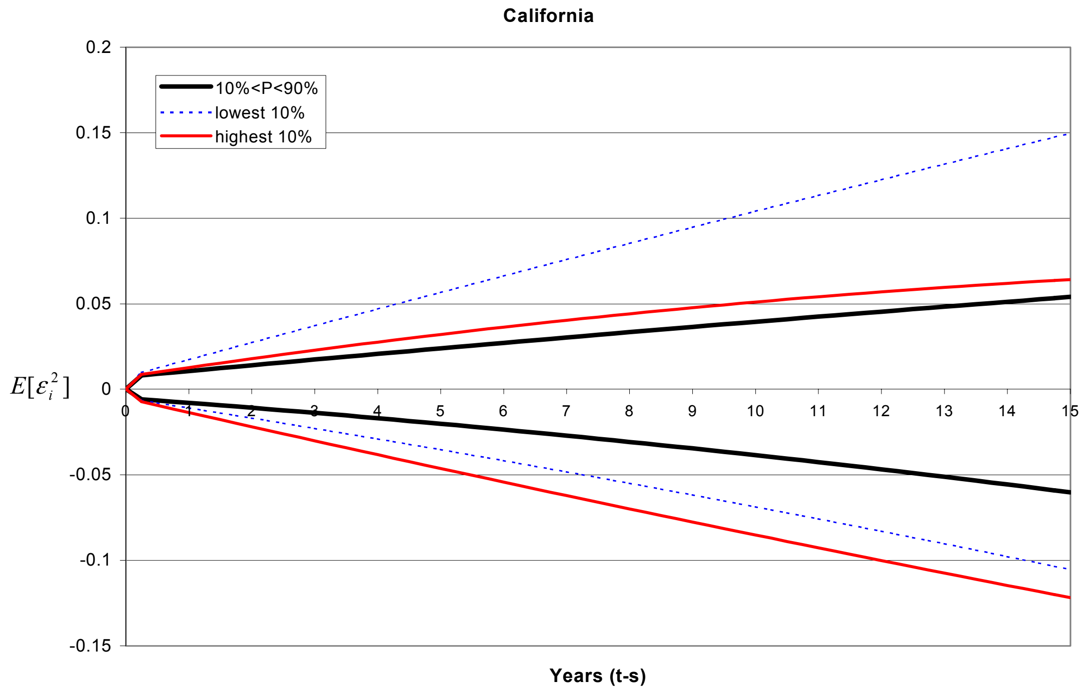
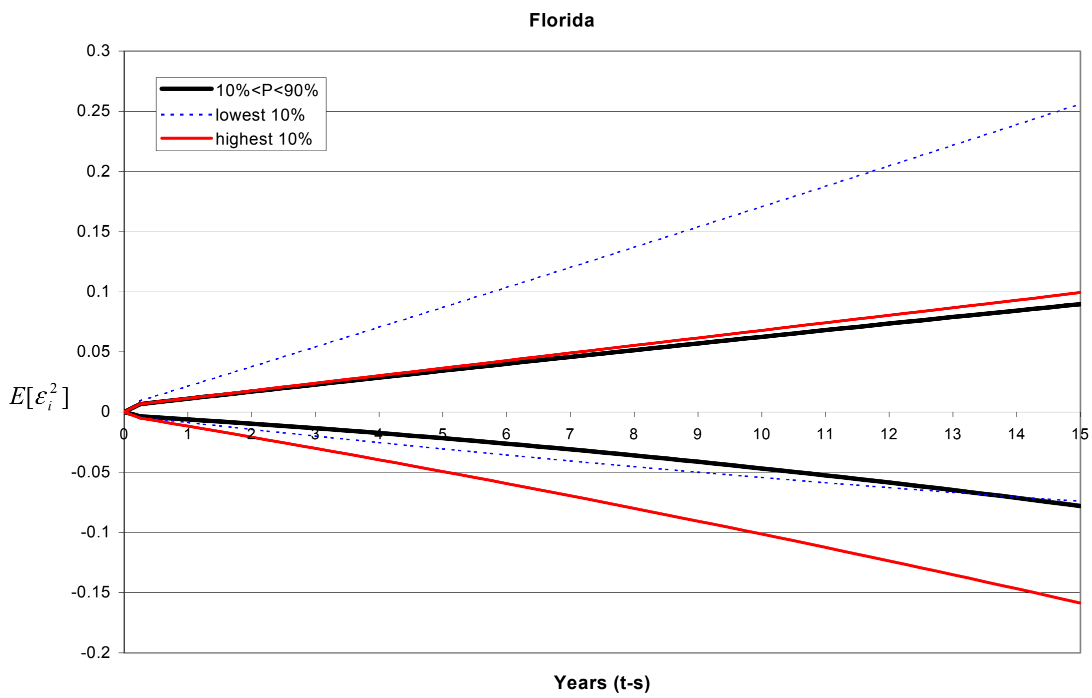


Figure 12: Expected Variance, By Time Between Transaction and Price Tier



* See the note accompanying Figure 1. In addition, the expected variance, $E(\varepsilon^2)$, uses different parameter estimates that allows $E(\varepsilon^2)$ to be different above and below the mean, across time, and across different price tiers. $E(\varepsilon^2)$ above the mean is depicted as a positive number. $E(\varepsilon^2)$ 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 $E(\varepsilon^2)=0$.

