Appendix C: Data for Loan Performance Simulations

This appendix describes the methodology we used to produce forecasts of future loan performance. We first describe how loan event data for future time periods were generated to project future loan performance and mortgage-related cash flows. This required creating future event data both for existing books of business and for future loan cohorts not yet originated. Then we summarize how the economic forecasts were applied. The forecasts of the economic factors are discussed in Appendix D. The derivation and application of the dispersion volatility estimates for the national average house price forecast is also explained here in detail.

I. Future Loan Event Data

The development of future loan event data was closely integrated with the development of the data used in the statistical estimation of loan performance described in Appendix A. As described in Appendix A, the process of building the historical loan event data entailed expanding FHA loan origination records into dynamic quarter-to-quarter event data from loan origination up to, and including, the period of loan termination. The loan event data were augmented with external economic data (house price indices and interest rates) to compute a number of time-varying predictors of conditional prepayment and claim rates.

For loans that did not terminate within the historical sample period (from FY 1975 Q1 to FY 2007 Q1), the process of building the period-by-period event data followed the same procedure as for terminating loans, but used forecasted values of the external economic factors to project future loan termination rates and cash flows.

In addition, we forecasted the loan performance of future FHA loan books originated through FY 2014. We based our actuarial projections of future loan books on estimates of total loan volume for FY 2007 through FY 2014 as provided by FHA from their internal demand model. To establish a baseline set of loan characteristics for computing predicted prepayment and claim rates, we utilized historical loan-level data for the last four complete quarters of the historical data sample (FY 2006 Q2-Q4 and FY 2007 Q1). In effect, we duplicated the loan origination records for these four quarters and assumed that the same loan composition will be originated during the ensuing quarters of FY 2007 to FY 2014.

We then updated the initial values of the economic variables for future loan originations based on the corresponding economic forecasts to reflect conditions at the time of origination. After we projected the future mortgage cash flows for the detailed loan stratifications, we scaled the results to conform to the projections of total future loan volume from the FHA demand model.

As described in Appendix A, the data used for statistical estimation comprise detailed loan stratifications grouped by age of loan, all possible combinations of the categorical outcomes for the explanatory variables, and additional categories such as mortgage product types. The data for future cohorts are organized in the identical manner.

II. Future Economic Forecasts

FHA received quarterly forecast data from Global Insight, Inc. These data served as the source of future interest rates and house price appreciation in our analysis.

For the projection of future changes in housing values, we used the Global Insight forecast of the national average sales price of existing single-family homes. Because the national-level house price series is an average of regional house price performance, it tends to smooth out differences in house price trends in the individual underlying regional indexes. This implies an additional layer of uncertainty with regard to the dispersion of individual house price appreciation rates around the market average, now represented by a national-level house price index (HPI). When using the national-level house price forecast to compute the probability of negative equity, it was important to take into account both sources of uncertainty.

To address this issue, we developed a methodology to estimate the historical dispersion of the OFHEO regional (Census division) and metropolitan HPIs in relation to the national average house price series. This analysis is described in greater detail in the next section of this appendix. To summarize, estimates of additional dispersion were combined with those for the underlying regional or metropolitan index to increase the overall estimated dispersion of the appreciation rates of individual houses within a geographical location. The additional volatility is added only from the start of the forecast period for future loan performance; that is, at the point that the computation of the probability of negative equity switches from using MSA-level indexes to the forecasted national-level HPI. The additional dispersion increases with time in a similar manner to the dispersion of individual property appreciation rates based on the MSA-level index.

Recall that the source of house price appreciation rates for historical loans was the HPIs obtained from OFHEO. The rule for assigning metropolitan area indexes was to use the Metropolitan Statistical Area Division (MSAD) index if one exists for the loan's Federal Information Processing Standards (FIPS) state-county code, otherwise we used the Core Based Statistical Area (CBSA) index if one is available. If no MSAD or CBSA index is available, we applied one of the nine Census-division HPIs.

As described in Appendix A, the indexes are used in conjunction with estimates of house price diffusion parameters to compute probabilities of negative equity at each loan age for individual borrowers. The volatility estimate reflects the uncertainty about the dispersion of individual house price appreciation rates around the national average appreciation rate.

