OFHEO House Price Indexes :

HPI Technical Description

by

Charles A. Calhoun*

March 1996

Office of Federal Housing Enterprise Oversight
1700 G Street, N.W.
Washington, D.C. 20552
(202) 414-3800

*Senior Economist, Office of Research, Analysis, and Capital Standards, OFHEO.

1. Introduction

The Office of Federal Housing Enterprise Oversight (OFHEO) estimates and publishes quarterly house price indexes for single-family detached properties using data on conventional conforming mortgage transactions obtained from the Federal Home Loan Mortgage Corporation (Freddie Mac) and the Federal National Mortgage Association (Fannie Mae).[1] The house price indexes published by OFHEO -- hereafter referred to collectively or individually as the "HPI" -- are based on a modified version of the weighted-repeat sales (WRS) methodology proposed by Case and Shiller (1989). This paper provides background and a technical description of the data and statistical methods used to estimate the HPI.

OFHEO commenced publication of the HPI in March 1996. Quarterly house price indexes are reported for the nation, the nine U.S. Census divisions, and the 50 states and the District of Columbia, and are scheduled for release approximately two months after the end of the previous quarter.[2] The HPI for each geographic area is estimated using repeated observations of housing values for individual single-family residential properties on which at least two mortgages were originated and subsequently purchased by either Freddie Mac or Fannie Mae since January 1975. As of December 1995 there were over 6.9 million repeat transactions in the national sample. The use of repeat transactions on the same physical property units helps to control for differences in the quality of the houses comprising the sample used for statistical estimation. For this reason the HPI is described as a "constant quality" house price index.

[1] Freddie Mac and Fannie Mae are government sponsored enterprises (GSEs) chartered by Congress to provide a secondary market in conventional mortgages and increase lending for home ownership. Title XIII of the Housing and Community Development Act of 1992, PL 102-550, known as the Federal Housing Enterprises Financial Safety and Soundness Act of 1992, established OFHEO as an independent office within the Department of Housing and Urban Development with responsibility for ensuring that Fannie Mae and Freddie Mac are adequately capitalized and operating in a safe and sound manner.

The HPI provides broad geographic coverage by virtue of the national operations of the two government sponsored housing enterprises. This is one of the key advantages of an index based on enterprise data. There are, however, some limits to the coverage of the HPI. The HPI is produced using data on single-family detached properties financed by conforming conventional mortgages purchased by the enterprises. Thus, mortgage transactions on attached and multi-unit properties, properties financed by government insured loans, and properties financed by mortgages exceeding the conforming loan limits determining eligibility for purchase by Freddie Mac or Fannie Mae are excluded.[3] The HPI is updated each quarter as additional mortgages are purchased by the Enterprises and used to identify additional repeat transactions for the most recent quarter and all previous quarters.

2. Background

Freddie Mac developed the first repeat transactions house price indexes using a national database of secondary mortgage market transactions in 1989 (Abraham and Schauman, 1991).[4] Fannie Mae followed with its own version of the index in 1991 (Calhoun, 1991).[5] The Enterprises subsequently applied repeat transactions house price indexes to the problem of calculating changes in

---

[2] The national index is a weighted combination of the nine Census division indexes based on the distribution of owner-occupied households in the 1990 Census of Population and Housing.

[3] The conforming loan limit for mortgages secured by one-unit single-family properties was $203,150 in 1995 (50 percent higher in Alaska and Hawaii). The conforming loan limit for 1996 is $207,000. Increases in the conforming loan limit are based on the annual October to October percent change in property values as reported by the Federal Housing Finance Board.

[4] Data on eight million loans that Freddie Mac had purchased and securitized over the preceding twenty years were used to develop a sample of approximately 200,000 repeat transactions. The sample of repeat transactions included approximately 60,000 transactions in which purchase money originations provided an observation of the actual sale price of a property at two points in time. Appraisal (refinance) data were used to increase the total sample size to 200,000.

[5] The Fannie Mae indexes reported in Calhoun (1991) were developed for the same four cities (Atlanta, Chicago, Dallas, and San Francisco) examined by Case and Shiller (1987).

loan-to-value ratios on historical and outstanding mortgage acquisitions in their internal pricing and capital adequacy models.

