U.S. Department of Housing and Urban Development

Office Of Policy Development & Research

PD&R

A Study of Market Sector Overlap and Mortgage Lending

A Study of Market Sector

Overlap and Mortgage Lending

Acknowledgments

PREFACE

In 1992 legislation, Congress mandated that HUD provide information on the affordable lending activities of Fannie Mae and Freddie Mac, the two major government-sponsored enterprises (GSEs) in the mortgage market. At that time, Congress said there was a "vacuum of information" surrounding the GSEs' mortgage purchases, and called for HUD to collect data and conduct research on the GSEs' mortgage purchase activity. One priority for HUD has been to initiate an active research program with respect to the GSEs, including both in-house and contract research. Much of this research was used in the development of the three Affordable Housing Goal Regulations issued in 1995, 2000, and 2004.

This report continues this line of GSE research by comparing the characteristics of GSE-purchased loans with mortgages originated or insured by other sectors of the mortgage market, such as the Federal Housing Administration (FHA). This so-called overlap analysis clarifies the role of the GSEs in providing credit support for low-income and minority families. Intuitively, overlap refers to the set of loans that could have gone to either market sector, for example, either insured by FHA or sold to the GSEs.

The overlap question is one of great interest in mortgage policy discussions but has not been satisfactorily dealt with because of the lack of a database covering FHA-insured, GSE-purchased, and other conventional loans that includes borrower credit history and other mortgage loan data, such as the loan-to-value ratio. Thus, this study extends previous research by including data on borrower credit history, loan-to-value ratios, and other underwriting variables for conventional loans, privately-insured loans, GSE-purchased loans, and FHA-insured loans. It examines the extent of overlap between the mortgage market sectors, particularly the FHA and the two GSEs, Fannie Mae and Freddie Mac. The study finds that different market sectors serve distinct segments of the population. For example, one of the main findings of this study is that only about ten percent of FHA-insured loans have risk characteristics similar to GSE-purchased loans. Compared with GSE-purchased loans, FHA-insured loans are characterized by lower borrower credit scores and higher loan-to-value ratios (i.e., lower downpayments), and are more targeted to lower-income and minority borrowers.

Table of Contents

List of Exhibits

Appendix Exhibits

Executive Summary

A Study of Market Sector Overlap and Mortgage Lending

This study examined the extent of overlap between the mortgage market sectors, particularly the Federal Housing Administration (FHA) and the two government-sponsored enterprises (GSEs), Fannie Mae and Freddie Mac. Intuitively, overlap refers to the set of loans that could have gone to either market, for example, either insured by FHA or sold to the GSEs. Home Mortgage Disclosure Act (HMDA) data were supplemented with Experian data to provide FICO credit scores and house values (used to calculate loan-to-value (LTV) ratios). The years covered were 1998 to 2000. There were 11 MSAs for which the match rate between Experian and HMDA loans was at least 56 percent. The matched, loan-level data were divided by mortgage market sector and tabulated to compare borrower, loan, property and neighborhood characteristics. For example, GSE loans have the highest average FICO score (726) followed by private mortgage insurance (PMI) loans (712), depositories (699), FHA (643) and subprime (637). The share of loans in low-income areas (tract income below 90 percent of median income) is nearly the reverse order: subprime (35 percent), FHA (33 percent), depositories (25 percent), PMI (21 percent) and GSE (16 percent).

In order to determine which loans were selected for each mortgage sector, a series of origination models were estimated. For the choice between FHA and GSE, the most important factors were LTV, FICO, payment-to-income ratio and borrower race/ethnicity. The model produces a predicted probability that a particular loan will be FHA insured; this predicted probability conveniently compresses all the credit and non-credit factors into a single dimension. The distribution of predictions for FHA loans was overlaid on the distribution of predictions for GSE loans and the overlap was measured to be 11 percent of the combined set of loans. In other words, 11 percent of the loans in the combined FHA/GSE market have very similar characteristics, such that the model could not distinguish whether they were FHA or GSE loans. In terms of the FHA portfolio and depending on the overlap methodology, between 10 and 14 percent of FHA loans fall in the overlap region. In other words, 10-14 percent of FHA loans have characteristics that are similar to GSE loans. On examination, there are minor differences in income, FICO and LTV among the overlap loans, but overall the FHA loans in the overlap region look remarkably similar to the GSE loans. It appears that the FHA overlap loans were as qualified as many GSE-purchased loans.

A default model was also estimated using separate FHA performance data (that included FICO scores) and then applied to the Experian/HMDA matched data. The coefficients from the default model were used to assign a risk score to the matched loans. This risk score was highly correlated with the FICO credit score, but the risk score alone could not explain the choice of FHA vs. GSE loans. The origination model with its full complement of variables (including the FICO score) does a better job of assigning loans between the market sectors. Still, with respect to overlap, virtually the same results are obtained when the risk score replaces the FICO score in the origination model.

Two methods of overlap measurement were tried. The confidence interval method determines the boundaries of the overlap region based on the 95 percent confidence interval around each loan's predicted probability. If that confidence interval does not include either 0 or 1 (for example, prediction not clearly GSE or FHA), the loan falls in the overlap region. The confidence interval approach has the advantage of being intuitive, in that, when the model cannot determine with 95

percent confidence that the loan is either FHA or GSE, it falls in the overlap region. As explained in the text, there are two issues with the confidence interval method. First, imprecise models with wide confidence intervals create very narrow overlap regions. Second, and more problematic, the overlap region may be distorted by the extreme values of the marginal distribution that do not stand out relative to the combined market distribution. In other words, some high risk GSE loans may look very similar to FHA loans, but may not be representative of the typical underwriting for GSE loans. The tolerance interval method solves the problem of outliers by trimming off the top and bottom five percent for each market participant. A parametric version of the tolerance interval is well-suited when the individual distributions are normal (Gaussian). The preferred method for measuring the overlap of non-normal distributions is the non-parametric tolerance interval method based on order statistics. This method, which is used below, trims the outliers and does not assume normality.

The following table summarizes the measurements of overlap regions for pairs of mortgage market sectors. For each pair, the first row provides the combined overlap (share of similar loans relative to the market or combined sample). The second row gives one participant's share (typically FHA) of the overlap as a percentage of the combined sample. The third row presents FHA's overlap loans as a percentage of the total number of FHA loans. For example, in the FHA vs. GSE market using FHA-eligible loans with LTV between 80 and 100 percent, the combined overlap by the confidence interval method is 20 percent — made up of 9 percent FHA loans and 11 percent GSE loans. The FHA overlap loans are 14 percent of the FHA portfolio. When measured by the non-parametric tolerance interval method, FHA overlap loans are 10 percent of the FHA portfolio. The summary table also provides overlap measurements between FHA and other market sectors. For example, 15 percent of FHA-insured loans are similar in characteristics to privately-insured (PMI) loans.

The implication of this research is that about 10 percent of FHA borrowers have risk characteristics similar to GSE borrowers. It appears that these FHA borrowers could qualify for conventional loans. The measures of overlap presented in this report can serve as a baseline for comparison over time. The GSEs have increased their purchases of LTV loans above 95 percent since HUD conducted the first GSE-FHA overlap study in 1995. In addition, recent GSE commitments to buying more subprime loans indicate there will likely be increased overlap between the FHA and GSE markets in the future.

Summary Table of Overlap Regions

Overlap in Originations		
FHA vs. GSE		
Combined Overlap	20%[a]	11%
FHA Share of Overlap	9%[b]	7%
FHA Overlap loans rel. to FHA Distribution	14%[c]	10%
FHA vs. PMI within GSE		
Combined Overlap		10%
FHA Share of Overlap		7%
FHA Overlap loans rel. to FHA Distribution		9%
FHA vs. all PMI		
Combined Overlap		18%
FHA Share of Overlap		10%
FHA Overlap loans rel. to FHA Distribution		15%
FHA vs. Subprime		
Combined Overlap		13%
FHA Share of Overlap		9%
FHA Overlap loans rel. to FHA Distribution		13%[d]
GSE vs. Depositories		
Combined Overlap		78%
Dep. Share of Overlap		37%
Dep. Overlap loans rel. to Depository Portfolio		73%

[a] Interpreted as follows: 20 percent of the FHA-eligible loans from the combined FHA and GSE distributions fall in the overlap region.

[b] Interpreted as follows: 9 out of the 20 percentage points in the combined overlap are FHA loans and the remaining 11 percentage points are GSE loans.

[c] Interpreted as follows: 14 percent of the FHA distribution of loans (not the combined set, but just the FHA loans) are in the overlap region.

[d] The overlap between FHA and subprime is smaller than expected because the LTVs for FHA loans (average 97 percent) are much higher than for subprime purchases (average 81 percent).

Section 1: Background and Literature Review

This document is the Final Report for HUD Contract C-OPC-21895, Task Order CHI-T0003. The purpose of this research is to investigate the extent of overlap between the mortgage market sectors, especially between the Federal Housing Administration (FHA) and the Government Sponsored Enterprises[1] (GSE) loans. Loan level data matched under a previous contract were used to estimate the degree of overlap among home purchase mortgage originations in 11 metro areas. Overlap is defined as home purchase loans that could have gone to either the FHA or the GSE sector. A second component of the research is to estimate a default risk score, based on FHA loan performance, and apply that model to a large set of Home Mortgage Disclosure Act (HMDA) loans reported between 1998 and 2000.

This study finds that about 11 percent[2] of the combined FHA-eligible GSE and FHA loans with loan-to-value (LTV) ratios between 80 and 100 percent are in the overlap region. In terms of the FHA portfolio, 10 percent of the FHA loans fall in the overlap region. The main implication from this work is that, although it remains rather modest, the overlap between the FHA and GSE mortgage sectors has been increasing as the GSEs strive to meet their housing goals. With more flexible underwriting, made possible by automated underwriting models, the GSEs are more likely to compete with FHA for high loan-to-value (LTV) and high payment-to-income (PTI) loans. As the GSE housing goals are increased, requiring the GSEs to purchase a higher share of loans from low-income and minority borrowers would be expected to increase the overlap between the sectors. With increased competition from the GSEs, FHA lenders may seek to preserve market share by competing with the subprime market for qualified loans. Overall, the increased competition should benefit consumers who will have more choice, if they are willing to shop for the best loan terms.

The Final Report is organized as follows. Section 1 contains the background and a brief literature review that motivates the research questions for the remaining sections. Section 2 describes the data sources and preparation. Home Mortgage Disclosure Act (HMDA) data include information on the loan amount, the income, race and ethnicity of the borrower, and the census tract location of the newly-mortgaged property. HMDA data do not contain credit scores, which are needed to assess the risk of the loan. HMDA also omits house values, which means an important measure of equity, the loan-to-value ratio, cannot be calculated. HUD purchased from Experian data that included credit scores and house values on over 1 million home purchase records from 24 MSAs that were matched to HMDA originations for the origination years 1998 to 2000. The match rates were not uniform across MSAs. The 11 MSAs with the best match rates were used,[3] and Section 2 describes how the loans were selected for the subsequent analysis.

---

[1]   In this document, the GSEs refer to Fannie Mae and Freddie Mac. The Federal Home Loan Banks are excluded. It is highly likely that some GSE loans are not conventional in terms of prime risk quality. As shown in Exhibit 7, there are 1,116 GSE loans that are also designated as subprime. The GSE categorization is based on purchaser type 1 or 3 in the HMDA data.

[2]   The 11 percent overlap is based on the preferred overlap method of non-parametric tolerance limits explained in Section 3. The corresponding amount from the confidence interval method is 20 percent, but the tolerance limits method has a stronger theoretical foundation and is less sensitive to model fit.

[3]   For the 11 included MSAs, the match rate ranged from 0.56 in Baltimore to 0.66 in Portland. The other MSAs are: Chicago, Cleveland, Denver, Los Angeles, Oakland, Philadelphia St. Louis, Tampa and Washington, DC.

The overlap is measured in several different ways and most of the approaches begin with an origination model presented in Section 3.[4] The origination model is a logistic regression in which the dependent variable is the probability of a loan being insured by FHA relative to the alternative of being sold to the GSEs. Other combinations, such as FHA vs. private mortgage insurance (PMI), FHA vs. subprime, and GSE vs. depository lenders, are also estimated. Taking the initial example of FHA vs. GSE, the overlap region is the subset of home purchase loans for which the 95 percent confidence interval around the predicted probability does not include either 0 or 1. In other words, if the model prediction is not 95 percent certain that the loan is insured by FHA or sold to a GSE, then that loan is defined to be in the overlap region. An alternative definition of overlap is derived from tolerance intervals in which the overlap is the set of loans between the lower limit of the FHA distribution and the upper limit of the GSE distribution. Both parametric and nonparametric methods for setting the tolerance intervals are described in detail.