Figure 13: Expected Variance, By Time Between Transaction and Price Tier

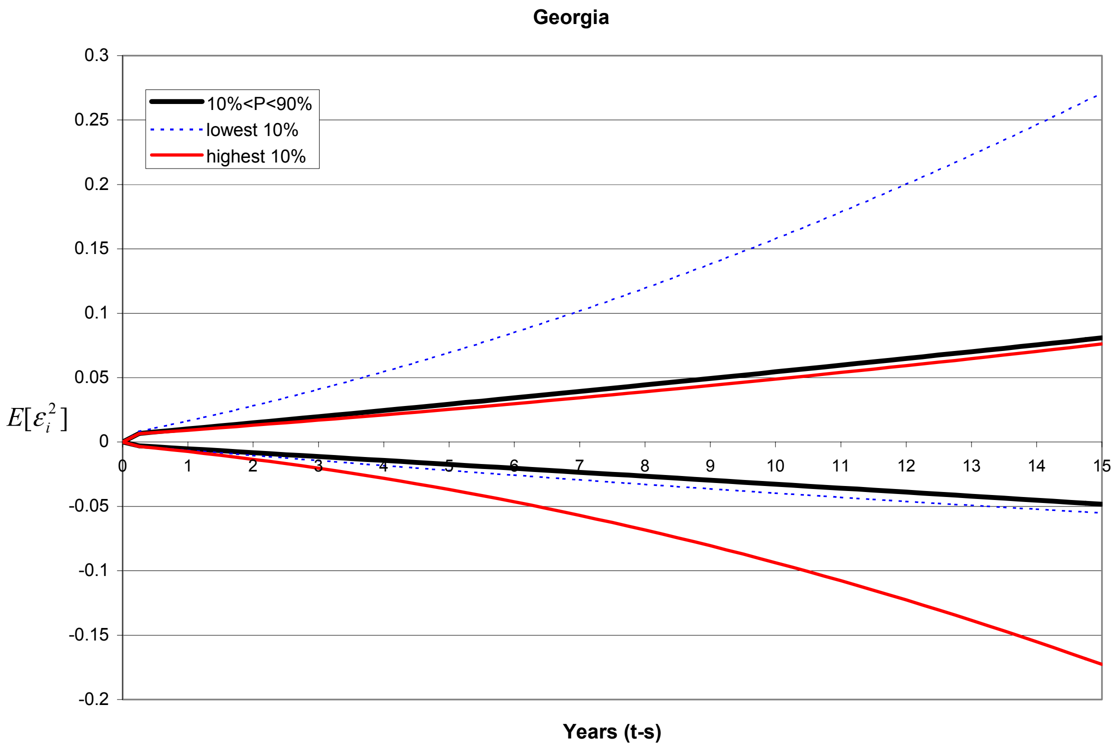


Figure 14: Expected Variance, By Time Between Transaction and Price Tier

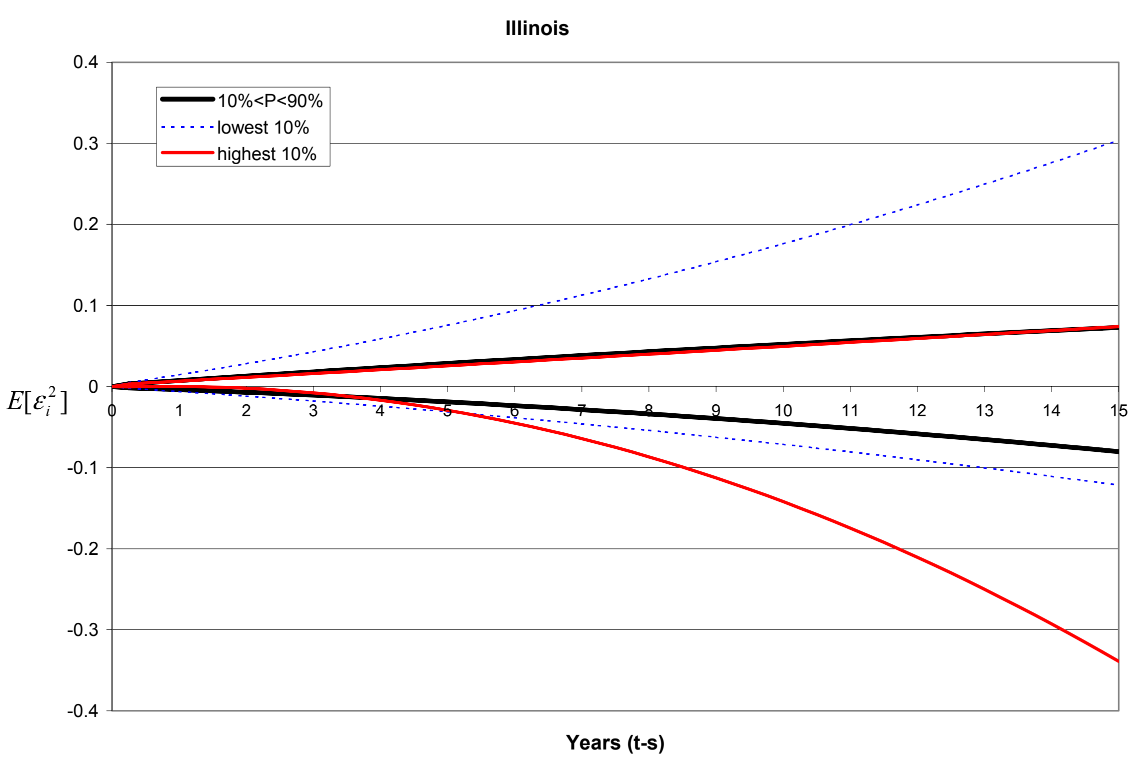


Figure 15: Expected Variance, By Time Between Transaction and Price Tier