In 1992, Fannie Mae and Freddie Mac began to explore the possibility of developing a joint version of their WRS indexes. An agreement was subsequently reached between the two companies for an exchange of data needed to produce a combined database of repeat transactions. A primary motivation for a joint index was the potential for improvements in statistical accuracy and geographic coverage resulting from a larger sample of repeat transactions. Although both Enterprises operate nationally, there are differences in the geographic coverage of their historical mortgage originations. Combining data produces a broader, more representative geographic distribution of properties. In addition, the WRS index methodology is inherently wasteful of data, as only properties that have appeared at least twice in their combined portfolios are eligible for inclusion in the sample of repeat transactions. Because of the potential for creating additional property matches across portfolios, combining data more than doubles the number of repeat transactions available for analysis.

The Enterprises have released quarterly house price indexes based on a version of the WRS methodology since January 1994. The Conventional Mortgage Home Price Index (CMHPI) has been published jointly by the Enterprises at the national and nine Census division levels (Stephens *et al.*, 1995). Freddie Mac has published additional indexes for 50 metropolitan statistical areas (MSAs).[6] As the financial safety and soundness regulator of the Enterprises, OFHEO determined that house price indexes based on Enterprise data would be more appropriate for the purposes of establishing risk-based

[6] The Conventional Mortgage Home Price Index is reported on the Bloomberg News and Information Service under the acronym CMHP.

capital standards for the Enterprises than the Constant Quality House Price Index published by the Secretary of Commerce, and that this required production and publication of the indexes by OFHEO.[7]

3. Repeat Transactions Data

The HPI is estimated using data on repeat transactions obtained by combining data on single-family mortgage acquisitions provided to OFHEO by the Enterprises. Approximately one month after the end of each quarter Enterprise data on all single-family mortgage acquisitions through the previous quarter are delivered to OFHEO on computer tape.[8] Prior to matching on property addresses, the data are processed using commercially available software designed to standardize street addresses to conform to U.S. Postal Service standards. This significantly improves the success rate of address matching and facilitates quality control of the resulting database of repeat transactions.

The mortgage origination records for the enterprises are combined and sorted by ZIP code, property address, and origination date. This places consecutive transactions on the same property in sequentially adjacent positions in the data set. Matches are then identified by comparing the address fields on consecutive records. When a match is identified, the information on the property values and transaction dates from both records is retained and written as a single record to a new data file of repeat transactions.

4. Statistical Methods

---

[7] The Federal Housing Enterprises Financial Safety and Soundness Act of 1992 ("the Act") prescribes the use of house price indexes to account for changes in the values of the properties securing mortgages guaranteed by the enterprises and their potential impact on credit risk. The Act requires that changes in property values be determined in accordance with the Constant Quality (CQ) Home Price Index published by the Secretary of Commerce. The statute also provides, subject to certain conditions, for the use of an alternative index. The specific requirements for the use of a house price index in the stress test are given in Section 1361 of the Act.

[8] The data include mortgage records for single-family, single-unit, detached properties, excluding condominiums, cooperatives, and planned urban developments (PUDs).

The repeat sales method was first proposed by Bailey, Muth, and Nourse (1963), and later extended by Case and Shiller (1987, 1989).[9] This approach limits the extent to which changes in the composition of the sample used for estimation can influence the estimated index. Utilizing information on the values of the same physical units at two points in time controls for differences in housing attributes across properties in the sample without directly estimating their marginal contribution to total value. A multivariate regression is employed to account for the fact that all properties do not transact in every period. The repeat sales method is the only approach that provides an opportunity to develop constant quality house price indexes using GSE data. The lack of information on detailed property characteristics in historical GSE data precludes the estimation of hedonic house price indexes.

The WRS method has been motivated by both practical econometric considerations and the characterization of house price changes as a stochastic diffusion process. The use of repeat transactions requires one to consider the implications of the elapsed time between observed transactions for the sampling properties of the statistical estimates. The sampling variability of observed changes in housing values is assumed to increase with the length of time between transactions, as factors other than market appreciation are increasingly likely to influence the values of individual housing units.[10]

---

[9]Research on house price indexes has expanded in recent years with the emergence of new large scale national databases and advancements in statistical methods. A number of recent efforts were presented in the Fall 1991 Special Issue of the *Journal of the American Real Estate and Urban Economics Association*. Of special interest are the paper by Abraham and Schauman (1991) on the Freddie Mac repeat sales index, and the paper by Case, Pollakowski, and Wachter (1991) on choosing among house price index methodologies. In April 1994, Fannie Mae hosted a research roundtable on the research and business uses of house price indexes. Many of the papers presented at the Fannie Mae conference were subsequently published in a special issue of the *Journal of Housing Research* (Megbolugbe, 1995). *The Journal of Real Estate Finance and Economics* will publish a two-volume special issue on house price indexes in 1996 (Thibodeau, 1996) .