Overlap can also be viewed in terms of risk of default, as presented in Section 4. In this case, Unicon estimated the probability of claim after 3 years from origination using FHA performance data. From that default model, a default risk score could be predicted for every loan in the matched data. The distribution of FHA loans is shifted to the right along the risk scale relative to the GSE distribution, but there is considerable overlap between the distributions. This suggests that a lot of FHA loans are no riskier than the loans purchased by the GSEs. However, predictions of FHA vs. GSE status based only on the risk score are not very accurate. If the risk score is used in place of the credit score in the origination model, the predictions are virtually the same as with the credit score and the overlap is back to the values in Section 3. Thus, the risk score is a good proxy for the credit score, but apparently non-credit factors are important in determining whether a loan is insured by FHA or sold to the GSEs. Non-credit factors (e.g., neighborhood factors such as center city location and average family income in the census tract) contribute significantly to choice of mortgage sector. These neighborhood variables may reflect the prospect for future property value appreciation.

Although the matched data have only 3 years of originations and 11 MSAs, the loan level data can be organized by census tract. Section 5 contains estimates of the variation in FHA market share across 4,240 tracts representing neighborhood housing markets. Much of the explanatory power of the models comes from the separate indicators for each MSA, but within an MSA the percent minority in the census tract, the percentage of household heads aged 15 to 24 years, the average FHA default rate of the census tract, and the median household income in the census tract are positively related to FHA market share. The median house value of the census tract and the percentage of elderly owners in the census tract tend to reduce FHA market share. These are tentative results that suggest more work needs to be done to fully explain the variation in FHA and GSE market shares.

The Appendix contains more extensive tables with variable definitions and calculation methods along with separate tabulations for each MSA.

---

[4] The GSEs do not originate loans, but rather purchase loans originated by the primary market lenders. FHA does not originate loans either, but rather insures loans originated by lenders. The text refers to the origination model to designate that the information about the loan comes from the origination and to distinguish it from the subsequent default model, where the designation is based on FHA claims.

Literature Review

1995 HUD Study. The baseline for this analysis on market overlap comes from a 1995 HUD report titled *An Analysis of FHA's Single-Family Insurance Program*. In that report overlap between FHA and PMI was measured as the loans with LTV between 80 and 95 percent with loan amounts below the FHA loan limits. Loans with LTV below 80 percent did not need private mortgage insurance and loans with LTV above 95 percent could not qualify for private mortgage insurance (at least, at that time). However, most FHA loans had LTV ratios above 95 percent. Moreover, conventional loans typically had to have a payment-to-income ratio below 28 percent and a debt-to-income (DTI) ratio below 36 percent. The corresponding guidelines for FHA were PTI of 29 percent and DTI of 41 percent, but there was considerable flexibility that allowed loans to exceed those guidelines. Given the difference in underwriting guidelines and the strictness of PMI guidelines at that time, most FHA loans did not qualify as conventional loans. However, about 1/3 of FHA loans did have LTV between 80 and 95 percent and thus were considered roughly comparable to conventional loans under the FHA loan limits. Overlap, in the 1995 HUD study, was based on the share of FHA loans that met conventional lending guidelines for LTV and PTI. In that sense, the study selected loans with the potential for overlap. The authors did not predict what subset of those FHA loans had the combination of characteristics that made them indistinguishable from conventional loans, at least for a statistical model's point of view.

In HUD's 1995 study, a linear probability model was estimated on the potential overlap loans to determine which characteristics affected the probability of the loan becoming insured by FHA versus PMI. The factors that increased the probability of a loan being FHA-insured were: payment-to-income ratio, first-time homebuyer, black race of borrower, Hispanic ethnicity of borrower, center city location of property, tract median family income relative to MSA median family income, and the percent minority households in the tract. Factors that decreased the probability of a loan being FHA-insured were: borrower age, loan amount relative to FHA loan limit, LTV, borrower income relative to MSA median family income, and other race of borrower (white being the reference group). Given that high LTV is a distinguishing feature of FHA loans, it is a little surprising that LTV had a negative effect on the probability of FHA. However, the selection of potential overlap loans excluded the high (over 95 percent) LTV loans in FHA, but included the mass of conventional loans at the maximum LTV of 95 percent. A further subdivision of loans into high, medium, and low cost areas based on FHA loan limits, showed that the coefficients on tract income and center city location became negative in the high cost areas. Unfortunately, the data on 1993 originations (used in the 1995 study) did not include credit scores.

The implication of the HUD report was that there existed little overlap between the FHA and conventional mortgage markets. FHA tended to serve low-income, minority, young and first-time homebuyers. Rather than treating FHA and conventional as separate markets, HUD promoted increased purchases of loans from low-income and minority borrowers through the GSE housing goals.

Other Studies. The following papers did not measure overlap, but they did provide valuable information about the specifications for loan choice models (including neighborhood effects) and they developed a theory about the FHA market sector relative to the conventional market.

Berkovec, Canner, Gabriel, and Hannan (1998) wrote a well-known paper on discrimination in which they used 1987-89 FHA performance data to test whether minorities have lower default rates than non-minorities. The theory, based on Becker (1971), was that minorities would have to meet a higher standard than whites for their loans to be accepted by a discriminating lender. The authors found that minorities have a higher probability of default, indicating that lenders did not discriminate. This study has some similarities to Berkovec et al. in that a logistic regression is estimated with FHA data and the impact of market concentration is measured using Herfindahl-Hirschmann indices. However, this study does not focus on discrimination, but rather the overlap between the FHA and conventional originations. In addition, the data are more recent (1998-2000) and include FICO scores, which were not available for the earlier discrimination study.

A subsequent study by Cotterman (2002) replicated the Berkovec et al. (1998) analyses on FHA data from a more recent time period, but also included credit scores in parallel analyses. When credit scores were introduced, the estimated minority coefficients tended to fall, often becoming statistically insignificant and sometimes changing sign. When credit scores were included, the empirical results no longer gave unambiguous support for the notion that lenders do not discriminate. The change in results also suggests that the original work suffered from omitted variable bias, despite considerable efforts to control for omitted variables.

In 1998, Onder investigated the neighborhood factors affecting FHA market share with a two-stage model. In the first stage the probability of a loan being FHA is regressed on loan level characteristics and a set of tract dummies. In the second stage, the coefficients on the tract dummies are regressed on a set of tract characteristics. The idea is to determine which characteristics are associated with the sign and size of the tract fixed effect. The research showed that minority composition was not significant, and there was a negative relation between tract median family income and the likelihood of being FHA-insured. Interestingly, census tract income had a positive effect on FHA for values below $30,000 and negative effect above $30,000. Also, although the level of minority composition was not significant, an upward change in minority share greater than 15 percent over the previous decade was a positive factor in a loan being insured by FHA. High rent levels had a negative relation to FHA, but rent increases were positively associated with FHA. Similarly, high vacancy rate had a negative relation to FHA, but vacancy rate increases had a positive association to FHA. On a national pool of 35,464 tracts, the R-square was 0.39, which jumped to 0.72 when the specification included 333 MSA dummies. Apparently, there remained significant differences between MSAs even after controlling for an extensive list of individual and neighborhood effects.

Pennington-Cross and Nichols (2000) filled in a gap left by earlier research by including new data on credit history. A national loan level sample included originations from 1995 and 1996 and represented 306 MSAs. They showed that there was a considerable overlap between FHA and conventional loans in terms of the distribution of credit scores. Their research also showed that the credit score was an important ingredient in loan choice. The probability of a loan being FHA declined with higher credit scores.

Using the same data, Ambrose and Pennington-Cross (2000) estimated an FHA market share equation using logistic regression. Concerned that LTV may be jointly chosen with FHA, the researchers estimated an instrumental variable equation as a first stage and used the predicted LTV in the FHA market share equation. The coefficient results showed the following metropolitan area factors had a positive association with FHA market share: unemployment rate, segregation of blacks, percent underserved, and FHA loan limit relative to the median house price for the MSA. The negative

coefficients were for: 1-year and 10-year house price change, annual volatility in house prices, and higher minority share. The unexpected result on minority share may be because underserved is also included in the specification and minority share is an important component in underserved status. From these results, the authors concluded that FHA market share was higher in cities with greater economic risk characteristics. GSE purchase rates were fairly insensitive to local economic conditions.

Ambrose, Pennington-Cross, and Yezer (2002) provided a more detailed theoretical explanation of the interface between FHA and conventional market sectors. They assumed that loans could be ordered by a single risk factor and conventional underwriting determined the upper limit on acceptable risk. FHA has more lenient underwriting standards, so the higher risk loans rejected by conventional lenders may be acceptable to FHA. They call this the FHA wedge. In this view, the amount of overlap between the markets is quite small. A few loans may go to FHA that could have qualified for conventional lending, but these are basically a mistake due to insufficient shopping by the borrower or steering by the lender. The reason it is considered a mistake is that mortgage insurance for conventional loans is less expensive than FHA mortgage insurance (at least it is for loans with an LTV ratio less than or equal to 95 percent). FHA charges a higher premium that corresponds to the higher risk and claim rate for most loans in the FHA wedge. So, if a loan could qualify as a conventional loan, it would be less expensive for the borrower to have a conventional loan than an FHA loan. Historically, private mortgage insurance was not available for such high-risk loans, though in recent years the insurance has become available at a higher rate than FHA charges. Thus, they conclude that the overlap between FHA and conventional loans comprises a small set of loans. If borrowers with low-risk FHA loans had conducted a more thorough search, they would have realized their mistake and pursued a conventional loan.

Building on the data and analysis in their 2000 paper, Ambrose, Pennington-Cross, and Yezer (2002) estimate an FHA market share model (at the metropolitan area level) with measures for cyclical risk and permanent risk. The cyclical risk factors include local unemployment rate and the percent change in delinquent bank loans. The permanent risk factors include volatility in house price appreciation, average default rate over the past six years, share of low-income households, and the percent of loans with loan amount relative to income greater than three. The findings showed positive coefficients for: change in unemployment rate, change in delinquency rate, average delinquency rate, volatility in house prices, share of incomes below $20,000, and percent black. The variables with negative coefficients were current and lagged house price change, loan-to-income greater than three, and black segregation (Gini coefficient). It is interesting to note that the house price volatility, black segregation, and percent black had reversed signs in the 2000 paper.

The authors conclude that conventional underwriting does not adjust to local risk factors in order to maintain market share. Rather, non-price credit rationing by conventional lenders leaves FHA with the role of maintaining the mortgage credit supply in declining housing markets. These effects from the 1995-1996 data may have been less apparent during the 1998-2000 period when housing markets were more uniformly strong. Indeed, both data sets relied on cross-sectional variation rather than a full cycle or major regional recession.

Freeman, Galster and Malega (2003) provide an in-depth empirical analysis of the secondary mortgage market impacts on underserved areas of Cleveland during 1993-1999. Based on single family home sales by census tract, the researchers found that secondary mortgage purchases, particularly by non-GSE buyers, had a positive effect on the number of sales transactions with a one

year lag. The increase in purchases did not affect sales prices, though there is some evidence that non-GSE purchases of refinances did boost prices one to two years later. Gyourko and Hu (2002) did a broader study in 20 major metropolitan areas and found a spatial mismatch between GSE purchases in low-income and minority areas and the demand for affordable housing. These studies indicate a modest degree of competition for loans from low-income and minority borrowers, but do not quantify the degree of overlap, i.e., how many loans could have qualified for GSE and non-GSE purchases.

To measure market overlap, this study follows the loan choice literature using logistic regression. However, an alternative approach using discriminant analysis is described by Amemiya (1985, pp. 281-285). The maximum likelihood approach is robust to non-normal covariates, though it may not be as efficient as discriminant analysis in some cases. This comparison is left for future research.

Recently there have been several media announcements from the GSEs that they intend to increase their purchases of loans from low-income and minority households, as well as subprime loans. For example, the National Mortgage News (Oct. 25, 2004) reported on an interview with the new CEO of Freddie Mac, Richard Syron, in which Mr. Syron is quoted as saying (p. 86), "I'd like to be more aggressive in the minority and Hispanic markets, yes. We will push. It's what we are supposed to be doing but it's good business." Regarding subprime, Mr. Syron said the Freddie Mac credit losses are about one percent on subprime loans and he feels the company can afford to take on more credit risk. Reported in Origination News (www.originationnews.com/plus/#4 on 10/21/2004), Eugene McQuade, Chief Operating Officer at Freddie Mac, announced at the America's Community Bankers convention in Washington, DC, that the company was simplifying it's A-minus loans and related low-downpayment products. At the same convention, Franklin Raines, Fannie Mae chairman and chief executive, declared that Fannie Mae intended to be more aggressive in serving the subprime market. He said the subprime market is estimated to be $323 billion and growing. Mr. Raines said, "We estimate that about half of the subprime borrowers have only slightly blemished credit and are just a notch away from qualifying for Fannie Mae's prime conventional financing." The main point is that as the GSEs become more aggressive about purchases of low-income and minority loans, it is highly likely that this will entail more overlap with FHA and subprime lenders that have traditionally served those borrowers.