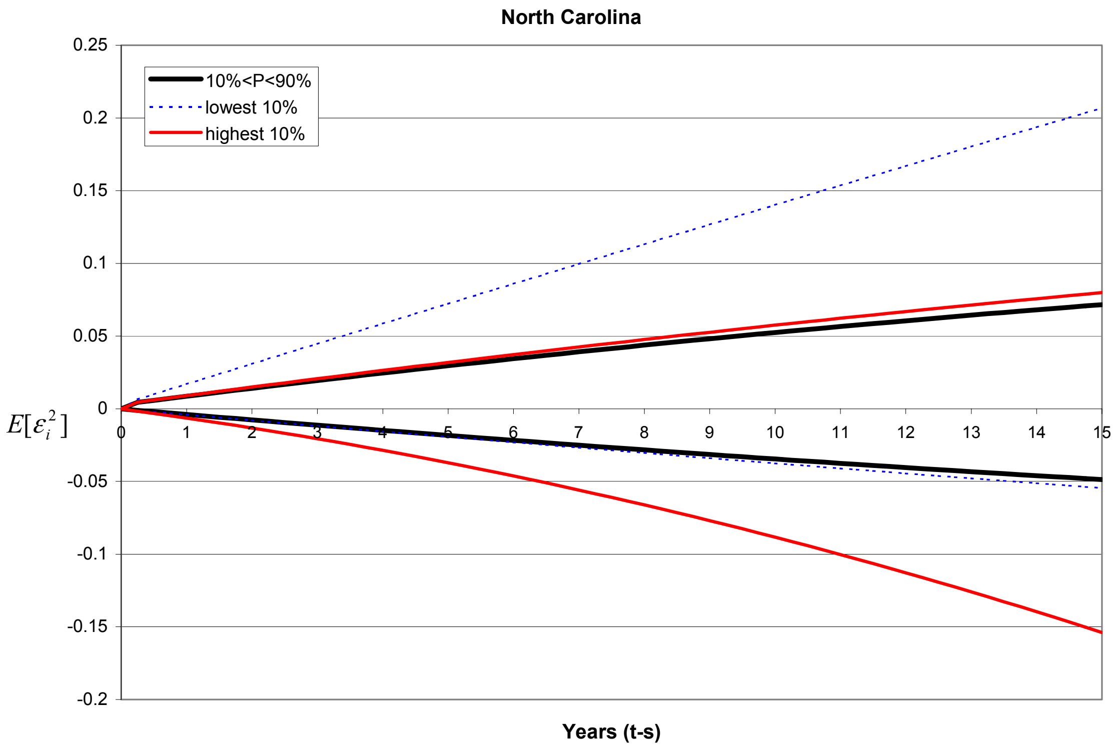


Figure 16: Expected Variance, By Time Between Transaction and Price Tier

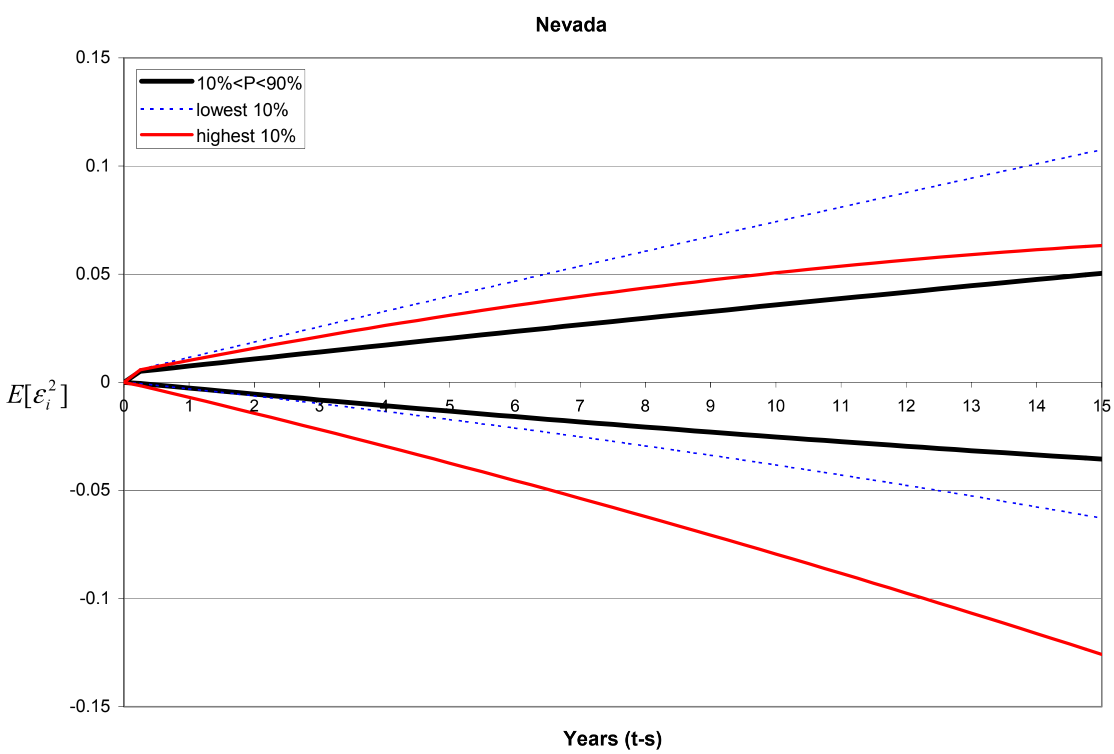


Figure 17: Expected Variance, By Time Between Transaction and Price Tier