[10] For example, differential rates of depreciation for otherwise similar properties are more likely over longer periods of time. In addition, localized value differences may appear over time as neighborhood identities emerge or evolve with changes in the socioeconomic and demographic characteristics. Differences in house price growth rates associated with sub-market heterogeneity is an additional source of variation around a market index.

Case and Shiller (1987) extended the basic approach by proposing the use of generalized least squares to account for heteroscedastic sampling errors whose variances were assumed to be proportional to the length of time between repeat transactions. This simple assumption about the sampling variances of changes in housing values for repeat transactions is consistent with the characterization of house prices as a stochastic log-normal diffusion process similar to those used to model uncertainty about interest rates and asset prices in financial simulation models.

It is now standard practice in the housing research literature to characterize individual house prices as arising from a stochastic process in which the average rate of change or drift in housing values is represented by a market index and the dispersion and volatility of values around the market average are modeled as a log normal diffusion process.[11] In this approach, one assumes that the 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

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

This implies that the total percentage change in price for house i transacting in time periods s and t is given by

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

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

The assumption that house prices obey a Gaussian diffusion and additional assumptions about the error process are summarized as follows:

[11] See, for example, Foster and Van Order (1984, 1985), Cunningham and Hendershott (1984), Cooperstein (*et al.*, 1991), and Quigley and Van Order (1992).

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

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

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

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

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

for all *i*, and *j*, and $t > s$. The market price index b_t represents the average behavior of housing values in a given market, and remains unrestricted. The Gaussian random walk H_{it} describes how variation in individual house price growth rates around the rate of change in the market index causes house prices to disperse over time.[12] 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. The N_{it} are assumed to be uncorrelated over time and across properties.

Estimates of variance parameters *A, B,* and *C* emerge as a by-product of the application of the WRS methodology. In a sample of repeat sales or mortgage transactions, the difference in the natural logarithm of the price of house *i* that is observed to transact at any two dates can be expressed more generally by:

$$\Delta V_i = \sum_{t=0}^{T} \ln(P_{it}) D_{it} \quad (9)$$

[12]The standard assumption that the error variance of the (log) change in house value is proportional t the length of time between repeat transactions was modified by Abraham and Schauman (1991) to allow for a diminution of the rate of increase in volatility through the use of a quadratic function of time.

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

$$\Delta V_i = \sum_{t=0}^{T}(b_t + H_{it} + N_{it})D_{it} \tag{10}$$

$$\Delta V_i = \sum_{t=0}^{T} b_t D_{it} + e_i \tag{11}$$

The b_t parameters for the market index can be estimated by ordinary least squares (OLS) regression.[13] When *A* or *B* are non-zero, the variance of e_i varies with the length of time between repeats sales or transactions, and more efficient estimates of b_t, $t = 0, 1, 2, \ldots, T$ can be obtained by using a generalized least squares (GLS) procedure. The results from the first-stage OLS regression of equation (11) on a sample of repeat transactions are used to construct estimates of the squared deviations of observed house prices around the estimated market index. Note that for house *i* with transactions in periods *s* and *t*, the predicted house price in period *t* given house price in period *s* is:

$$\ln(\hat{P}_{it}) = \ln(P_{is}) + (\hat{b}_t - \hat{b}_s). \tag{12}$$

Thus, the predicted price in period *t* is the original price plus expected market appreciation, and the squared deviations of observed house prices from the market index are given by:

$$\begin{aligned} d_i^2 &= [\ln(P_{it}) - \ln(\hat{P}_{it})]^2 \\ &= [\ln(P_{it}) - \ln(P_{is}) - \hat{b}_t + \hat{b}_s]^2 \end{aligned} \tag{13}$$