Research Questions

Given this background from the literature, the goal of this study was to update and broaden the overlap findings from the literature. Previous work was updated by estimating FHA market share models on more recent data that included controls for credit scores. The results are broader because they included mortgage market sectors for GSEs, depository lenders, FHA, and subprime. A limited attempt at explaining FHA market shares based on tract level information was also attempted. The research questions were the following:

1) What are the borrower, loan, property and neighborhood characteristics associated with each mortgage market sector?

2) What factors determine the market shares captured by each mortgage sector?

3) Is there a significant degree of overlap between FHA and GSE sectors such that those loans could have gone to either FHA or GSE?

4) How much overlap is there in terms of default risk?

5) If the credit score is replaced by a default risk score in the origination model, does that attenuate the importance of non-credit factors?

Section 2: Data Preparation and Description

The data used in this analysis come from three sources: HMDA (1998-2000), Experian (1998-2000) and Census (1990). HMDA data provide nearly a complete set of loans for metropolitan areas.[5] Although HMDA data provide the loan amount, borrower race, tract location, and much more, it does not include credit scores or house values (needed to calculate LTV). To bridge this gap, loan record data were purchased from Experian for a select set of MSAs and HUD merged the Experian data with the HMDA data. Additional neighborhood information was obtained from the 1990 Census. This section describes the process of selecting, merging and cleaning the data along with tabulations of the data used in the origination models. More detailed information about individual variables or tabulations by MSA can be found in the Appendix.

HMDA Data

Depository and other financial institutions report their mortgage loan activity to the Federal Financial Institutions Examination Council (FFIEC), which makes a subset of the data available to the public for analysis.[6] The HMDA data comprise the most comprehensive source for mortgage lending information and HMDA data were used as the benchmark for weights. For each loan, there is information on loan type (especially conventional vs. FHA), loan purpose (home purchase, improvement, refinancing or multifamily dwelling), action taken (originated, denied, withdrawn) and type of purchaser (GSE, Ginnie Mae, commercial bank, etc.). For this research, newly-originated home purchase loans were selected. Besides a listing of loans, HMDA data provide the MSA and tract location, which makes it possible to merge in other data, particularly Census data, at the tract level. In addition, HMDA date include income and race/ethnicity of the borrower, which enabled this study to focus on low-income and minority borrowers.

Experian Data

Unfortunately, HMDA data do not include two pieces of information crucial to the assessment of risk and the underwriting process, namely credit score and house value. These data were obtained from a private vendor, Experian Information Solutions, Inc., under a previous HUD contract.[7][8] Twenty-four metropolitan areas were selected for their diversity of geographic location, housing appreciation rates, housing prices, and broad representation of the nation. Within each MSA, the census tracts were

---

[5] According to the HMDA website (http://www.ffiec.gov/hmda/default.htm), in 2000 all depository institutions with assets exceeding $30 million and a metropolitan office that originated a home purchase loan or refinancing secured by a single family home must report their loans. Other for-profit mortgage lending institutions with home purchase loan originations at least 10 percent of total loan originations, a metro office, and assets of $10 million or originated at least 100 loans, must also report their mortgage loans to HMDA.

[6] More complete description of HMDA data can be found in, *A Guide to HMDA Reporting: Getting It Right!* Published and updated frequently by the Federal Financial Institutions Examination Council (http://www.ffiec.gov/hmda/guide.htm).

[7] The previous HUD contract was C-OPC-18571, Task Order 9.

[8] Much more information on the Experian data is recorded in the "Experian Data Report," HUD, Policy Development and Research, May 30, 2003.

stratified according to underserved status (based on income and percent minority in 1999). Underserved tracts typically have fewer mortgage loans per year, so those tracts were sampled at a higher rate than the served tracts. A prioritized list of tracts was given to Experian and they extracted all the home purchase loans from those tracts that originated in 1998 to 2000 (over 1 million loans). The loan information came from county recorders. The credit score is the FICO score based on the borrower characteristics at approximately the time of the origination.

Researchers at HUD merged the Experian loan data with the HMDA data using all the loan level data available. [9] First, the match is based on geography. Both data sets have state, county, tract and MSA information geocoded. In addition, both databases have variables for loan amount, race, gender and loan type (conventional, FHA, Veterans Administration and Farmers Home Administration). The matching process goes through six iterations in which the best matches are removed and the remaining records are compared using fewer variables or wider bounds for a match. For example, in the first iteration, race, loan amount, gender, and loan type must all be equivalent for the loans to qualify as a match. In the next iteration, the race variable is dropped from the matching requirement. In the following round, race is brought back and loan amount is dropped. By iteration four, race and loan amount are dropped. Then in iteration five, race, loan amount and gender are dropped. Finally in iteration six, race, loan amount, gender and loan type are dropped. The matches are screened for unacceptable matches (race/ethnicity is different, loan amount differs by more than $3000, gender is different or both missing, or loan type does not match). There is also a tie-breaking protocol used in case where more than one loan record qualifies for a match. Out of the original 24 MSAs, the 11 MSAs with the best match rates were selected for analysis.

Weights are assigned to the matched loans so that the sum of the matched loans equaled the sum of the HMDA loans in each tract. The weight starts with a base weight according to the probability of selecting the tract multiplied by the probability of selecting the loan within a particular tract. Separate weights are needed for the served and underserved strata. To correct for missing loans and non-matches, adjustment factors are assigned to ensure the weighted total of matched loans equals the HMDA totals. For example, after the matching is completed, if the weighted total of loans for a tract is 90 percent of the HMDA total, then an adjustment factor of 1.11 is applied so that the weighted sample matches the HMDA total. Final weights are associated with each matched loan and used in the tabulations presented below. Separate weights have not been designed for the FHA-eligible subset or the FHA vs. GSE subset of loans. The assumption is that the tract level weights for the full, matched sample is adequate for subsets of loans drawn from the full sample.

A comparison between the matched and non-matched loans is shown in Exhibit 1 for the variables used to select the sample and conduct the regression analysis. The analogous tables by MSA are in the Appendix. To the extent that HMDA is representative of the universe of loans and the weights are designed to match HMDA, then the averages and distributions for the matched data were representative of the home purchase loans in the 11 MSAs during 1998-2000. Differences between the unmatched and matched data do not imply that the matched data were incorrect or unrepresentative because the unmatched data were unweighted. If a higher percentage of Experian loans had been successfully matched, that would have resulted in different unweighted values and

---

[9]   The data matching was done by Ismail Mohamed and Ron Hanson of Titan Systems, Inc. and Jay Schultz of the Office of Policy Development and Research, HUD.

different weights, but not necessarily different weighted values. Lacking more complete data, the weighted, matched sample provides the most representative estimates for the population values.

For the pooled sample of 11 MSAs, the median income in the matched sample, $55,000, is similar to the median income in HMDA, even though the weighted mean in the matched sample is much higher. As noted, there are extreme values in the reported incomes. The average FICO score in the matched sample (696) is higher than the unmatched Experian data, but a quarter of the unmatched data were missing FICO. FICO was a required field for the matching process. To make the distributions of unmatched Experian data more comparable to the matched data, the distributions are calculated on the non-missing loans. The matched data have a higher share of whites, but a lower share of race missing. The age distribution is very similar after adjusting for the 29 percent missing in the unmatched data. The average loan amount is higher in the matched data. However, the ratio of the loan amount to the FHA loan limit is essentially the same in the matched and unmatched data.

The distribution of LTV is shifted higher in the matched data, which has an average LTV of 84 (vs. 71 percent in the unmatched).[10] It is possible that the higher average LTV in the matched data accentuate the degree of overlap between FHA and GSE. Another notable difference is the lower percentage of new construction in the matched data compared to the unmatched data, 11 vs. 19 percent respectively.

The neighborhood characteristics in the matched and HMDA data are generally close. This is no surprise because the weights are at the tract level and should eliminate any substantial differences at the neighborhood level. The unmatched sample has a lower share of minority neighborhoods than the Experian data and a higher share of loans in low-cost MSAs.

Overall, the largest difference that would affect the study results is the higher average LTV in the matched data. Unfortunately, there are no LTV data in HMDA for comparison. If the weighted matched data have an upward bias for LTV, it is likely that the estimates for overlap are also biased upwards.

---

[10]   As shown in Exhibit 1, 22 percent of the matched sample had an LTV greater 98 percent, compared with 9 percent in the unmatched sample.

Exhibit 1: Analysis of Matched and Unmatched Experian and HMDA Loans from 11 MSAs (1998-2000)

Charateristics	Experian Unmatched (Unweighted)	Experian/HMDA Matched (Unweighted)	Experian/HMDA Matched (Weighted)	HMDA Unmatched	All HMDA
Borrower Characteristics					
Unweighted Number of Borrowers	239,529	347,732	347,732	1,589,133	1,936,865
Weighted Number of Borrowers			1,980,080		
Average Annual Income	$71,052	$60,986	$63,145	$69,041	$67,565
Median Annual Income	$62,500	$52,000	$55,000	$59,000	$58,000
Average Annual Income (Trimmed Top 1%)	$69,527	$58,440	$58,458	$65,827	$64,485
% Estimated Income Information	0%	2%	2%	2%	2%
Average FICO	690	694	696		
% With FICO < 620	22%	20%	20%		
% With FICO 620 - 679	18%	17%	17%		
% With FICO => 680	61%	62%	63%		
% Missing FICO Information	25%	0%	0%		
% White		71%	67%	63%	65%
% Black		10%	11%	9%	9%
% Hispanic		10%	12%	10%	10%
% Other		6%	7%	8%	7%
% Missing Race Information		3%	3%	10%	8%
% Female	21%	28%	28%	26%	26%
% Male	79%	72%	72%	67%	68%
% Missing Gender Information	7%	0%	0%	6%	5%
% Age 19-34	31%	32%	32%		
% Age 35-49	51%	49%	50%		
% Age 50-64	15%	15%	14%		
% Age >65	4%	4%	3%		
% Missing Age Information	29%	0%	0%		
Loan Characteristics					
Average Loan Amount	$124,847	$122,385	$128,684	$119,205	$119,776
Average LTV %	71%	84%	84%		
% With LTV <= 90	72%	56%	55%		
% With LTV 90 - 96	12%	15%	15%		
% With LTV 97 - 98	6%	7%	8%		
% With LTV > 98	9%	22%	21%		
% Missing LTV Information	37%	0%	0%		
Average Ratio of Loan Amount to FHA Loan Limit	74%	72%	73%		
% With LoantoFHA Ratio <=.5	24%	25%	23%		
% With LoantoFHA Ratio of .6 - 1	56%	58%	59%		
% With LoantoFHA Ratio of 1.1 - 1.2	12%	10%	11%		
% With LoantoFHA Ratio > 1.2	8%	7%	6%		

Exhibit 1 (cont.): Analysis of Matched and Unmatched Experian and HMDA Loans from 11 MSAs (1998-2000)

Charateristics	Experian Unmatched (Unweighted)	Experian/HMDA Matched (Unweighted)	Experian/HMDA Matched (Weighted)	HMDA Unmatched	All HMDA
Average PTI	19%	20%	20%	17%	18%
% Originated in 1998	23%	25%	32%	32%	31%
% Originated in 1999	35%	35%	34%	34%	34%
% Originated in 2000	43%	40%	34%	34%	35%
Mortgaged Property Characteristics					
% Old Construction	81%	89%	89%		
% New Construction	19%	11%	11%		
% Missing Construction Information	9%	0%	0%		
% Unit Size 1	96%	96%	95%		
% Unit Size 2	1%	1%	2%		
% Unit Size 3	2%	2%	2%		
% Unit Size 4	1%	1%	1%		
Borrower Neighborhood Characteristics (1990 Census)					
% In Underserved Tracts		34%	32%	31%	32%
% Not in Underserved Tracts		66%	68%	67%	67%
% Missing Underserved Tracts		0%	0%	2%	1%
% in High Cost Cities	10%	10%	12%	13%	12%
% in Average Cost Cities	78%	84%	85%	85%	85%
% in Low Cost Cities	12%	6%	3%	2%	3%
% In Center City	25%	28%	27%	27%	27%
% Not in Center City	75%	72%	73%	71%	71%
% Missing Center City Information	0%	0%	0%	2%	1%
Average 5-year Appreciation Lagged 1 year	117%	117%	113%	112%	113%
% In Area with Depreciation	4%	5%	11%	12%	11%
% In Area with Appreciation up to 20%	63%	63%	71%	73%	71%
% In Area with Appreciation over 20%	33%	33%	18%	15%	18%
% In <90% Relative Income Tracts	26%	26%	23%	24%	24%
% In 90-120% Relative Income Tracts	32%	33%	34%	32%	32%
% In >120% Relative Income Tracts	42%	41%	43%	44%	44%
% In <10% Minority Tracts	51%	50%	42%	40%	41%
% In 10-30% Minority Tracts	29%	30%	32%	34%	33%
% In >30% Minority Tracts	20%	20%	26%	25%	24%
% Missing Minority Tract Information	0%	0%	0%	2%	1%

Note: Loans in Boston MSA and Jumbo loans are excluded from this analysis. Share of missing is so large for the unmatched Experian loans that the distributions are calculated for the nonmissing loans in column 1.