III. Dispersion of Local House Price Indices

As also described in Appendix A, the distributions of individual housing values relative to the value at loan origination were computed using estimates of house price drift and diffusion volatility estimated by OFHEO.

To forecast the future changes in housing values, we utilized Global Insight's national-level HPI forecast. However, the Global Insight's national average house price data and forecast do not provide estimates of the diffusion volatility between a single property and the national index. Although OFHEO publishes a national-level HPI, it is based on a weighted average of indexes for the nine Census divisions, and no separate diffusion parameters are produced at the national level. Thus, direct estimates of house price dispersion around a national index do not exist. Since we used Census-division-level and metropolitan-level indexes for estimating historical loan performance models, we elected to adopt an approach that would build upon the OFHEO local house price volatility estimates and modify them to be consistent with the forecasting period when a national-level index is used.

We estimated the variance of the geometric growth rates of housing values implied by the division indexes around the geometric growth rates of the national-level index. The following discussion uses the case of MSA indexes as an example, but the same approach is applied in the case of the Census-division indexes.

The growth rate for property *i* between time periods *t* and *s* relative to its MSA index is given by:

$$\ln(G_{i,t,s}) = \ln(H_{i,t}) - \ln(H_{i,s}) = \ln(H_{MSA,t}) - \ln(H_{MSA,s}) + \varepsilon_{i,t,s} \quad (1)$$

where $H_{i,t}$ is the value of the house *i* at time *t*, $H_{MSA,t}$ is the HPI of the surrounding MSA at time *t*, and $\varepsilon_{i,t,s}$ is the deviation of the growth rates of the house *i* and its surrounding MSA index from time *s* to *t*. Similarly, the growth rate implied by the MSA index relative to the national average forecast can be decomposed as follows:

$$\ln(G_{MSA,t,s}) = \ln(H_{MSA,t}) - \ln(H_{MSA,s}) = \ln(H_{N,t}) - \ln(H_{N,s}) + \varepsilon_{MSA,t,s} \quad (2)$$

where $H_{N,t}$ is the national HPI at time t, and $\varepsilon_{MSA,t,s}$ is the deviation of the growth rates of the MSA index and the national index from time *s* to *t*. Intuitively, one can think of the growth rate of a particular MSA as the national average growth rate plus a deviation term ($\varepsilon_{MSA,t,s}$). Similarly, the growth rate of a particular house is equal to the MSA-level average growth rate plus another deviation term ($\varepsilon_{i,t,s}$).

Plugging equation (2) into equation (1), we find that the individual house price growth rate equals the national average HPI growth rate and the sum of the dispersions of individual property around MSA levels, and the specific MSA levels around the national average growth rate:

$$\ln(G_{i,t,s}) = \ln(H_{N,t}) - \ln(H_{N,s}) + \varepsilon_{i,t,s} + \varepsilon_{MSA,t,s} \tag{3}$$

Notice that the variance of the first component of dispersion error given by $\varepsilon_{i,t,s}$ can be computed directly from the "*a*" and "*b*" parameters estimated by OFHEO using the three-stage weighted-repeat-sales methodology:

$$E\left(\sigma^2\left(\ln G_{i,t,s} - \ln G_{MSA,t,s}\right)\right) = E\left(\sigma^2\left(\varepsilon_{i,t,s}\right)\right) = a \cdot (t-s) + b \cdot (t-s)^2 \tag{4}$$

where E(•) is the expectation operator, (*t* - *s*) is the number of quarters since the loan origination time *s*. We estimated the variance of the second component error $\varepsilon_{MSA,t,s}$ for the dispersion of the MSA index growth rate around the national index growth rate forecast by a linear regression in the following form:

$$\sigma^2\left(\ln G_{MSA,t,s} - \ln G_{N,t,s}\right) = \sigma^2\left(\varepsilon_{MSA,t,s}\right) = c \cdot (t-s) + e \tag{5}$$

where e is the residual of the regression. Because equation (4) was estimated by OFHEO as a residual term when estimating the MSA HPI using all houses within that location, the individual property diffusion term must be orthogonal to the diffusion term between the MSA and the national HPIs. That is, the noise term $\varepsilon_{i,t,s}$ is independent of $\varepsilon_{MSA,t,s}$, or:

$$\rho\left(\varepsilon_{\mathrm{i,t,s}}, \varepsilon_{\mathrm{MSA,t,s}}\right) = 0 \tag{6}$$