Using equations (5) and (8) it can be shown that this expression has expectation given by:

$$E[d_i^2] = A(t-s) + B(t-s)^2 + 2C. \tag{14}$$

Estimating a second-stage regression of d_i^2 on *(t-s)*, $(t\text{-}s)^2$, and a constant term provides consistent estimates of *A*, *B*, and *C*, and direct evidence about the volatility and dispersion of individual house prices around the market index. The predicted values of the squared deviations, $\hat{d}_i^2$, are used to derive the weights needed to obtain GLS estimates of the b_t parameters in the following regression:

$$\frac{\Delta V_i}{\sqrt{\hat{d}_i^2}} = \sum_{t=0}^{T} b_t \frac{D_{it}}{\sqrt{\hat{d}_i^2}} + \frac{e_i}{\sqrt{\hat{d}_i^2}} \tag{15}$$

Equation (15) can be estimated for selected geographic areas to derive WRS house price indexes. Index numbers for periods $t = 1, 2, 3, \ldots, T$ are given by:

$$I_t = 100 \cdot e^{\hat{\hat{b}}_t} \tag{16}$$

where $\hat{\hat{b}}_t$, $t = 1, 2, 3, \ldots, T$ are the GLS parameter estimates.[14]

In practice, the estimates of C obtained from the second-stage regression may turn out to be negative. This can result in predicted values of d_i^2 that are negative for properties with sufficiently short waiting times between repeat transactions. This outcome is inconsistent with the assumption that this term represents the error variance in the log difference in house values and precludes the use of the estimated error variance for these observations to derive the weights for the third stage of the GLS procedure. An alternative approach that avoids this problem is to

---

[13] It is necessary to restrict one of the market index parameters to avoid perfect collinearity among the explanatory variables. It is convenient to use $b_r = 0$, where *r* is the base period of the reported index.

[14] If the restriction $b_1 = 0$ is imposed in estimation, then $I_1 = 100$.

assume that N_{it} is a constant, N_i, for each house. Under this assumption, the N_{it} terms cancel along with all other fixed terms in equation (1). The second-stage OLS regression now becomes:

$$E[d_i^2] = A(t-s) + B(t-s)^2. \qquad (17)$$

This approach constrains the estimated error variance associated with the distribution of individual house price appreciation rates to be positive. This is consistent with the interpretation of changes in individual housing values as a diffusion process -- and the fact that the actual change in housing values must be zero until some time has elapsed.[15] This constraint is imposed in the estimation of the HPI.

5. Arithmetic Versus Geometric Indexes

An alternative to the geometric weighting implicit in the GLS procedure in equation (15) is the arithmetic repeat sales index procedure described by Shiller (1991). Shiller noted that a geometric index will underestimate the percentage change in the average value of a portfolio comprising individual real estate assets with different values. He proposed an arithmetic WRS estimator that is theoretically superior for the portfolio valuation problem. However, the estimation procedure underlying the arithmetic WRS index is more complicated to apply than the geometric WRS method, and is not directly related to a model of individual housing values. Goetzmann (1992) proposed the following simple modification to the geometric WRS index that approximates the appreciation rates obtained from the arithmetic index:

[15] One interpretation of the N_{it} values is that they represent idiosyncratic differences in values associated with random events in the search process or the behavior of real estate agents (Case and Shiller, 1990, p. 255). Thus, they may be interpreted as differences in offers made on any given date, in which case no time must elapse for differences in N_{it} values to result. Obtaining negative estimates of the variance the N_{it} suggests that it may not be possible to identify this relatively subtle effect in the indirect manner proposed by Case and Shiller.

$$\tilde{I}_t = 100 \cdot e^{\hat{\hat{b}}_t + \frac{1}{2}\hat{s}_t^2} \tag{18}$$

where $\hat{\hat{b}}_t$, $t = 1, 2, 3, \ldots, T$ are the GLS parameter estimates from the geometric WRS procedure, and $\hat{s}_t^2$ is an estimate of the variance in house price growth rates associated with the diffusion of house prices after t periods, given by:

$$\hat{s}_t^2 = \hat{A} \cdot t + \hat{B} \cdot t^2 \tag{19}$$

where $\hat{A}$ and $\hat{B}$ are the OLS estimates of A and B from the second stage of the WRS procedure.[16]

OFHEO publishes geometric WRS indexes along with estimates of growth rate diffusion parameters A and B for each region that can be used to derive the Goetzmann correction for house price changes between any two dates in the index series. This enables HPI users who are primarily concerned with portfolio valuation to make adjustments that are appropriate to specific holding periods that may not correspond to the beginning and ending dates of the published index series.

[16]The CMHPI includes the Goetzmann correction for indexes reported from the base period (1987 Q1) to the most recent quarter.