Data Cleaning and Record Selection

In building analysis files from the matched Experian-HMDA data, there are several selections, as shown in Exhibit 2. The total number of matched records for 12 MSAs is 393,643, but Boston did not have loan type and jumbo loans are excluded from all analysis, leaving 347,732 loans in 11 MSAs. After excluding non-fixed-rate mortgages and loan amounts greater than the FHA loan limits, the FHA-eligible file contains 238,158 loans. A final selection is made for LTV between 80 and 100 percent and FHA or GSE loans to reach a working file of 114,780 loans. The LTV is calculated simply as the reported loan amount divided by the reported house value. The FHA designation is based on a match to the complete set of FHA loans. The GSE designation is based on the purchaser type in HMDA being either 1 (Fannie Mae) or 3 (Freddie Mac). Although there are FHA loans in the GSE portfolios, in the regression analysis the FHA loans are taken out of the GSE group.

Exhibit 2: Methodology for Limiting HUD Data sets

	Count of Loans		
	Add	Subtract	Net
Start with 'hmda_ex_selected_pmsa_1998_2000'	315,625		315,625
Append 'hmda_ex_chicago_1998_2000'	51,900		367,525
Append 'hmda_ex_la_1998_2000'	26,118		393,643
Exclude Boston loans (MSA<>1120)		12,009	381,634
Exclude remaining jumbo loans (Conform=1)		33,902	347,732
	393,643	45,911	347,732
Exclude remaining non-fixed rate mortgages (ex_rate_type<>B, U, or V)		61,847	285,885
Exclude remaining loans above FHA limit		47,727	238,158
		109,574	238,158
Exclude remaining loans not between 80-100% LTV		80,146	158,012
Limit to FHA loans or loans in pur_type=1 (FNMA) or 3 (FHLMC)		43,232	114,780
		123,378	114,780

In the large analysis file (347,732 loans), there are 61,922 PMI loans. Traditionally, mortgage insurance was required by the GSEs for loans with LTV greater than 80 percent. Although most PMI loans have LTV greater than 80 percent, 28 percent of the PMI loans have LTV less than or equal to 80 percent, as shown in Exhibit 3. The lower panel shows that most of the below-80 percent loans are in the range of 70 to 80 percent, as expected, but there are some much lower. No attempt was made to impute or exclude those loans in the PMI analysis; that is, they were included.

Exhibit 3
PMI Loans by LTV Ratio
On All Matched, Conforming Loans, Weighted
(n=347,732; weighted sample=1,980,080)

LTV Ratio	Number of Loans with PMI	Percent of Loans with PMI
<=80	17,615	28%
81-85	3,280	5%
86-90	9,740	16%
91-95	17,403	28%
96-97	10,033	16%
>97	3,851	6%
Total PMI Loans	61,922	100%

Low LTV PMI Loans by LTV Ratio
On All Matched, Conforming Loans, Weighted
(n=347,732; weighted sample=1,980,080)

LTV Ratio	Number of Loans with PMI
<=10	17
11-20	19
21-30	60
31-40	172
41-50	394
51-60	848
61-70	1,443
71-80	14,662

The intersection of PMI with GSE loans is shown in Exhibit 4. Out of 115,798 total GSE loans, 69 percent or 79,777 are without PMI and 31 percent or 36,021 have PMI. Most of the GSE loans without PMI have LTV less than or equal to 80 percent, and most of the GSE loans with PMI have LTV greater than 80 percent. However, there are many cases (10,206) of GSE loans below 80 percent LTV with PMI and GSE loans above 80 percent LTV without PMI (23,975). There are even a substantial number (1,797) of GSE loans with PMI above 97 percent LTV. Apparently, PMI is not a hard constraint on the GSEs for their high-LTV purchases and the GSEs are using other forms of credit enhancement. It is likely that there are more than 48 percent of GSE loans above 80 percent with PMI, but they were not successfully matched either between Experian and HMDA or PMI and HMDA.

Exhibit 4: GSE Loans[1] With & Without PMI by LTV Ratio
On All Matched, Conforming Loans, Weighted
(n=347,732; weighted sample=1,980,080)

LTV Ratio	Without PMI	Percent Without PMI	With PMI	Percent With PMI
<=80	55,802	85%	10,206	15%
>80	23,975	48%	25,815	52%
Total GSE Loans	79,777	69%	36,021	31%

LTV Ratio	Without PMI	Percent Without PMI	With PMI	Percent With PMI
<=80	55,802	85%	10,206	15%
81-85	6,205	76%	1,933	24%
86-90	4,325	43%	5,793	57%
91-95	6,625	39%	10,458	61%
96-97	3,575	38%	5,834	62%
>97	3,245	64%	1,797	36%
Total GSE Loans	79,777	69%	36,021	31%

1. Percent of GSE Loans with an LTV above 80% that do not have PMI: 48%

There are 93,606 total FHA loans with dwelling unit size as reported by Experian in Exhibit 5. All of these loans are supposed to be associated with single-family properties (1 to 4 units), but a sizable portion of the units is reported to be in buildings with 5+ units. Given that out of the entire set of FHA loans, only 347 exceed the loan limits for a single unit that do not exceed the loan limits for 4 units, the loans in 5+ unit buildings are treated as single-unit loans. Loans in 2- to 4-unit buildings are assumed to be loans for multiple units and the FHA loan limits are adjusted accordingly.

Exhibit 5: FHA Loans by Experian Dwelling Unit Size

Dwelling Unit Size	"EX_UNIT_SIZE" Category	Number of FHA Loans	Percent of FHA Loans
Single	C, Missing	79,533	85%
Duplex	D	1,332	1%
3-unit	E	1,568	2%
4-unit	F	848	1%
Larger	G	4,920	5%
Larger	H	797	1%
Larger	I	148	0%
Larger	J	613	1%
Larger	K	71	0.1%
Larger	L	1,569	2%
Larger	M	922	1%
Larger	N	576	1%
Larger	O	709	1%
TOTAL FHA LOANS		93,606	100%

Percent of FHA Loans for Units in Dwellings Larger than 4 units: 11%

"EX_UNIT_SIZE" Category	Loan Limit Size	Number of FHA Loans	Percent of FHA Loans
C, G-O, Missing	1	89,858	96%
D	2	1332	1%
E	3	1568	2%
F	4	848	1%
TOTAL FHA LOANS		93,606	100%

FHA eligibility is based on FHA loan limits, which can change throughout the year. This analysis uses the loan amount reported in HMDA, which is rounded to the nearest $1000. The FHA designation comes from a match to FHA records, so those loans are expected to be below the FHA loan limits at the time of origination. However, as shown in Exhibit 6, there are 2,345 FHA loans (2.5 percent of 93,606 total FHA loans) above the FHA loan limits for the date and MSA of origination. The data appear to be inconsistent. The limits could have been increased until nearly all the FHA loans were included, but the concern was that would include many GSE loans that were actually ineligible at their time of origination. Therefore, the more conservative approach for FHA eligibility of applying the FHA loan limit as of the origination date was chosen, even though this excludes 2,345 FHA loans.

Exhibit 7 shows the number of missing observations associated with a few of the variables that are not completely available. By far, the biggest problem of incomplete data is due to race/ethnicity not being reported on HMDA.

Exhibit 6: Number of FHA Loans Above FHA Loan Limits[1]
 On All Matched, Conforming Loans, Weighted
(n=347,732; weighted sample=1,980,080)

MSA	FHA Loans Above FHA Loan Limits Unrounded	FHA Loans Above FHA Loan Limits2 Rounded Up to Nearest $1000
0720	90	77
1600	213	190
1680	122	117
2080	613	553
4480	285	252
5775	144	125
6160	131	99
6440	320	281
7040	97	80
8280	405	334
8840	272	237
Total	2,692	2,345

FHA Loan Limit Source: FHA_limits_holly.xls.

1. Number of FHA loans above loan limits, excluding loans not in ex_unit_sizes C, D, E, F or missing. Loan limits assigned as follows by ex_unit_size: if C or missing, then Unit Size 1, if D then Unit Size 2, if E then Unit Size 3, if F then Unit Size 4, if G or higher then Unit Size 1 (assumed single unit in larger building). In 1998, FHA loan limits for units in categories D, E and F are estimated based on loan limits for category C. D limit = C limit*1.28. E limit = C limit*1.55. F limit = C limit*1.92.

2. Loan limits based on loan limits as of October 21, 1998; December 1999; and January 2000. Note that 2000 loan limits do not change. Loans are compared to loan limits according to HMDA Action Date.

Exhibit 7: Observations with Missing Data
 On All Matched, Conforming Loans, Weighted
(n=347,732; weighted sample=1,980,080)

Missing Data	Number of Loans	Reason
OBS_POST_WGHT	169	No weights. Model ignores these observations.
Race	10,665	BO_Race in categories 0, 7 or 8.
Center City	86	No information provided by Unicon for 2 tracts in our data.
Age	721	No age or age range information.

There are 19,330 subprime loans in the large analysis file, as shown in Exhibit 8. The subprime designation is based on the lender name (HUD determined which agencies were predominantly subprime lenders in each origination year). Most of the subprime loans (53 percent) have a purchaser type of life insurance company or other, although 41 percent are held by depositories. A small share (5 percent) are insured by FHA and 6 percent are purchased by GSEs. It is quite possible there are errors in the designation of subprime loans and that the subprime mortgage sector is considerably larger than represented in this data set. Most subprime loans are refinance loans, which are not considered in this research.

Exhibit 8: Subprime Loans[1] by Mortgage Market Sector[2]
 On All Matched, Conforming Loans, Weighted
 (n=347,732; weighted sample=1,980,080)

Sector	Number of Subprime Loans	Percent of Subprime Loans
FHA	994	5%
PMI	1,507	8%
Other Investor	10,333	53%
Depository	7,833	41%
GSE	1,116	6%
TOTAL Subprime Loans[3]	19,330	

1. Subprime data downloaded from www.huduser.org/datasets/manu.html. Merged on compressed AGENCY and RESP_ID. SUBPRIME=1 if lender classified as a primarily subprime lender in year of loan origination.

2. Sectors assigned as follows:
FHA   EX_LOAN_TYPE_HUD2=F
PMI   PMI_FLAG="Y"
Other Investor   PUR_TYPE=7, 9
Depository   PUR_TYPE=0, 5, 6, 8
GSE   PUR_TYPE=1,3

3. Total does not equal sum of subprime loans in each sector because the sectors overlap, e.g, PMI overlaps with GSE.

The cross tabulation of HMDA purchaser type with FHA, PMI and subprime mortgage sectors is shown in Exhibit 9. The FHA column shows that 1.6 percent of the FHA loans were purchased by Fannie Mae (usually as part of a batch sale). Typically most FHA loans are insured by Ginnie Mae, but according to purchaser type only 46.5 percent are reported under Ginnie Mae. The PMI column shows that most PMI loans are purchased by the GSEs (33.7 percent by Fannie Mae and 24.1 percent by Freddie Mac), but a substantial fraction (42 percent) are held by other institutions.

Exhibit 9: Distribution of Loans by Purchaser Type
 On All Matched, Conforming Loans, Weighted
 (n=347,732; weighted sample=1,980,080)

	HMDA Purchaser Type	FHA	PMI	Subprime
0	Loan not originated or sold in calendar year	10.6%	20.2%	24.9%
1	Federal National Mortgage Association	1.6%	33.7%	2.3%
2	Government National Mortgage Association	46.5%	0.0%	0.2%
3	Federal Home Loan Mortgage Association	0.1%	24.1%	4.7%
4	Farmers Home Administration	0.2%	0.0%	0.0%
5	Commercial Bank	2.9%	1.1%	1.5%
6	Savings Bank or Association	0.6%	0.8%	0.4%
7	Life Insurance Company	0.1%	0.0%	0.0%
8	Affiliate Institution	2.5%	7.5%	11.0%
9	Other	34.9%	12.6%	54.9%

Explanatory Variables

The goal was to use the combined Experian/HMDA data, along with census tract level data and OFHEO MSA level house price appreciation data, to capture the most important factors in the underwriting decision. The FICO score provided credit history information, and house value made it possible to calculate LTV. In addition, borrower income made it possible to calculate the payment-to-income ratio, and demographics provided measures of age, gender, and race. The following description provides more details about how each explanatory variable was calculated and then how these variables were used to subset the data for analysis of mortgage sectors. A data dictionary is also provided in Appendix Exhibit A.1.