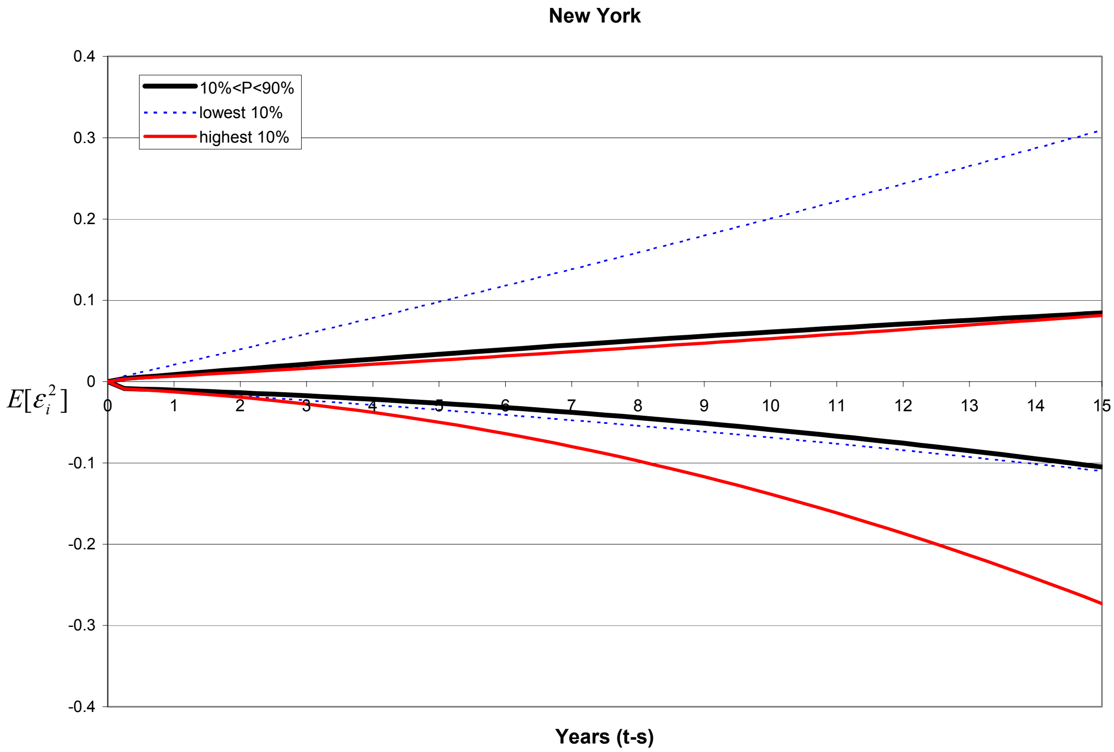


Figure 18: Expected Variance, By Time Between Transaction and Price Tier

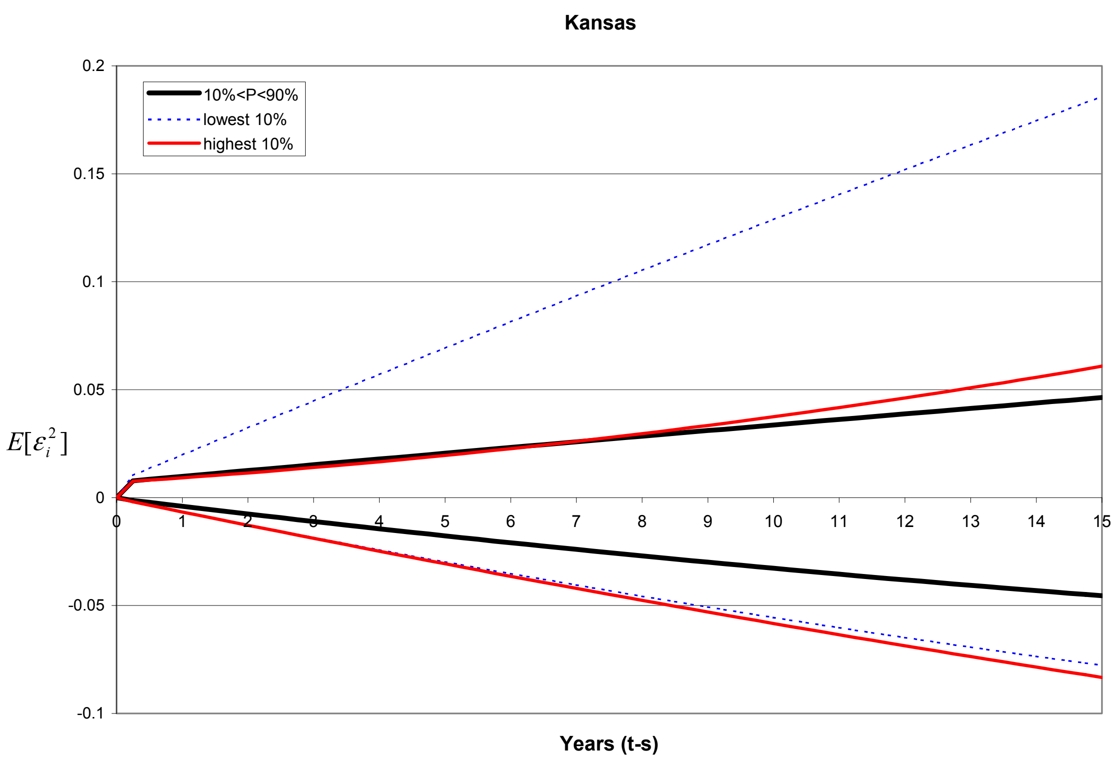


Figure 19: Expected Variance, By Time Between Transaction and Price Tier

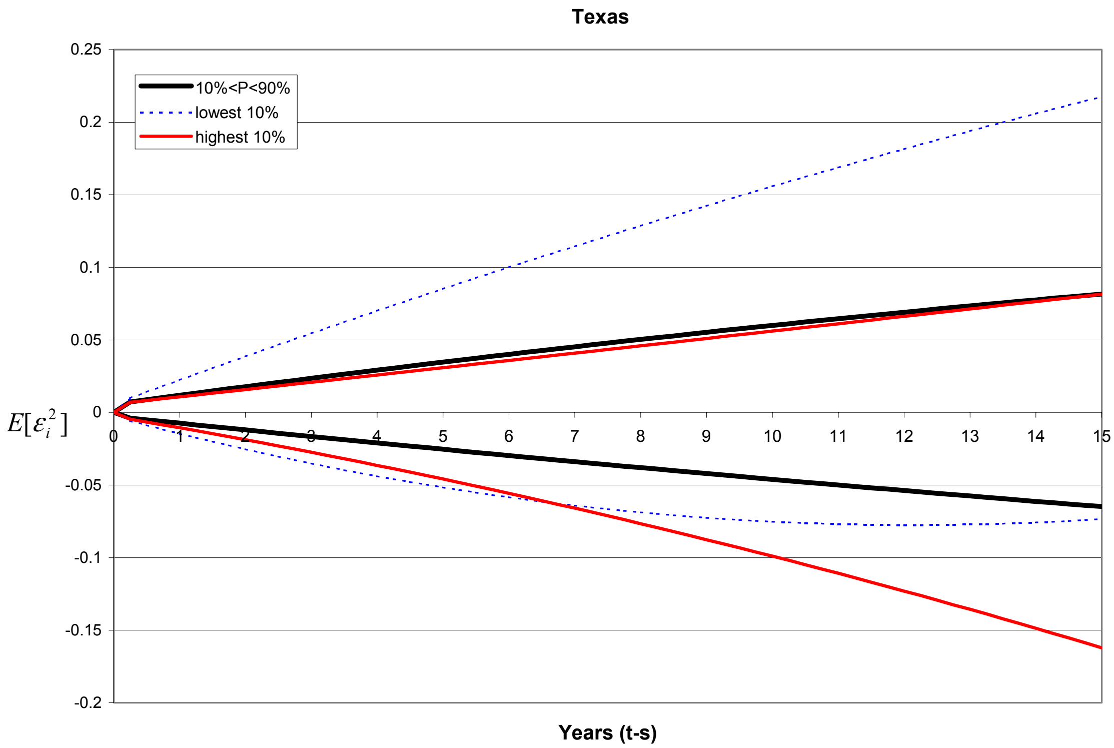