6. Standard Errors for Repeat Transactions Indexes

HPI quarterly index numbers based on the estimated market index parameters are reported for the period from 1980 to the most recent quarter. Asymptotic standard errors for the quarterly index numbers, valid in large samples, are derived from the standard errors for the b_t estimates. If I_t is the index number for period t, and $s_{\hat{b}_t}$ is the standard error of the GLS estimate of b_t, then the standard error of the index number is given by:

$$s_{I_t} = I_t \cdot s_{\hat{b}_t} \quad . \tag{20}$$

Estimates of index standard errors are obtained by replacing $s_{\hat{b}_t}$ with its estimated value from the third stage of the GLS estimation procedure.

7. Examples

The attached statistical supplement to the quarterly HPI release provides examples of the HPI index series and growth rate diffusion parameter estimates that are released each quarter. Quarterly index numbers are reported for the period from 1980 Q1 -- the base period of the HPI -- to the most recent quarter (1995 Q4 in the sample output). Estimated standard errors are reported in parentheses after each index number. Estimates of growth rate diffusion parameters A and B and the implied annualized rate of house price volatility associated with each index series are also reported in a separate table. The annualized volatility estimates are computed as the standard deviation of the distribution of the percentage change in housing values one year (4 quarters) following mortgage origination:[17]

---

[17] The annualized rate of house price diffusion (volatility) is computed for the first year after a house value is assumed to have been observed. Because the volatility parameters were estimated using quarterly data, the estimated volatility after 4 quarters gives the annualized volatility for year one. The estimated volatility will not

$$Volatility = \sqrt{\hat{A} \cdot 4 + \hat{B} \cdot 16} \qquad (21)$$

8. Future Research

OFHEO is conducting research to further investigate the statistical properties of repeat transactions house price indexes. Many of these issues are discussed in the paper by Stephens *et al.* (1996). Topics for future research include: geographic and temporal aggregation, revision volatility, the use of appraisal values, sample selection, and comparisons with alternative methods.

---

increase linearly with time because of the use of the quadratic specification in estimating the error variance of house price changes for repeat transactions.

Literature Cited

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 M.J., Muth, R.F. and Nourse, H.O. (1963). "A Regression Method for Real Estate Price Index Construction," *Journal of American Statistical Association*. 58, 933-942.

Calhoun, C.A. (1991) "Estimating Changes in Housing Values from Repeat Transactions," presented at the 1991 Western Economic Association International meetings, Seattle, Washington, June 30-July 3.

Case, Bradford, Henry Pollakowski, and Susan Wachter,(1991). "On Choosing Among House Price Index Methodologies," *AREUEA Journal*, 19, 286-307.

Case, K.E. and Shiller, R.J. (1987). "Prices of Single Family Real Estate Prices," *New England Economic Review*. 45-56.

Case, K.E. and Shiller, R.J. (1989). "The Efficiency of the Market for Single-Family Homes," *The American Economic Review*. 79, 125-137.

Cooperstein, R.L. , F. S. Redburn, and H.G. Meyers, "Modelling Mortgage Terminations in Turbulent Times," *AREUEA Journal*, 19(4):473-494, Winter 1991.

Crone, Theodore M. and Richard P. Voith, (1992) "Estimating House Price Appreciation: A Comparison of Methods", *Journal of Housing Economics*. 2(4) 324-338.

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

Foster, C. and R. Van Order, "FHA Terminations: A Prelude to Rational Mortgage Pricing," *AREUEA Journal*, 13(3):273-91, Fall 1985.

Goetzmann, William, (1992). "The Accuracy of Real Estate Indexes: Repeat Sales Estimators," *Journal of Real Estate Finance and Economics*, 5, 5-53.

Quigley, J. and R. Van Order (1992). "More on the Efficiency of the Market for Single Family Homes: Default," Mimeo. University of California, Berkeley.

Shiller R.J. (1991) "Arithmetic Repeat Sales Price Estimators," *Journal of Housing Economics*. 1, 110-126.

Stephens, William; Ying Li, Vassilis Lekkas, Jesse Abraham, Charles Calhoun, and Thomas Kimner, "Conventional Mortgage Home Price Index," *Journal of Housing Research*, 6(3):389-418, 1996.

Thibodeau, T.G. (ed.) (1996). "House Price Indices: Part I," special issue of *The Journal of Real Estate Finance and Economics*, forthcoming.