In Exhibits 10 through 14, borrower, loan, property, and neighborhood characteristics are evaluated to determine how the pools of loans change as the data were categorized according to the area of greatest potential overlap between FHA and the GSEs. Note that Exhibits 10 to 13 continue on a second page.

Exhibit 10: Comparison of Characteristics by Dataset
Weighted

	Matched, Conforming, Fixed-Rate, FHA-Eligible FHA & GSE Loans with LTV 80-100%	Matched, Conforming, Fixed-Rate, FHA-Eligible Loans	Matched, Conforming Loans
Borrower Characteristics			
Unweighted Number of Borrowers	114,780	238,158	347,732
Weighted Number of Borrowers	674,238	1,369,923	1,980,080
Average Annual Income	$52,438	$55,749	$63,145
Median Annual Income	$47,000	$49,000	$55,000
Average Annual Income (Trimmed Top 1%)	$50,697	$53,596	$60,552
% Estimated Income Information	1%	2%	2%
Average FICO	668	693	696
% With FICO <620	29%	21%	20%
% With FICO 620-680	21%	17%	17%
% With FICO =>680	50%	62%	63%
% Missing FICO Information	0%	0%	0%
% White	57%	65%	67%
% Black	16%	12%	11%
% Hispanic	19%	14%	12%
% Other	5%	6%	7%
% Missing Race Information	3%	3%	3%
% Female	31%	31%	28%
% Age 19-34	38%	33%	32%
% Age 35-49	49%	49%	50%
% Age 50-64	11%	14%	14%
% Age >65	2%	4%	3%
% Missing Age Information	0%	0%	0%
Loan Characteristics			
Average Loan Amount	$116,716	$113,209	$128,684
Average LTV %	95	85	84
% With LTV<=80	6%	38%	42%
% With LTV 80-90	12%	11%	13%
% With LTV 90-96	17%	15%	15%
% With LTV 96-98	14%	9%	8%
% With LTV>98	51%	27%	21%
% Missing LTV Information	0%	0%	0%
Average Ratio of Loan Amount to FHA Loan Limit	64%	63%	73%
% With LoantoFHA Ratio <=.5	25%	28%	23%
% With LoantoFHA Ratio of .5 - 1	75%	72%	59%
% With LoantoFHA Ratio of 1 - 1.2	0%	0%	11%
% With LoantoFHA Ratio of >1.2	0%	0%	6%

Exhibit 10 (cont.): Comparison of Characteristics by Dataset
Weighted

	Matched, Conforming, Fixed-Rate, FHA-Eligible FHA & GSE Loans with LTV 80-100%	Matched, Conforming, Fixed-Rate, FHA-Eligible Loans	Matched, Conforming Loans
Average PTI	21%	20%	20%
% Originated in 1998	31%	32%	32%
% Originated in 1999	35%	34%	34%
% Originated in 2000	34%	34%	34%
Mortgaged Property Characteristics			
% New Construction	8%	9%	11%
% Unit Size 1	95%	95%	95%
% Unit Size 2	2%	2%	2%
% Unit Size 3	2%	2%	2%
% Unit Size 4	1%	1%	1%
Borrower Neighborhood Characteristics (1990 Census)			
% In Underserved Tracts	43%	36%	32%
% In High Cost MSA	11%	11%	12%
% In Medium Cost MSA	86%	86%	85%
% In Low Cost MSA	3%	3%	3%
% In Center City	29%	28%	27%
Average 5-yr House Price Appreciation, Lagged 1 year	112%	113%	113%
% In Area with Depreciation	12%	10%	11%
% In Area with Appreciation up to 20%	74%	73%	71%
% In Area with Appreciation over 20%	14%	17%	18%
% In Tracts with Income <90% of MSA Income	29%	26%	23%
% In Tracts with Income 90 - 120% of MSA Income	39%	36%	34%
% In Tracts with Income =>120% of MSA Income	31%	38%	43%
% In <10% Minority Tracts	34%	41%	42%
% In 10-30% Minority Tracts	33%	32%	32%
% In =>30% Minority Tracts	32%	27%	26%

Exhibit 11: Analysis of Loans In Mortgage Market Sectors

On All Matched, Conforming Loans, Weighted
(n=347,732; weighted sample=1,980,080)

	GSE Purchased Loans			Loans Held by Depository Lenders			Other Investors			Not Mutually Exclusive		
	All GSE	GSE With PMI	GSE No PMI	All Depository	Depository With PMI	Depository No PMI	All Other Investors	Other Investors With PMI	Other Investors No PMI	All with PMI	All FHA	All Subprime
	pur_type=1,3			pur_type=0,5,6,8			pur_type=7,9			pmi_flag=1	fha_loan=1	subprime=1
Share of Loans	33%	10%	23%	33%	5%	28%	21%	2%	19%	17%	27%	5%
Borrower Characteristics												
Unweighted Number of Borrowers	115,798	36,021	79,777	114,010	17,999	96,011	72,626	7,901	64,725	61,922	93,606	19,330
Weighted Number of Borrowers	658,228	198,779	459,449	650,925	101,875	549,050	417,906	43,236	374,670	343,900	542,138	101,464
Average Annual Income	$68,721	$63,199	$71,110	$64,963	$60,127	$65,861	$59,734	$64,034	$59,237	$62,394	$49,325	$61,337
Median Annual Income	$62,000	$60,000	$63,000	$55,000	$55,000	$55,000	$52,000	$60,000	$51,000	$58,000	$45,000	$53,000
Average Annual Income (Trimmed Top 1%)	$66,334	$61,889	$68,383	$61,873	$58,778	$62,470	$57,371	$62,654	$56,774	$61,034	$47,634	$58,527
% Estimated Income Information	1%	4%	2%	1%	7%	1%	2%	17%	2%	2%	1%	7%
Average FICO	726	714	731	699	706	698	674	712	670	712	643	637
% With FICO <620	9%	12%	8%	19%	14%	20%	28%	12%	29%	12%	38%	43%
% With FICO 620-680	13%	15%	12%	16%	17%	16%	20%	16%	20%	16%	25%	20%
% With FICO =>680	78%	73%	80%	65%	69%	64%	53%	72%	51%	72%	38%	36%
% Missing FICO Information	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
% White	77%	75%	78%	68%	70%	68%	60%	75%	58%	73%	50%	52%
% Black	5%	6%	4%	10%	8%	11%	15%	6%	16%	7%	21%	22%
% Hispanic	7%	10%	6%	11%	13%	10%	17%	10%	17%	11%	23%	12%
% Other	8%	6%	9%	8%	6%	9%	6%	5%	6%	6%	3%	7%
% Missing Race Information	3%	4%	3%	3%	3%	3%	3%	3%	3%	3%	2%	8%
% Female	24%	23%	25%	29%	26%	29%	30%	25%	30%	24%	32%	34%
% Age 19-34	29%	34%	27%	30%	34%	30%	33%	34%	33%	34%	40%	27%
% Age 35-49	51%	51%	50%	50%	50%	50%	51%	51%	51%	51%	49%	55%
% Age 50-64	16%	13%	18%	15%	13%	16%	13%	13%	13%	13%	10%	15%
% Age >65	4%	2%	5%	4%	3%	4%	3%	3%	3%	2%	2%	3%
% Missing Age Information	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Loan Characteristics												
Average Loan Amount	$139,103	$138,163	$139,510	$125,531	$132,717	$124,198	$126,350	$141,740	$124,574	$136,999	$114,311	$112,359
Average LTV %	80	88	76	82	88	81	87	88	87	88	97	81
% With LTV<=80	56%	25%	69%	49%	26%	53%	33%	24%	34%	25%	6%	53%
% With LTV 80-90	16%	22%	13%	15%	20%	14%	11%	23%	10%	22%	2%	22%
% With LTV 90-96	21%	43%	12%	15%	38%	11%	12%	44%	8%	42%	5%	14%
% With LTV 96-98	4%	8%	3%	8%	11%	7%	9%	6%	9%	9%	16%	3%
% With LTV>98	3%	2%	3%	13%	5%	15%	34%	2%	38%	3%	71%	7%
% Missing LTV Information	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%

Exhibit 11(cont.): Analysis of Loans In Mortgage Market Sectors

On All Matched, Conforming Loans, Weighted
(n=347,732; weighted sample=1,980,080)

	GSE Purchased Loans			Loans Held by Depository Lenders			Other Investors			Not Mutually Exclusive		
	All GSE	GSE With PMI	GSE No PMI	All Depository	Depository With PMI	Depository No PMI	All Other Investors	Other Investors With PMI	Other Investors No PMI	All with PMI	All FHA	All Subprime
	pur_type=1,3			pur_type=0,5,6,8			pur_type=7,9			pmi_flag=1	fha_loan=1	subprime=1
Average Ratio of Loan Amount to FHA Loan Limit	80%	81%	80%	71%	75%	70%	70%	80%	68%	79%	63%	62%
% With LoantoFHA Ratio <=.5	18%	15%	18%	27%	21%	29%	25%	16%	26%	17%	27%	38%
% With LoantoFHA Ratio of .5 - 1	57%	59%	56%	56%	59%	55%	62%	60%	62%	59%	70%	50%
% With LoantoFHA Ratio of 1 - 1.2	16%	16%	16%	11%	13%	11%	9%	17%	8%	15%	3%	8%
% With LoantoFHA Ratio of >1.2	10%	9%	10%	6%	6%	5%	4%	8%	3%	8%	0%	3%
Average PTI	19%	20%	19%	19%	20%	19%	20%	20%	20%	20%	22%	18%
% Originated in 1998	36%	38%	35%	28%	29%	28%	31%	31%	31%	34%	30%	26%
% Originated in 1999	31%	30%	31%	37%	40%	36%	34%	31%	34%	33%	36%	34%
% Originated in 2000	33%	32%	34%	35%	31%	36%	35%	38%	35%	32%	34%	39%
Mortgaged Property Characteristics												
% New Construction	12%	10%	12%	10%	10%	10%	11%	11%	10%	10%	8%	7%
% Unit Size 1	96%	95%	96%	95%	94%	95%	96%	95%	96%	95%	95%	95%
% Unit Size 2	1%	1%	1%	2%	2%	2%	2%	1%	2%	2%	2%	1%
% Unit Size 3	2%	2%	2%	2%	3%	2%	2%	2%	2%	2%	2%	2%
% Unit Size 4	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%
Borrower Neighborhood Characteristics (1990 Census)												
% In Underserved Tracts	22%	26%	21%	33%	34%	33%	40%	29%	42%	29%	48%	47%
% In High Cost MSA	10%	8%	11%	14%	12%	14%	14%	11%	15%	10%	11%	17%
% In Medium Cost MSA	87%	88%	87%	83%	85%	83%	84%	86%	83%	87%	87%	79%
% In Low Cost MSA	3%	3%	3%	3%	3%	3%	2%	2%	2%	3%	3%	4%
% In Center City	24%	27%	23%	29%	31%	28%	28%	27%	28%	28%	29%	36%
Average 5-yr House Price Appreciation, Lagged 1 year	114%	114%	114%	113%	114%	113%	112%	114%	112%	114%	111%	113%
% In Area with Depreciation	10%	10%	10%	10%	9%	10%	14%	12%	14%	10%	11%	14%
% In Area with Appreciation up to 20%	72%	71%	72%	70%	70%	70%	69%	68%	69%	70%	75%	64%
% In Area with Appreciation over 20%	19%	19%	18%	21%	21%	21%	17%	20%	16%	20%	14%	22%
% In Tracts with Income <90% of MSA Income	16%	19%	14%	25%	26%	25%	27%	20%	28%	21%	33%	35%
% In Tracts with Income 90 - 120% of MSA Income	32%	34%	30%	33%	33%	33%	36%	34%	36%	34%	40%	33%
% In Tracts with Income =>120% of MSA Income	53%	47%	55%	42%	41%	43%	37%	46%	36%	45%	28%	33%
% In <10% Minority Tracts	49%	48%	50%	44%	43%	44%	35%	44%	34%	46%	31%	31%
% In 10-30% Minority Tracts	33%	33%	33%	30%	31%	30%	32%	34%	32%	32%	32%	31%
% In =>30% Minority Tracts	18%	20%	17%	26%	26%	26%	33%	23%	34%	22%	37%	38%