Figure 20: Expected Variance, By Time Between Transaction and Price Tier

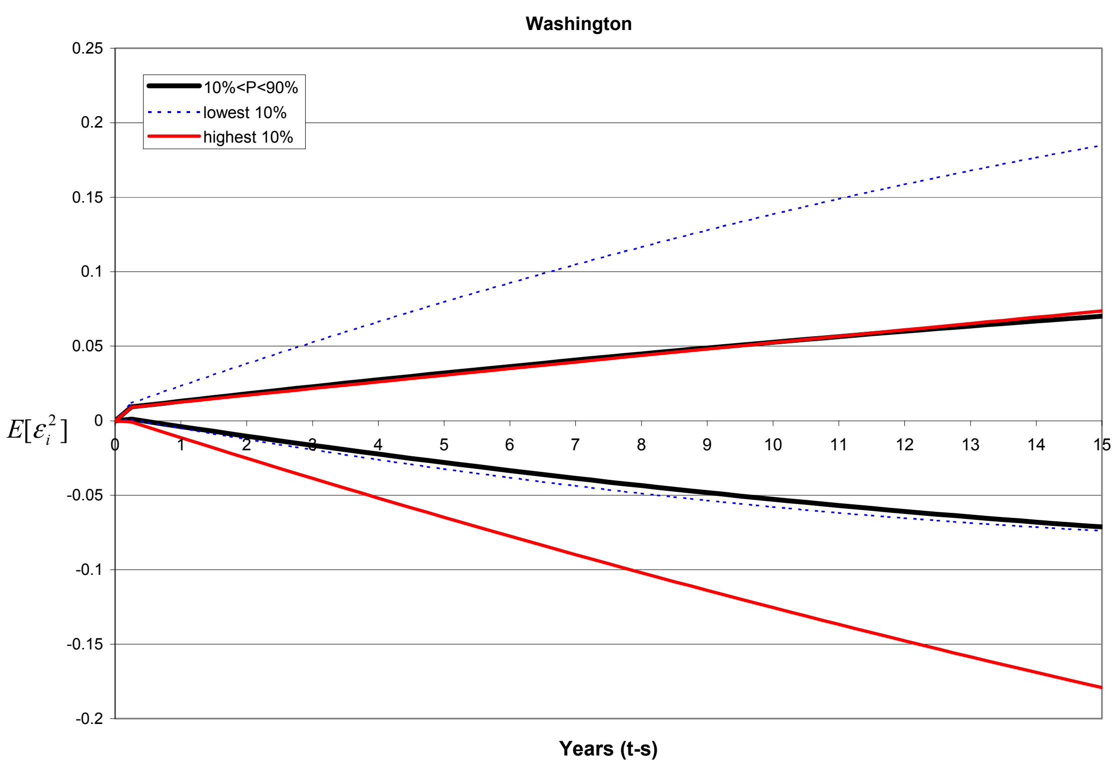


Figure 21: House Price Indices Under Various Weighting Schemes

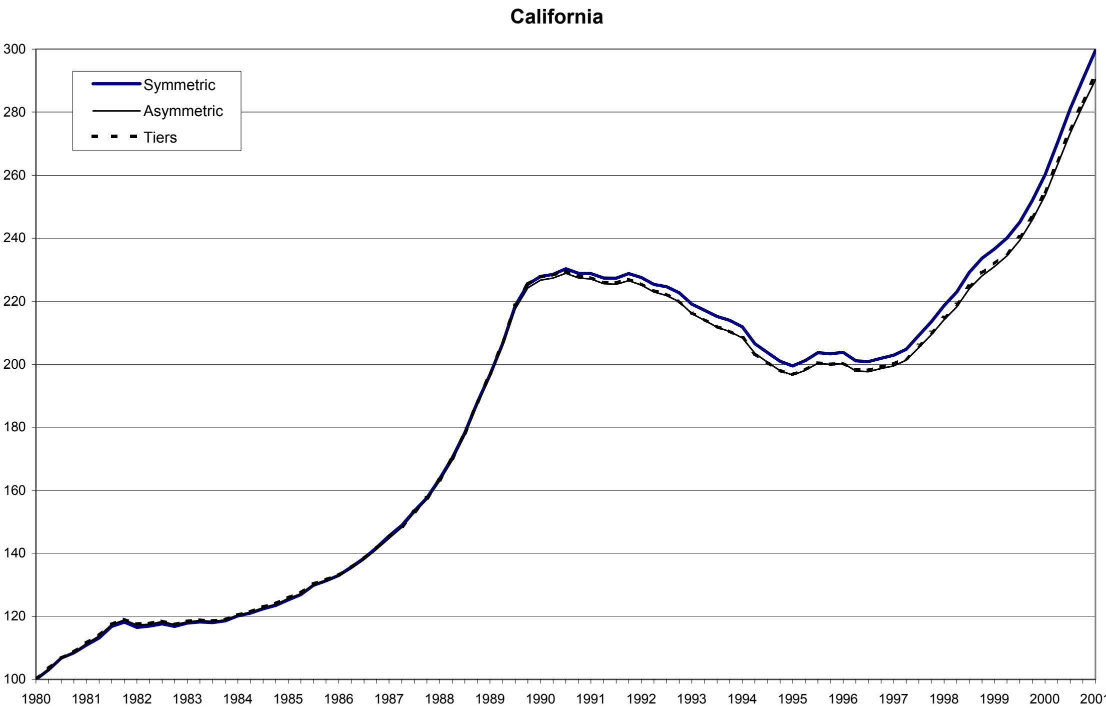


Figure 22: House Price Indices Under Various Weighting Schemes