This implies the following model for the variance of individual house price cumulative appreciation rates around the national average forecast:

$$E\left(\sigma^2\left(\ln G_{i,t,s}\right)\right) = a\cdot(t-s) + b\cdot(t-s)^2 + c\cdot(t-s) \tag{7}$$

The parameter "*c*" required for projecting the additional dispersion of the MSA index around the national average forecast was estimated as follows. For each quarter *t* we computed the cross-sectional (across MSAs) dispersion variance (MSA versus national) for each possible value of $(t-s)>0$, which corresponds to time since loan origination, *i.e.*, mortgage age. We then computed the average dispersion variance according to each age of loan:

$$\sigma^2(\bar{\varepsilon}_{t-s}) = \frac{1}{N}\sum_{\{\mathrm{MSA}\}}\left[\ln(G_{MSA,t,s}/G_{N,t,s})\right]^2 \quad t = s+1, s+2, \ldots, T \tag{8}$$

where *N* is the number of MSAs used in the estimation sample, and *T* is the most recent quarter that HPIs are available (first quarter of 2007 in this Review). This gives us a cross-section/time-series sample of average MSA index dispersion variance around the national average forecast that we assume is a linear function of (*t* - *s*):

$$\sigma^2(\bar{\varepsilon}_{t-s}) = c\cdot(t-s) + u_{t-s} \qquad t = s+1, s+2, \ldots, T; \quad s = 0,1,\ldots; s<t \tag{9}$$

where u_{t-s} is the regression residual. We estimated the unknown parameter "*c*" using a weighted least square regression with the number of average variance observations at each value of *t* - *s* as weights. The estimated quarterly standard deviations ($\sqrt{c}$) at age one, i.e., (*t* - *s*) = 1, were 2.77 percent for MSA indexes and 2.72 percent for Census division indexes.

One of the following two formulas was applied depending on whether the time period was historical or future:

$$\begin{cases} E\left(\ln G_{i,t,s}\right) = \ln G_{MSA,t,s} \\ \sigma^2\left(\ln G_{i,t,s}\right) = a\cdot(t-s) + b\cdot(t-s)^2 \end{cases} \qquad \text{if } s < t \le T \tag{10a}$$

$$\begin{cases} E\left(\ln G_{i,t,s}\right) = \ln G_{MSA,T,t} + \ln G_{N,t,T} \\ \sigma^2\left(\ln G_{i,T,s}\right) = a\cdot(t\text{-}s) + b\cdot(t\text{-}s)^2 + c\cdot(t-T) \end{cases} \qquad \text{if } s < T \le t \tag{10b}$$

Equations (10a) were applied to historical sample time periods when either an MSAD or CBSA index was used to update expected housing values, and equations (10b) were applied during future simulated time periods when the national average forecast was used to update expected housing values.

For future loan originations only a single formula set is required:

$$\begin{cases} E\left(\ln G_{i,t,s}\right) = \ln G_{N,t,s} \\ \sigma^2\left(\ln G_{i,t,s}\right) = a \cdot (t\text{-}s) + b \cdot (t\text{-}s)^2 + c \cdot (t - s) \end{cases} \quad \text{if } T \le s < t \tag{11}$$

Equations (11) were applied to future loan originations when only the national average forecast was used to update the expected housing values.