Exhibit 12: Comparison of Fixed-Rate vs. Adjustable-Rate Mortgage
On All Matched, Conforming Loans, Weighted
(n=347,732; weighted sample=1,980,080)

	Fixed-Rate Loans	Adjustable-Rate Loans
	ex_rate_type <> B,U,V	ex_rate_type = B,U,V
Share of Loans	83%	17%
Borrower Characteristics		
Unweighted Number of Borrowers	285,885	61,847
Weighted Number of Borrowers	1,649,056	331,024
Average Annual Income	$62,203	$67,838
Median Annual Income	$54,000	$59,000
Average Annual Income (Trimmed Top 1%)	$59,723	$64,785
% Estimated Income Information	2%	2%
Average FICO	700	677
% With FICO <620	19%	26%
% With FICO 620-680	16%	20%
% With FICO =>680	65%	55%
% Missing FICO Information	0%	0%
% White	67%	67%
% Black	10%	11%
% Hispanic	12%	10%
% Other	7%	8%
% Missing Race Information	3%	3%
% Female	28%	29%
% Age 19-34	32%	33%
% Age 35-49	50%	50%
% Age 50-64	15%	14%
% Age >65	3%	3%
% Missing Age Information	0%	0%
Loan Characteristics		
Average Loan Amount	$127,247	$135,844
Average LTV %	84	83
% With LTV<=80	41%	49%
% With LTV 80-90	12%	19%
% With LTV 90-96	16%	12%
% With LTV 96-98	8%	5%
% With LTV>98	23%	15%
% Missing LTV Information	0%	0%

Exhibit 12 (cont.): Comparison of Fixed-Rate vs. Adjustable-Rate Mortgage
On All Matched, Conforming Loans, Weighted
(n=347,732; weighted sample=1,980,080)

	Fixed-Rate Loans ex_rate_type <> B,U,V	Adjustable-Rate Loans ex_rate_type = B,U,V
Average Ratio of Loan Amount to FHA Loan Limit	72%	74%
% With LoantoFHA Ratio <=.5	24%	23%
% With LoantoFHA Ratio of .5 - 1	59%	59%
% With LoantoFHA Ratio of 1 - 1.2	11%	13%
% With LoantoFHA Ratio of >1.2	6%	5%
Average PTI	20%	20%
% Originated in 1998	34%	19%
% Originated in 1999	33%	39%
% Originated in 2000	32%	42%
Mortgaged Property Characteristics		
% New Construction	11%	11%
% Unit Size 1	95%	95%
% Unit Size 2	2%	2%
% Unit Size 3	2%	3%
% Unit Size 4	1%	1%
Borrower Neighborhood Characteristics (1990 Census)		
% In Underserved Tracts	33%	32%
% In High Cost MSA	11%	17%
% In Medium Cost MSA	86%	80%
% In Low Cost MSA	3%	3%
% In Center City	27%	28%
Average 5-yr House Price Appreciation, Lagged 1 year	113%	115%
% In Area with Depreciation	11%	10%
% In Area with Appreciation up to 20%	72%	65%
% In Area with Appreciation over 20%	17%	25%
% In Tracts with Income <90% of MSA Income	23%	23%
% In Tracts with Income 90 - 120% of MSA Income	34%	33%
% In Tracts with Income =>120% of MSA Income	43%	44%
% In <10% Minority Tracts	43%	40%
% In 10-30% Minority Tracts	32%	33%
% In =>30% Minority Tracts	25%	27%

Exhibit 13: Analysis of FHA-Eligible Loans In Mortgage Market Sectors

On Matched, Conforming, Fixed-Rate FHA-Eligible Loans, Weighted
(n=238,158; weighted sample=1,369,923)

	GSE Purchased Loans			Loans Held by Depository Lenders			Other Investors			Not Mutually Exclusive		
	All GSE	GSE With PMI	GSE No PMI	All Depository	Depository With PMI	Depository No PMI	All Other Investors	Other Investors With PMI	Other Investors No PMI	All with PMI	All FHA	All Subprime
	pur_type=1,3			pur_type=0,5,6,8			pur_type=7,9			pmi_flag=1	fha_loan=1	subprime=1
Share of Loans	33%	10%	23%	30%	5%	25%	22%	2%	20%	17%	34%	4%
Borrower Characteristics												
Unweighted Number of Borrowers	77,895	24,452	53,443	70,646	10,789	59,857	51,735	5,193	46,542	40,435	79,175	10,419
Weighted Number of Borrowers	445,372	136,499	308,873	407,958	61,981	345,977	302,835	28,708	274,127	227,199	465,118	54,874
Average Annual Income	$59,710	$55,723	$61,472	$56,296	$52,258	$57,019	$54,226	$56,820	$53,955	$54,916	$48,384	$57,065
Median Annual Income	$54,000	$52,000	$55,000	$48,000	$48,000	$48,000	$48,000	$52,000	$47,000	$51,000	$44,000	$49,000
Average Annual Income (Trimmed Top 1%)	$57,731	$54,730	$59,171	$53,756	$51,146	$54,299	$51,968	$55,605	$51,613	$53,866	$46,718	$54,443
% Estimated Income Information	1%	1%	1%	3%	4%	3%	3%	7%	3%	3%	1%	13%
Average FICO	726	714	731	698	706	696	673	713	669	712	642	653
% With FICO <620	9%	12%	8%	20%	14%	20%	28%	12%	30%	12%	38%	37%
% With FICO 620-680	13%	16%	12%	16%	17%	16%	20%	16%	20%	16%	24%	18%
% With FICO =>680	78%	73%	80%	64%	69%	63%	52%	73%	50%	72%	37%	45%
% Missing FICO Information	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
% White	76%	73%	77%	66%	66%	66%	57%	73%	56%	71%	49%	51%
% Black	5%	7%	4%	12%	9%	13%	16%	7%	17%	7%	22%	22%
% Hispanic	8%	11%	7%	12%	16%	11%	19%	11%	20%	13%	24%	12%
% Other	8%	6%	9%	7%	5%	7%	5%	5%	5%	6%	3%	6%
% Missing Race Information	3%	3%	3%	3%	3%	3%	3%	3%	3%	3%	2%	8%
% Female	28%	27%	29%	32%	30%	32%	31%	29%	32%	28%	33%	34%
% Age 19-34	30%	34%	27%	31%	35%	31%	34%	34%	34%	34%	39%	27%
% Age 35-49	49%	50%	49%	49%	49%	49%	50%	50%	50%	50%	49%	54%
% Age 50-64	17%	13%	18%	15%	13%	16%	13%	13%	13%	13%	10%	15%
% Age >65	5%	3%	5%	4%	3%	5%	3%	3%	3%	3%	2%	4%
% Missing Age Information	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Loan Characteristics												
Average Loan Amount	$118,706	$120,645	$117,849	$107,541	$114,740	$106,251	$114,949	$124,029	$113,998	$119,462	$111,413	$98,472
Average LTV %	79	88	75	83	89	81	89	88	89	88	97	82
% With LTV<=80	55%	23%	69%	44%	24%	48%	29%	22%	29%	23%	5%	51%
% With LTV 80-90	15%	22%	13%	12%	18%	11%	9%	23%	7%	21%	2%	17%
% With LTV 90-96	21%	44%	11%	15%	36%	11%	11%	46%	8%	42%	5%	16%
% With LTV 96-98	5%	10%	3%	10%	15%	9%	10%	7%	11%	11%	15%	4%
% With LTV>98	3%	2%	4%	18%	7%	20%	41%	2%	46%	3%	72%	11%
% Missing LTV Information	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%

Exhibit 13 (cont.): Analysis of FHA-Eligible Loans In Mortgage Market Sectors

On Matched, Conforming, Fixed-Rate FHA-Eligible Loans, Weighted
(n=238,158; weighted sample=1,369,923)

	GSE Purchased Loans			Loans Held by Depository Lenders			Other Investors			Not Mutually Exclusive		
	All GSE	GSE With PMI	GSE No PMI	All Depository	Depository With PMI	Depository No PMI	All Other Investors	Other Investors With PMI	Other Investors No PMI	All with PMI	All FHA	All Subprime
	pur_type=1,3			pur_type=0,5,6,8			pur_type=7,9			pmi_flag=1	fha_loan=1	subprime=1
Average Ratio of Loan Amount to FHA Loan Limit	66%	68%	65%	60%	64%	59%	63%	68%	62%	67%	62%	55%
% With LoantoFHA Ratio <=.5	24%	20%	25%	34%	27%	35%	28%	21%	29%	22%	29%	44%
% With LoantoFHA Ratio of .5 - 1	76%	80%	75%	66%	73%	65%	72%	79%	71%	78%	71%	56%
% With LoantoFHA Ratio of 1 - 1.2	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
% With LoantoFHA Ratio of >1.2	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Average PTI	19%	20%	19%	19%	20%	19%	21%	20%	21%	20%	21%	17%
% Originated in 1998	34%	35%	34%	30%	31%	30%	31%	29%	31%	33%	31%	28%
% Originated in 1999	32%	32%	32%	36%	40%	35%	33%	31%	34%	34%	35%	35%
% Originated in 2000	34%	33%	34%	34%	29%	34%	35%	39%	35%	33%	34%	37%
Mortgaged Property Characteristics												
% New Construction	9%	8%	10%	9%	7%	9%	9%	8%	9%	8%	8%	7%
% Unit Size 1	94%	94%	94%	95%	93%	95%	95%	95%	95%	94%	95%	94%
% Unit Size 2	2%	2%	2%	2%	2%	2%	2%	1%	2%	2%	2%	2%
% Unit Size 3	3%	3%	2%	2%	3%	2%	2%	2%	2%	3%	2%	3%
% Unit Size 4	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%
Borrower Neighborhood Characteristics (1990 Census)												
% In Underserved Tracts	26%	31%	24%	37%	39%	37%	44%	32%	45%	33%	49%	48%
% In High Cost MSA	10%	8%	11%	12%	12%	12%	14%	12%	14%	10%	11%	15%
% In Medium Cost MSA	88%	89%	87%	85%	85%	85%	84%	86%	84%	87%	86%	81%
% In Low Cost MSA	3%	3%	2%	3%	3%	3%	2%	2%	2%	3%	3%	4%
% In Center City	26%	29%	24%	30%	34%	29%	29%	30%	29%	31%	30%	36%
Average 5-yr House Price Appreciation, Lagged 1 year	114%	115%	114%	113%	114%	112%	111%	114%	111%	114%	111%	112%
% In Area with Depreciation	9%	9%	10%	9%	9%	9%	14%	12%	14%	9%	12%	13%
% In Area with Appreciation up to 20%	73%	72%	73%	73%	72%	73%	71%	69%	71%	72%	75%	68%
% In Area with Appreciation over 20%	18%	19%	17%	19%	19%	19%	15%	19%	14%	19%	13%	19%
% In Tracts with Income <90% of MSA Income	19%	22%	17%	28%	31%	28%	30%	23%	30%	25%	34%	37%
% In Tracts with Income 90 - 120% of MSA Income	35%	37%	34%	35%	35%	35%	38%	37%	38%	36%	40%	34%
% In Tracts with Income =>120% of MSA Income	47%	40%	49%	37%	35%	37%	33%	40%	32%	39%	26%	30%
% In <10% Minority Tracts	48%	46%	49%	44%	41%	44%	33%	42%	33%	44%	30%	33%
% In 10-30% Minority Tracts	33%	32%	33%	29%	30%	29%	32%	34%	32%	32%	32%	30%
% In =>30% Minority Tracts	20%	22%	19%	27%	29%	27%	35%	24%	36%	24%	38%	36%

Exhibit 14: Analysis of 80-100% LTV Loans In FHA and GSE Market Sectors

On Matched, Conforming, Fixed-Rate FHA-Eligible FHA & GSE Loans with LTV 80-100%, Weighted
(n=114,780; weighted sample=674,238)