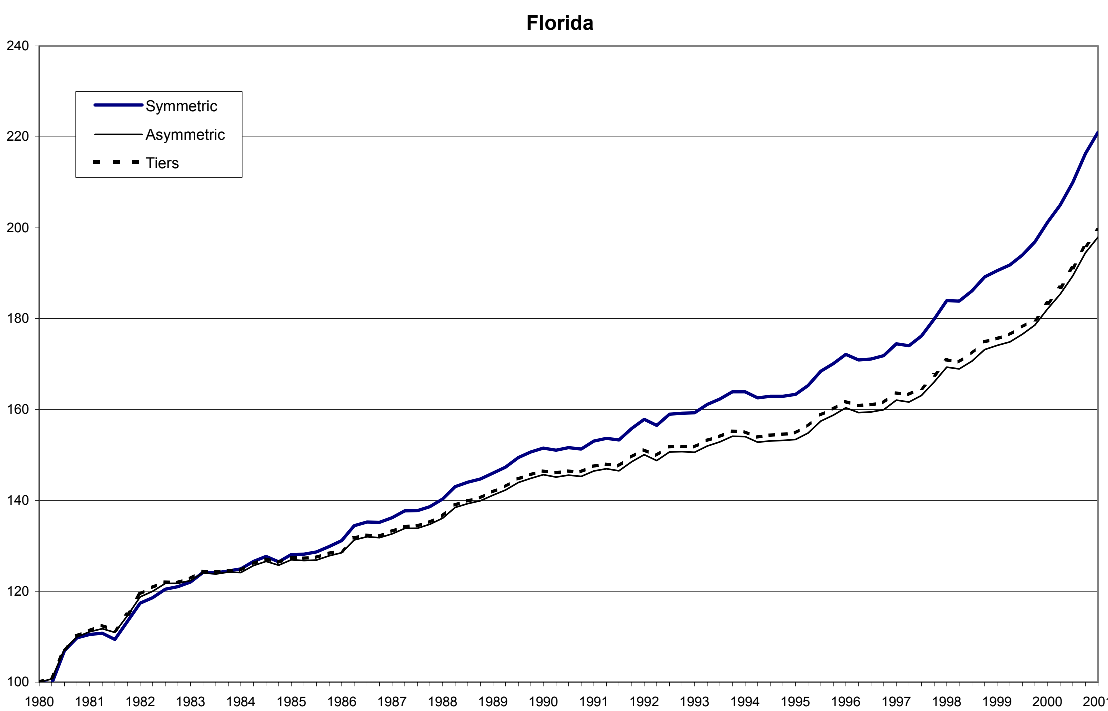


Figure 23: House Price Indices Under Various Weighting Schemes

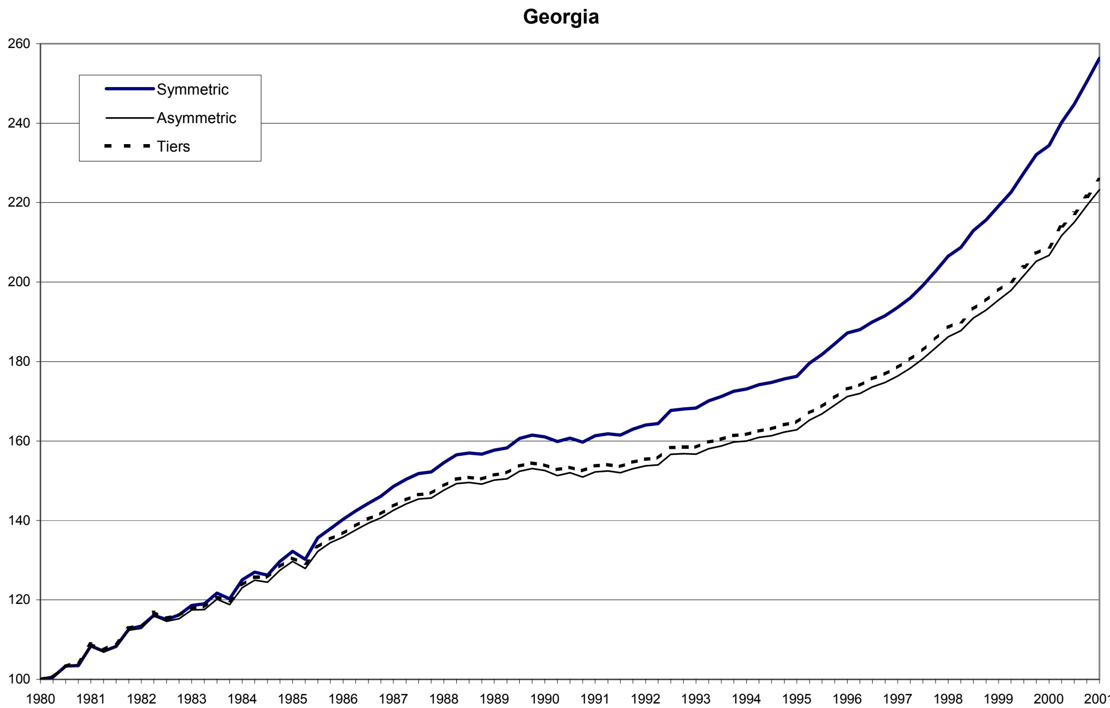


Figure 24: House Price Indices Under Various Weighting Schemes

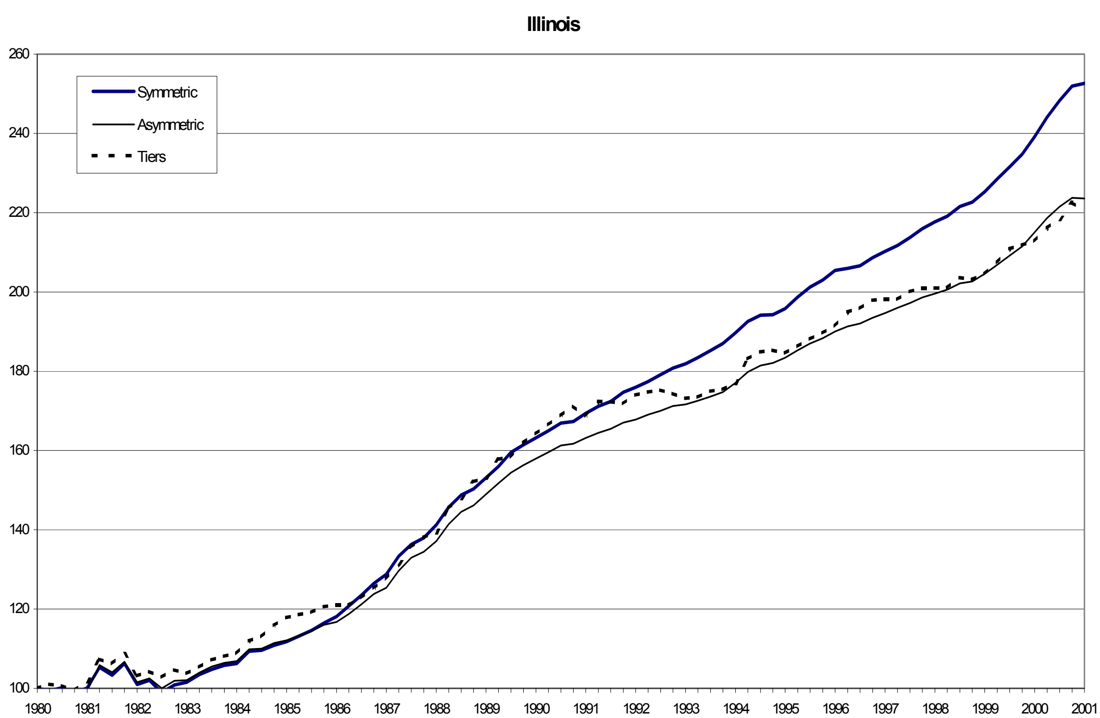


Figure 25: House Price Indices Under Various Weighting Schemes

