

Appendix C: Data for Loan Performance Simulation

This appendix describes technical details of the application of the estimated econometric models to produce forecasts of future loan performance. We begin by describing how loan event data for future time periods were generated for use in projecting 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 future 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 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 indexes 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 (FY 1975 Q1 to FY 2004 Q2) the process of building the period-by-period event data was continued for the scheduled remaining term of the loan. The key difference in this case is that forecasted values of the external economic data are required. Otherwise, the data creation process followed the same steps.

In addition, we were required to forecast the future loan performance of future FHA loan books originated through FY 2011. We were required to base our actuarial projections for future loan books on estimates of total loan volume of FY 2004 through FY 2011 books of business 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 2003 Q2-Q4 and FY 2004 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 corresponding quarters of FY 2004 to FY 2011, except that the proportion between the purchase and refinance loans are allocated based on FHA's specific forecasts. We then updated the initial values of economic variables for the future loan originations based on the corresponding economic forecasts to reflect conditions at the time of origination. Following projection of 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 strata 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. Although we have discussed the development of historical and future loan event data as separate processes, in practice we undertook an integrated approach whereby both the historical and future loan event data were generated at the same time and stored in the same database. This greatly simplified the projection of future conditional claim and prepayment probabilities by taking advantage of statistics software's internal routines for prediction following estimation. Once a logit model has been estimated, the program can automatically generate predicted values, in this case predicted logit probabilities, for all records in the data, regardless of whether they were used in estimation. Additional fields are automatically added to the dataset that contain the predicted logit probabilities for all records with non-missing values of the explanatory variable fields used in estimation. Automatic generation of predicted probabilities in this manner greatly reduced the chances for human error in computing the predicted probabilities.

II. Future Economic Forecasts

FHA provided TAC with quarterly forecast data from Global Insights, Inc. These data served as the source of future interest rates and appreciation in housing values in our analysis.

For the projecting future changes in housing values, TAC used the Global Insights' national average sales price of existing single-family homes forecast. 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 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 TAC 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 section 3 of this Appendix C. To summarize, estimates of additional dispersion (volatility) were combined with those for the underlying regional or metropolitan index to increase the overall estimated volatility of the appreciation rate of a specific house. The additional volatility is added only from the start of the future forecast of loan performance, that is, at the point that the computation of the probability of negative equity switches to using the national average house price forecast. The additional volatility increases with time in a similar manner to the underlying regional volatility estimate.

Recall that the source of house price appreciation rates for historical loans was house price indexes (HPIs) published by OFHEO. We applied OFHEO MSA-level HPIs for a subset of 43 representative MSAs. For all other loans outside of these MSAs, we applied one of the nine OFHEO Census division HPIs depending on the location of the mortgage.

As described in Appendix A, the indexes are used in conjunction with estimates of house price diffusion volatility 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. Estimation of Dispersion for the Regional House Prices

The Global Insights national average house price data and forecast do not provide estimates of diffusion volatility. Although OFHEO publishes a national-level HPI, this is based on a weighted average of indexes for the nine Census divisions, and no separate volatility parameters are published at the national level. Thus, direct estimate of house price dispersion around a national index does not exist. Since TAC used regional-level and metropolitan-level indexes for estimating historical loan performance models, we elected to adopt an approach that would build upon on the OFHEO local house price volatility estimates and modify it appropriately for the forecasting period.

For consistency with the OFHEO estimates of house price volatility, we required estimates of the variance in the geometric growth rates of housing values implied by the regional indexes around the geometric growth rates implied by 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:

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

Similarly, the growth rate implied by the MSA index relative to the national average forecast can be decomposed as follows:

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

Therefore, relative to the national average forecast, the individual house price growth rate equals the growth rate implied by the national index and the sum of the dispersion of individual around MSA growth rates and the specific MSA around national average growth rates:

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

Recall 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 published by OFHEO:

$$\sigma^2(\ln G_i / \ln G_{MSA}) = \sigma^2(\varepsilon_{i,t,s}) = a \cdot (t-s) + b \cdot (t-s)^2 \quad (4)$$

where $(t-s)$ is the number of quarters since loan origination. For consistency with the OFHEO formulation we required an estimate of the variance in the second component error $\varepsilon_{MSA,t,s}$ for the dispersion of the MSA indexes around the national average forecast that would also be linear in time, as follows:

$$\sigma^2(\ln G_{MSA} / \ln G_N) = \sigma^2(\varepsilon_{MSA,t,s}) = c \cdot (t-s) \quad (5)$$

Because equation (4) was estimated by OFHEO as residual term when estimate the MSA house price index using all houses within that location, it must be orthogonal to the volatility of the MSA level house price indexes. That is the noise term $\varepsilon_{i,t,s}$ is independent of the the $\varepsilon_{MSA,t,s}$, or:

$$\rho(\varepsilon_{i,t,s} + \varepsilon_{MSA,t,s}) = 0 \quad (6)$$

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

$$\sigma^2(\ln G_i / \ln G_N) = a \cdot (t-s) + b \cdot (t-s)^2 + c \cdot (t-s) \quad (7)$$

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 MSA) average dispersion variance (MSA versus national) for each possible value of $(t-s) > 0$, which corresponds to time since loan origination, *i.e.*, mortgage age:

$$\sigma^2(\bar{\varepsilon}_{t,s}) = \frac{1}{n_{MSA}} \sum_{MSA} [\ln(G_{MSA,t,s} / G_{N,t,s})]^2 \quad t = 2, 3, \dots, T; \quad s < t \quad (8)$$

This gives us a cross-section/time-series sample of average MSA index dispersion variance around the national average forecast which we assume is a linear function of $t-s$:

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

We estimated the unknown parameter “ c ” using a weighted least square regression using the number of average variance observations at each value of $t-s$ as weights. The estimated quarterly standard deviation (\sqrt{c}) values were 2.7682% for MSA indexes and 2.7192% for Census division indexes.

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

$$\sigma^2(\ln G_i / \ln G_{MSA}) = a \cdot (t-s) + b \cdot (t-s)^2 \quad \text{if } t \leq T \quad (10a)$$

$$\sigma^2(\ln G_i / \ln G_N) = a \cdot (t-s) + b \cdot (t-s)^2 + c \cdot (t-T) \quad \text{if } t > T \quad (10b)$$

where T is the last historical time period (FY 2004 Q1). Equation (10a) was applied to historical sample time periods when the MSA index was used to update expected housing values; and equation (10b) was applied during future time periods when the national average forecast was used to update expected housing values.

For future loan originations only a single formula is required:

$$\sigma^2(\ln G_i / \ln G_N) = a \cdot (t-s) + b \cdot (t-s)^2 + c \cdot (t-s) \quad \text{if } s > T \quad (11)$$

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

The additional term associated with dispersion of an MSA or Census division index around the national average forecast increases the overall dispersion volatility and results in higher probabilities of negative equity. This is counterbalanced by reduced relative frequency of low expected HPI values when using a national average house price forecast instead of the more volatile local or regional indexes.