	GSE Purchased Loans			Not Mutually Exclusive	
	All GSE	GSE With PMI	GSE No PMI	All FHA	All Subprime*
		pur_type=1,3		fha_loan=1	subprime=1
Share of Loans	36%	16%	20%	65%	1%
Borrower Characteristics					
Unweighted Number of Borrowers	40,600	18,673	21,927	75,462	1,361
Weighted Number of Borrowers	242,176	108,889	133,287	440,104	8,033
Average Annual Income	$59,624	$55,899	$62,667	$48,467	$48,145
Median Annual Income	$54,000	$52,000	$56,000	$44,000	$42,500
Average Annual Income (Trimmed Top 1%)	$57,860	$54,945	$60,390	$46,792	$46,977
% Estimated Income Information	1%	1%	1%	1%	18%
Average FICO	713	710	715	642	642
% With FICO <620	12%	12%	12%	38%	39%
% With FICO 620-680	16%	17%	15%	24%	21%
% With FICO =>680	73%	71%	73%	38%	40%
% Missing FICO Information	0%	0%	0%	0%	0%
% White	71%	70%	72%	50%	44%
% Black	7%	7%	6%	21%	25%
% Hispanic	11%	13%	10%	24%	21%
% Other	7%	6%	9%	3%	6%
% Missing Race Information	3%	4%	3%	2%	4%
% Female	28%	27%	28%	33%	40%
% Age 19-34	35%	36%	34%	39%	32%
% Age 35-49	49%	50%	49%	49%	55%
% Age 50-64	13%	12%	14%	10%	11%
% Age >65	3%	2%	3%	2%	2%
% Missing Age Information	0%	0%	0%	0%	0%
Loan Characteristics					
Average Loan Amount	$124,978	$122,685	$126,852	$112,017	$104,497
Average LTV %	90	92	88	98	95
% With LTV<=80	17%	3%	28%	0%	6%
% With LTV 80-90	28%	27%	29%	2%	12%
% With LTV 90-96	39%	55%	26%	5%	23%
% With LTV 96-98	10%	12%	8%	16%	9%
% With LTV>98	6%	2%	8%	76%	51%
% Missing LTV Information	0%	0%	0%	0%	0%

Exhibit 14 (cont.): Analysis of 80-100% LTV Loans In FHA and GSE Market Sectors
On Matched, Conforming, Fixed-Rate FHA-Eligible FHA & GSE Loans with LTV 80-100%, Weighted
(n=114,780; weighted sample=674,238)

	GSE Purchased Loans			Not Mutually Exclusive	
	All GSE	GSE With PMI	GSE No PMI	All FHA	All Subprime*
	pur_type=1,3			fha_loan=1	subprime=1
Average Ratio of Loan Amount to FHA Loan Limit	69%	68%	69%	62%	57%
% With LoantoFHA Ratio <=.5	19%	20%	18%	28%	41%
% With LoantoFHA Ratio of .5 - 1	81%	80%	82%	72%	59%
% With LoantoFHA Ratio of 1 - 1.2	0%	0%	0%	0%	0%
% With LoantoFHA Ratio of >1.2	0%	0%	0%	0%	0%
Average PTI	20%	20%	20%	21%	20%
% Originated in 1998	33%	34%	33%	30%	26%
% Originated in 1999	33%	32%	33%	36%	34%
% Originated in 2000	34%	34%	34%	34%	40%
Mortgaged Property Characteristics					
% New Construction	8%	7%	9%	8%	4%
% Unit Size 1	94%	94%	94%	95%	94%
% Unit Size 2	2%	2%	2%	2%	2%
% Unit Size 3	3%	3%	3%	2%	3%
% Unit Size 4	2%	2%	2%	1%	2%
Borrower Neighborhood Characteristics (1990 Census)					
% In Underserved Tracts	31%	33%	29%	49%	56%
% In High Cost MSA	11%	9%	12%	11%	15%
% In Medium Cost MSA	86%	87%	85%	86%	83%
% In Low Cost MSA	3%	3%	2%	3%	1%
% In Center City	29%	31%	27%	29%	39%
Average 5-yr House Price Appreciation, Lagged 1 year	113%	114%	113%	111%	108%
% In Area with Depreciation	11%	10%	11%	12%	16%
% In Area with Appreciation up to 20%	74%	73%	75%	75%	79%
% In Area with Appreciation over 20%	15%	17%	14%	13%	5%
% In Tracts with Income <90% of MSA Income	22%	24%	20%	33%	44%
% In Tracts with Income 90 - 120% of MSA Income	37%	38%	37%	40%	38%
% In Tracts with Income =>120% of MSA Income	41%	38%	43%	26%	18%
% In <10% Minority Tracts	41%	42%	41%	30%	32%
% In 10-30% Minority Tracts	35%	34%	35%	33%	31%
% In =>30% Minority Tracts	24%	25%	23%	37%	37%

Notes:
* The subprime classification in our analysis is not mutually exclusive from other categories.

The average annual income[11] of the borrowers among all of the matched conforming[12] loans is over $63,000 (see Exhibit 10). However, the distribution of incomes in this data set suggests that a number of outliers artificially raise this average. Both a median income, as well as an average income derived from a trimmed data set is provided to offer a potentially more accurate picture of borrowers' incomes.[13] Additionally, the borrowers' average credit score, or FICO, is provided along with a distribution of FICO scores.[14] Breakdowns of the race, age, and sex of the borrowers are also displayed.

In addition, the data were classified based on several key characteristics of the loans. The average loan amount is calculated based on the HMDA loan amount variable. However, to get a better sense of the potential risk of the loans in each data set, several variables are provided, including the average loan to value ratio (LTV), the average ratio of loan to FHA loan limit, and the average payment to income ratio (PTI). A distribution of LTV and loan to FHA loan limit ratios is provided, as well as a breakdown of originations by year.

The loan-to-value ratio was calculated by dividing the Experian loan amount by the Experian sale amount (i.e., purchase value). The ratio of loan amount to a rounded[15] FHA loan limit is calculated by comparing the HMDA loan amount to the corresponding FHA loan limit based on origination date, assigned dwelling size,[16] and MSA of origination. In the first column of Exhibit 10, the 51 percent of loans with LTV greater than 98 percent may seem high, but the sample of FHA-eligible loans is dominated (65 percent) by FHA loans. Only 6 percent of GSE loans in the FHA-eligible sample have such high LTVs and only 2 percent of the full sample of GSE loans has LTVs above 98 percent.

The payment-to-income ratio is a measure of the annual loan payment relative to the annual income of the borrower. In order to calculate the payment-to-income ratio, the annual payment of the mortgage must be estimated. It is assumed that all loans were 30-year mortgages with a fixed interest rate based on the national average contract interest rate for fixed-rate mortgages in the month of sale.[17] This estimated annual payment is then compared to the borrower's annual income. (In fact,

---

[11]   Annual income is based on the HMDA Annual Income variable. However, roughly two percent of the loans were missing annual income information. All but seven of these loans did, however, have an income range code provided in the Experian Income variable. The midpoint of each range was used as an estimate of the borrowers' income. Those in the top income range of >$250,000 were assigned an income of $250,000.

[12]   Conforming loans in this context refers to loans below the GSE conforming loan limit (i.e., non-jumbo loans), which was $252,700 in 2000.

[13]   The outliers are not excluded from the rest of the analysis.

[14]   FICO is not collected in HMDA and is only available because of the merging of HMDA with the Experian data.

[15]   The HMDA loan amounts are reported rounded to the nearest $1,000, so the FHA loan limits are rounded up to the nearest thousand to avoid inadvertently classifying loans as ineligible for FHA.

[16]   See above discussion of dwelling size assignments.

[17]   Contract Interest Rates-Monthly National Averages for All Homes, Fixed-Rate Mortgages in 1998, 1999, and 2000. Federal Housing Finance Board Monthly Interest Rate Survey: www.fhfb.gov/mirs/mirs_downloads.htm.

17% of the matched conforming loans are not on fixed-rate terms. The above annual payment estimation does not capture the changing annual payment associated with adjustable-rate or balloon loans).

Highlighted mortgaged property characteristics include the share of loans that are for new construction as well as the assumed unit size.

In addition to the above characteristics of the borrowers, loans, and properties, information is provided about the metropolitan areas and neighborhoods of the mortgaged properties, including tract "underserved area" status, relative cost of the MSA based on FHA eligibility limits, center city location, MSA property value growth, and tract minority percentage and tract income relative to the MSA median.

In the matched, conforming data set, 27 percent of the loans were originated in center city tracts[18] and 32 percent were originated in tracts that were "underserved" [19] in the 1990 Census.

The distribution of loans into high, average and low cost MSAs is based on whether loans originated in a MSAs with the highest or lowest FHA loan limits at the time of origination.[20] Only three percent of the loans in the data originated in "Low Cost" MSAs, meaning MSAs that had the lowest FHA loan limits at the time of origination.

In order to evaluate whether loans originated in areas of high or low property value growth, the average 5-year house price change (lagged one year) was calculated. MSAs with higher and lower levels of appreciation were based on the OFHEO House Price Index.[21] Over 70 percent of loans in the largest analysis file were originated in MSAs with property value appreciation up to 20 percent over five years (lagged one year) compared to only 11 percent of loans originated in areas with negative appreciation over five years.

In addition, 26 percent of the loans were originated in tracts with a population at least 30 percent minority in the 1990 Census, and 23 percent were originated in tracts with a median household income that was less than 90 percent of the MSA median household income.

Exhibit 11 examines the pool of loans associated with each mortgage market "sector" in the matched conforming data. As described above, the GSE sector is based on the HMDA secondary purchaser type (Purchaser Type=1 or 3). The HMDA purchaser type is also used to determine the loans that were held by depository lenders (Purchaser Type =5, 6, 8 or 0) and those that were sold to other types

---

[18]   Tracts were assigned as "center city" tracts by Unicon based on the 1990 Census.

[19]   Tracts are designated by HUD as underserved if they have a median household income no more than 90 percent of the MSA median household income or if they have a population that is at least 30 percent minority with a median household income that is no more than 120 percent of the MSA median household income.

[20]   In 1998, FHA loan limits varied within an MSA by county. From 1999 on, HUD indexed the base FHA loan limit at 48 percent of the conforming loan limit and the maximum FHA one-family loan limit for "high cost" areas at 87 percent of the conforming loan limit, depending on the median house price in the county or MSA. FHA loan limits from FHA_limits_holly.xls.

[21]   OFHEO House Price Index, Q3 1993.

of investors (Purchaser Type =7 or 9). Additionally, the loans are segmented into those with private mortgage insurance (PMI) or with FHA insurance, and those originating with lenders designated as primarily subprime lenders in the year of origination.[22] It is important to note that some loans in the last three columns of Exhibit 11, labeled "Not Mutually Exclusive," are also included in the first nine columns. In other words, there are a few FHA loans in GSE portfolios or held by depositories. There are also FHA loans originated by lenders designated as subprime lenders. In the context of Exhibit 11, those loans are allowed to appear in both categories. In the origination models, a mutually exclusive categorization is created by excluding FHA loans from the GSE category or excluding subprime loans from the FHA category.

Using this method of segmentation, the GSE sector has borrowers with the highest FICO scores (average FICO of 726) and the highest incomes (median income of $62,000), and the loans with the lowest LTV ratios (average LTV of 80 percent). In contrast, FHA-insured borrowers have the lowest annual incomes (median income of $45,000) and the loans with the highest LTV ratios (average LTV of 97 percent). The pool of FHA-insured loans also contains the highest proportion (48 percent) of loans originated in underserved tracts.

Subsetting the Data

In order to evaluate the loans with the greatest likelihood of being in the FHA and GSE overlap, the analysis first excluded the 61,847 adjustable rate loans from the matched conforming data set (see Exhibit 12). This choice was made, in part, because the PTI calculation for this set of loans would not capture the changing annual payments over time and there was a desire for a more consistent set of loans for this analysis. Furthermore, these loans should be modeled separately because they are not directly comparable to fixed-rate loans. However, the set of adjustable loans is already fairly small and, after cutting it into competing market sectors, would create subsets that were too small for reliable results.

An examination of the loans with adjustable rates shows that these loans tend to go to borrowers with higher incomes and lower FICO scores, tend to be for a larger loan amount, and are more likely to originate in high cost MSAs than fixed-rate loans. For example, California is a high-cost area with a disproportionate share of ARMs. Borrowers prefer ARMs in high-cost areas because, given a certain income, they can qualify for a larger loan amount under an ARM than a FRM. Overall, the remaining subset of fixed-rate loans has very similar characteristics to the full matched conforming set.

The matched conforming loans were further subdivided to include only fixed-rate, FHA-eligible loans, that is, loans under the corresponding FHA loan limit. This remaining set of conforming, fixed-rate, FHA-eligible loans contains 238,158 observations (See Exhibit 10). The loans in this pool tend to go to borrowers with lower incomes and that are more likely to be female than those in the set of all conforming loans. The loans in this set are also smaller, on average. However, segmenting this data into mortgage market sectors reveal that the pool of FHA-insured loans in the set of fixed-rate, FHA-eligible loans is not substantially different from the pool of FHA-insured loans in the set of all conforming loans (see Exhibit 13). But, not surprisingly, the characteristics of the loans in the other sectors changed more dramatically. For example, the median income of GSE borrowers dropped

---

[22] Subprime data were downloaded from www.huduser.org/datasets/manu.html and merged into this study's data on compressed Agency and Resp_ID variables. Subprime=1 if lender was classified as a primarily subprime lender in the year of loan origination.

from $62,000 to $54,000, the average GSE loan amount dropped from $139,103 to $118,706, and the GSE average ratio of loan amount to FHA loan limit dropped from 80 to 66 percent.