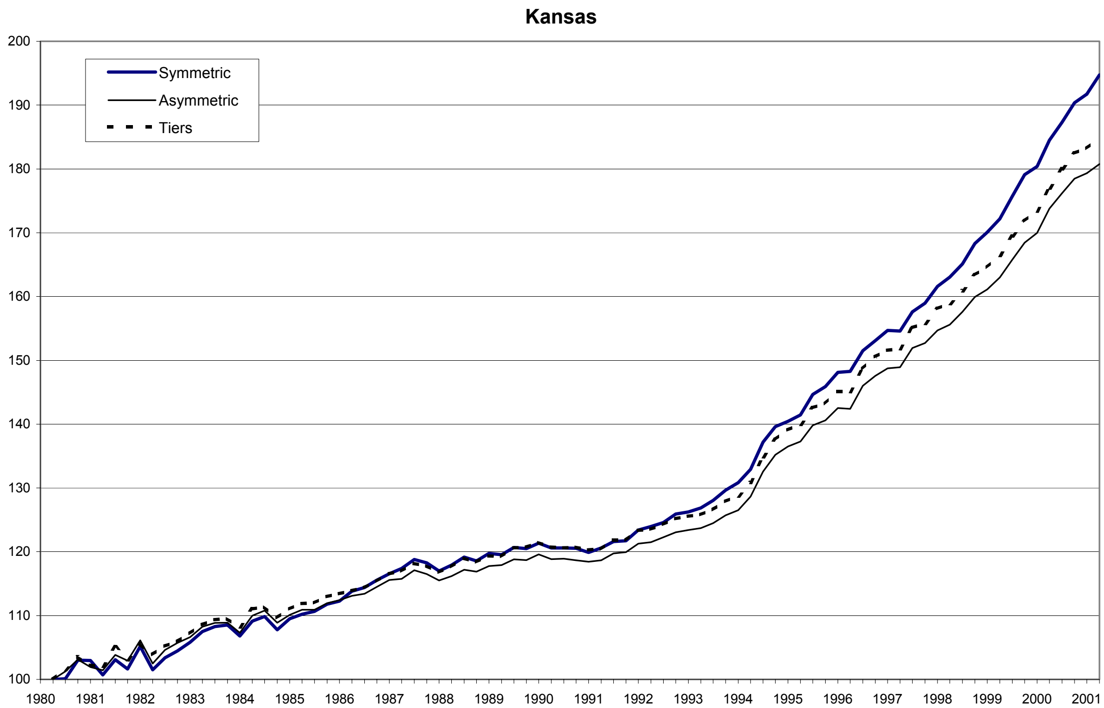


Figure 26: House Price Indices Under Various Weighting Schemes

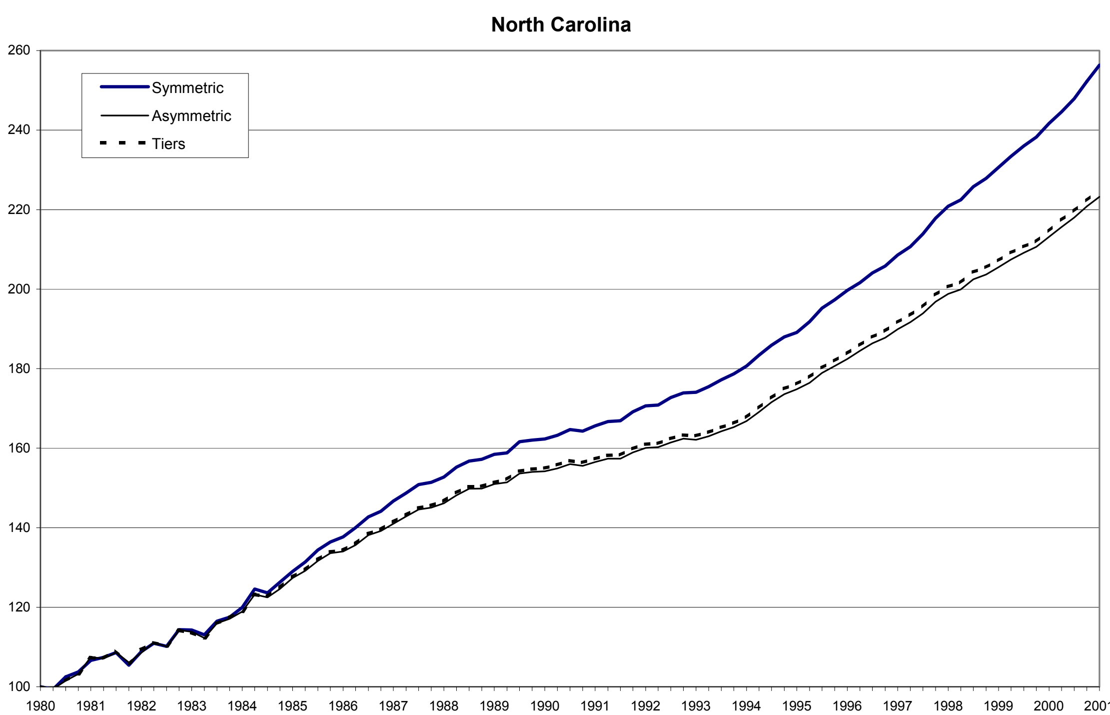


Figure 27: House Price Indices Under Various Weighting Schemes

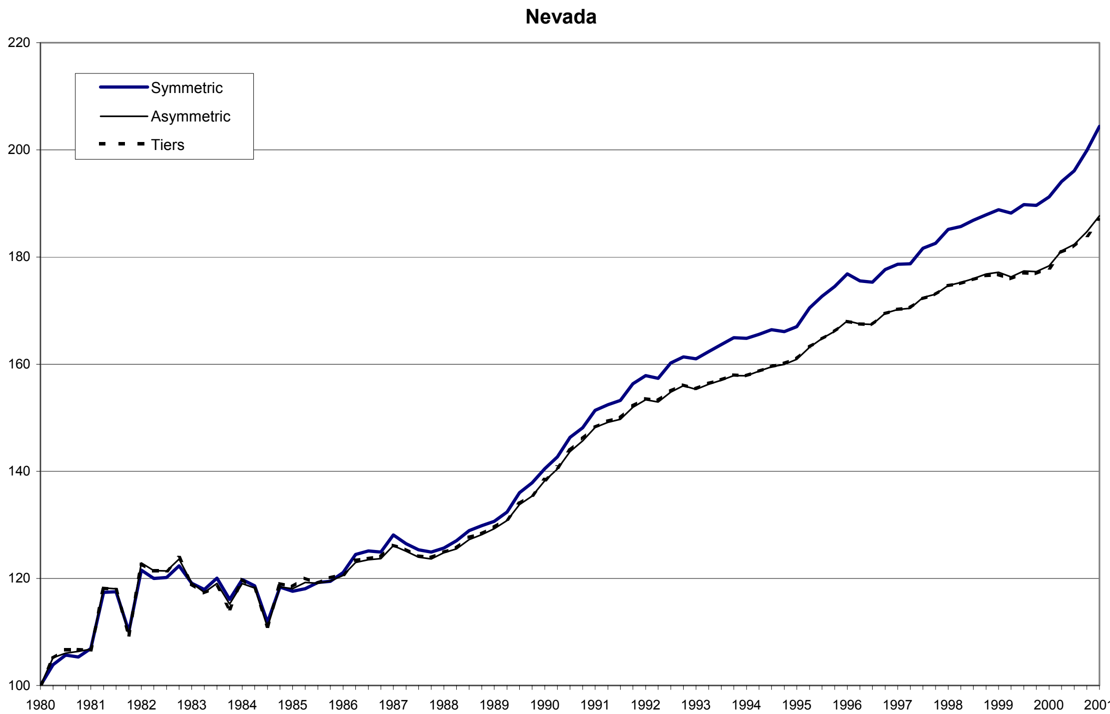