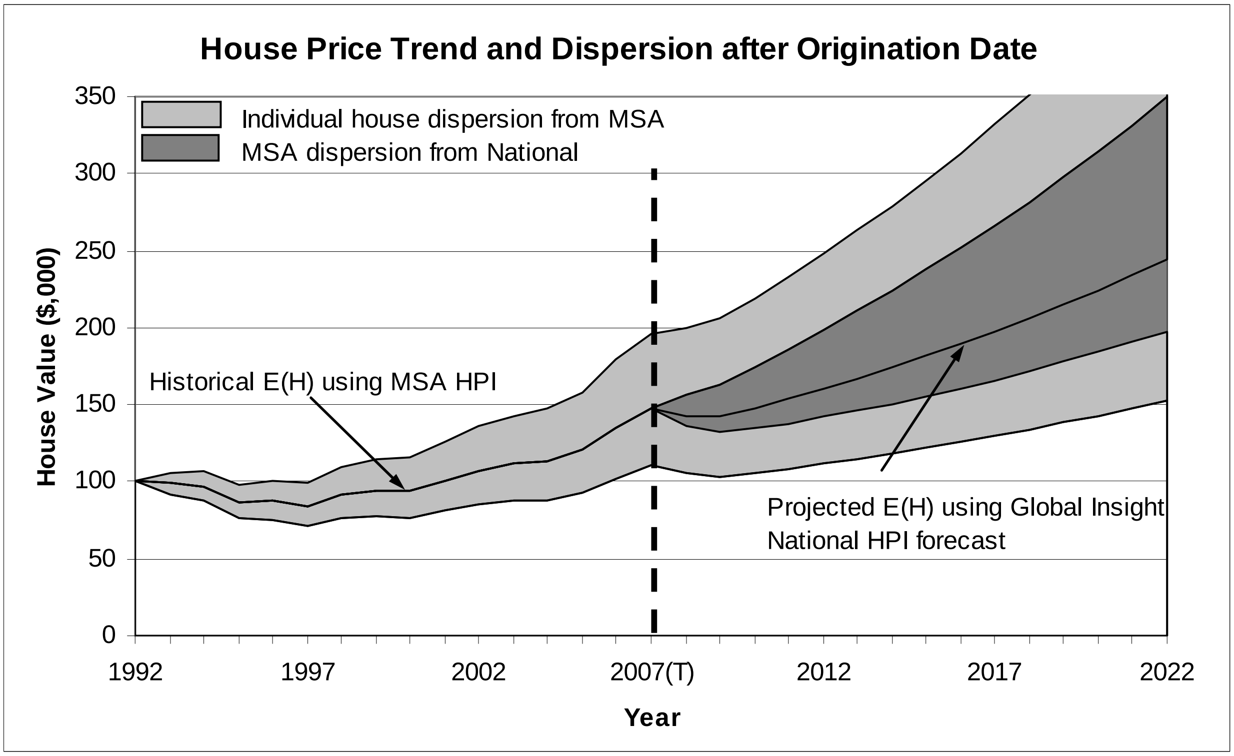
The additional term associated with the dispersion of the growth rate of an MSA HPI around the growth rate of the national HPI increases the overall dispersion volatility and results in higher probabilities of negative equity. This is counterbalanced by the reduced relative frequency of low expected HPI values when using a national average house price forecast instead of the more volatile metropolitan area HPIs.

IV. A Numerical Example

Exhibit C-1 presents an example to depict the expected value and the dispersion of the growth of the value of a house [E(H)] over time. Let's assume that the sales price of an underlying house of a mortgage loan originated in a particular MSA at the first quarter of FY 1992 (s) was $100,000. We further assume that the last historical HPIs available are as of the first quarter of FY 2007 (T).

The expected value and the dispersion standard deviation of this house at each exposure year (t) can be computed following equations 10b (as shown in section III). The center line in Exhibit C-1 demonstrates the expected value of the house at each exposure year. This expected value was updated by using the OFHEO HPI for the particular MSA up to T. After T, the future expected value of the house is then updated according to the national house price growth rate forecasted by Global Insight, Inc. since a forecast of the HPI growth rate for the particular MSA was not provided.

Exhibit C-1



The house price growth rate of an MSA could be higher or lower than the national average for the next 15 years. The red shaded area represents the possible future value of the house if the MSA growth rate is one standard deviation above or below the national average growth rates. Note that, prior to *T*, such MSA-level dispersion did not exist in all historical distributions since the expected house price was updated by using the MSA-specific index.

The gold shaded area represents the additional dispersion of one standard deviation due to the individual house-level volatility around the MSA average. The size of this source of dispersion starts immediately after the house was sold in the first quarter of 1992 and continues to grow over the age of the mortgage loan. The OFHEO dispersion volatility parameters were used in computing this source of dispersion.

As a result, the outer boundary of the shaded area in Exhibit C-1 provides a visual sense of the dynamics of the house price distribution over the life of the mortgage. The probability of negative equity variable is directly computed as the probability that the house price may fall below the UPB of the mortgage at any point in time.