Figure 28: House Price Indices Under Various Weighting Schemes

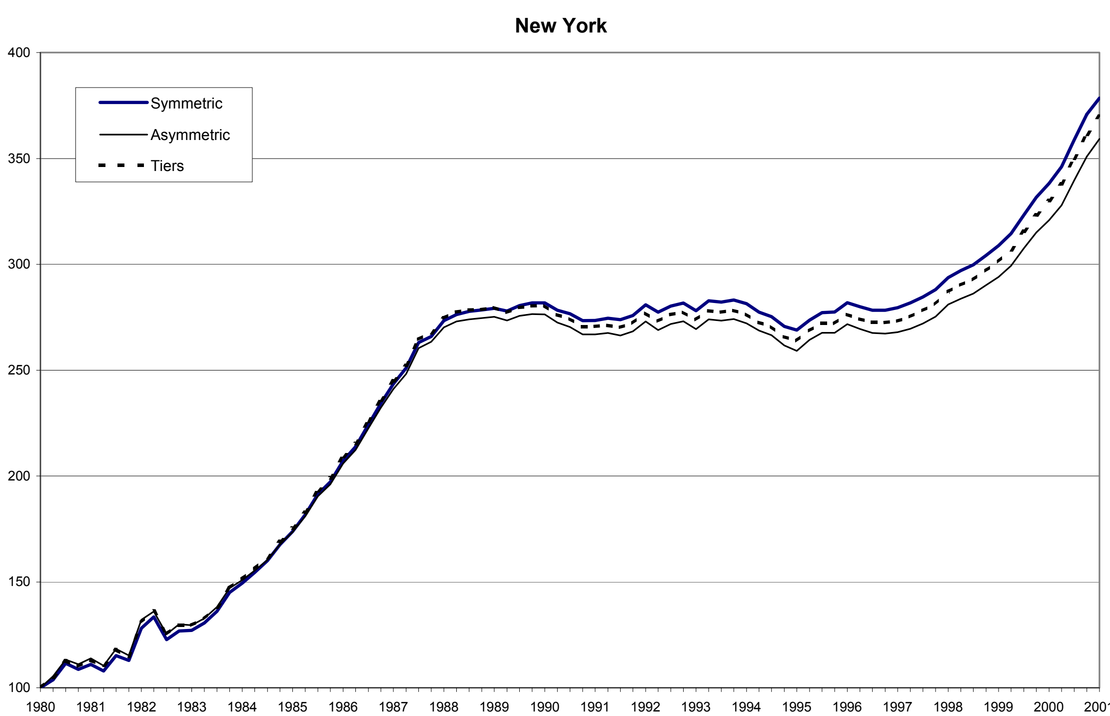


Figure 29: House Price Indices Under Various Weighting Schemes

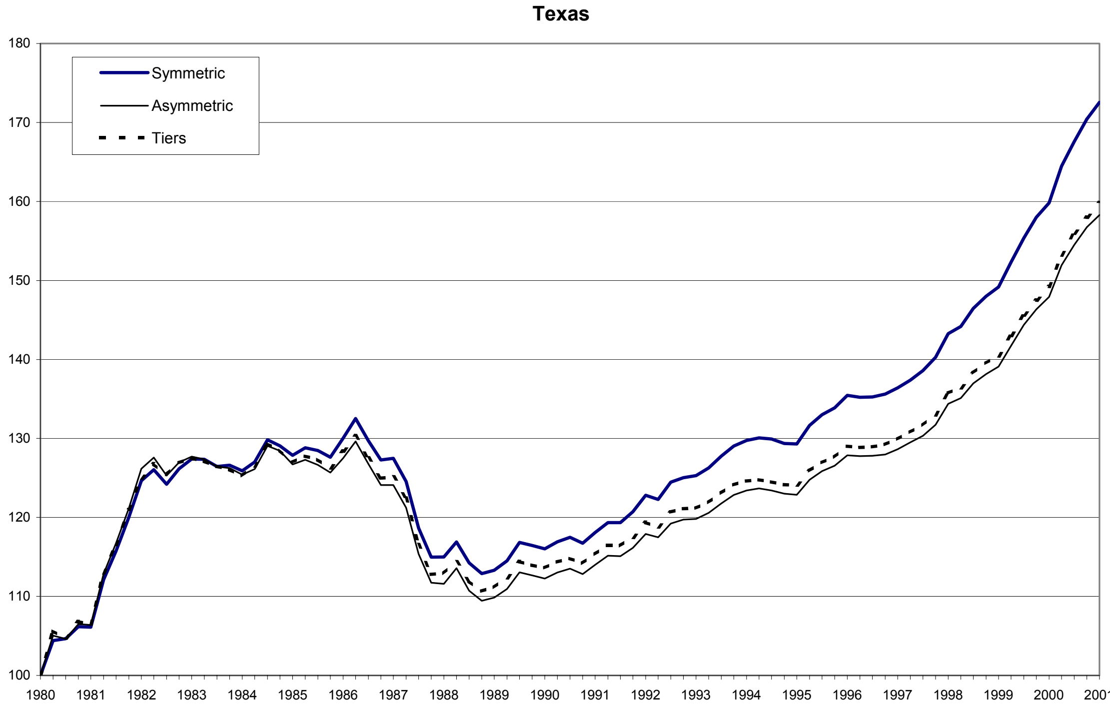


Figure 30: House Price Indices Under Various Weighting Schemes