While all of these loans are eligible for FHA, the fixed-rate, FHA-eligible loans were further subdivided to include only FHA and GSE loans with an LTV of 80–100 percent in order to focus on the loans most likely to be considered by both FHA and the GSEs. The 114,780 loans in this data set have an average FICO score of 668, down substantially from the average of 696 in the set of all matched, conforming loans (see Exhibit 10). The average LTV of 95 percent is appreciably higher than the average of 84 percent in the larger set. Furthermore, 43 percent of the pool of FHA and GSE loans in the 80-100 percent LTV range originated in underserved tracts. In contrast, only 32 percent of all matched, conforming loans originated in underserved tracts. The borrowers in the FHA/GSE working data set also have lower incomes, are more likely to be a minority, and are younger than the borrowers in the larger set of all conforming loans.

Overall, the pool of loans in the FHA and GSE data sets appear to have higher potential risk characteristics. The pool of FHA-insured loans in this data set (see Exhibit 14) does not appear to be significantly different from the pool of FHA-insured loans in the larger data set.[23] However, the pool of GSE loans in this data set is substantially different from GSE loans in the larger set along many different dimensions. FICO scores are 13 points lower, loans are over $14,000 smaller, the average LTV increased from 80 to 90 percent, the average ratio of loan amount to FHA loan limits fell from 80 to 69 percent, and the share of GSE loans originating in underserved tracts increased from 22 to 31 percent.

Given the focus on overlap, it is interesting to see how much overlap there is in credit scores and LTV. Exhibit 15 shows the high degree of overlap in FICO scores (94 percent by the non-parametric overlap method explained below or a KS statistic[24] of 0.37). Certainly, GSE loans have higher FICO scores on average, but there are many FHA loans with higher FICO scores than some GSE loans. Exhibit 16 shows much less overlap in the LTV distributions. The GSE loans are bunched around 80, 90 and 95 percent LTV, whereas FHA loans are almost entirely 96 percent and above. Many FHA borrowers who appear to be low-risk based on credit score, may actually be high-risk based on LTV. A regression model is needed to control for all these differences and to estimate the predicted probability of being FHA along a single dimension.

---

[23]   Note that Exhibit 14 includes only loans that are either GSE or FHA loans. The subprime loans listed are also either GSE or FHA loans. The data in Exhibit 14 are the same as those used for the FHA vs. GSE origination models to follow.

[24]   The KS (Kolmogorov-Smirnov) statistic measures the degree of separation between two distributions. It is sensitive to small differences in distributions, so all the KS statistics are significant. Larger values indicate more separation and correspond to less overlap.

Exhibit 15: Weighted Distribution of FICO Scores
In FHA & GSE Dataset
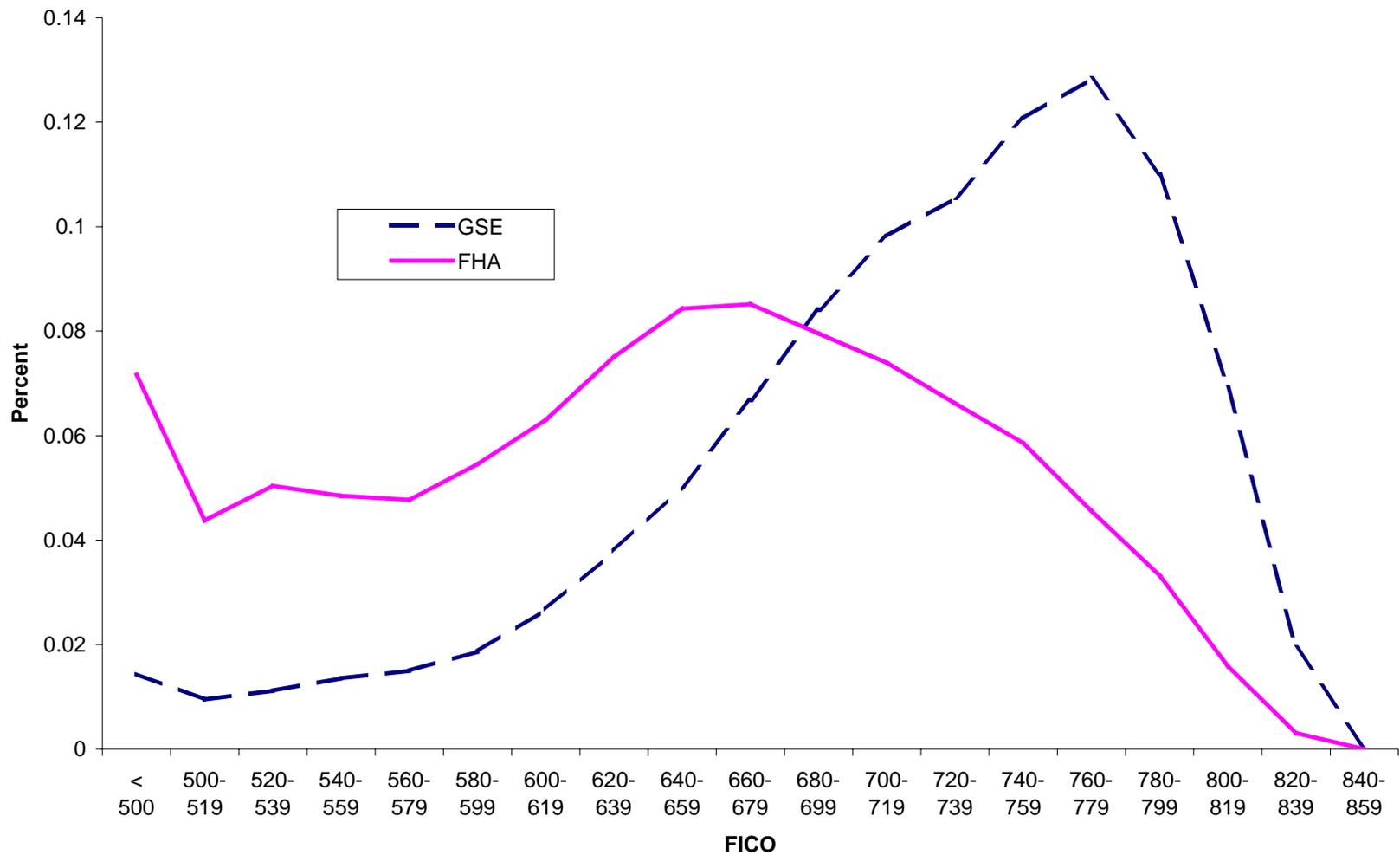
Exhibit 16: Weighted Distribution of LTV Ratios
In FHA & GSE Dataset
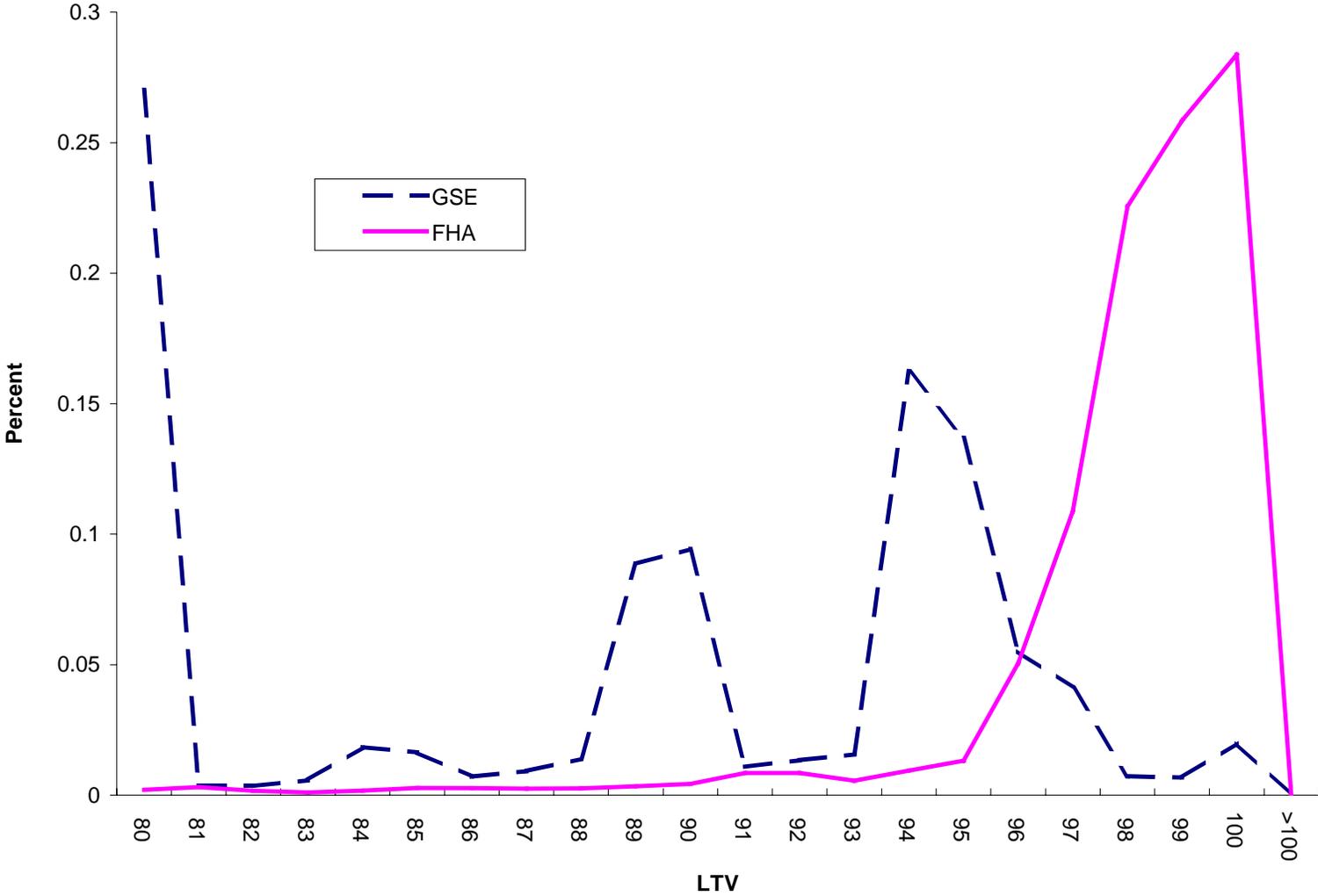
Section 3: Origination Model and Market Sector Overlap

To measure an overlap between FHA and GSE, or any two mortgage market sectors, it is useful to summarize all the different characteristics of a loan, borrower, property, and neighborhood into a single dimension. The origination model is a regression of the probability of a loan being insured by FHA relative to being purchased by a GSE, controlling for the observed characteristics that could significantly affect the predicted probability.[25] The predicted probability (or the log odds translation of the predicted probability[26]) is the single dimension summary that represents the FHA wedge described by Ambrose, Pennington-Cross, and Yezer (2002). If a sharp distinction exists between low-risk GSE loans and higher-risk FHA loans, there would be little overlap between the predicted probabilities for GSE and FHA loans. However, if loans with the same characteristics are nearly as likely to go to GSE as FHA, then there is considerable overlap. Perhaps the borrowers did not shop carefully or were steered into FHA when they could have qualified for a conventional loan. The immediate aim is not to explain why there is overlap, but rather to devise methods to measure the degree of overlap. This section describes three methods for measuring overlap and then applies those methods to the interfaces between FHA vs. GSE, FHA vs. PMI, FHA vs. subprime, and GSE vs. depository lenders.

Confidence Interval Measure of Overlap

The first method for measuring overlap is based on the confidence interval around the predicted probability. The predicted probability of FHA can range from zero to one and each prediction has a standard error for each prediction. A 95 percent confidence interval (+/-1.96 times the standard error) indicates the reliability of the prediction. If the confidence interval around the predicted probability of a loan being FHA contains zero (or nearly zero), the true probability of a loan being FHA is statistically indistinguishable from zero. To be 95 percent certain that a loan has a non-zero probability of being FHA, the lower bound of the confidence interval should be greater than zero. Conversely, if the confidence interval contains one, the true probability of a loan being FHA is not statistically distinguishable from one. In between those extremes are loans with confidence intervals that contain neither zero nor one. The loans that are statistically the same as definitely FHA or definitely GSE have been excluded. The remaining loans could have gone to either FHA or GSE and, thus, those loans are defined as the overlap region.

Technically, the confidence interval should not fall outside of the zero-one interval, but practically the interval measured as 1.96 times the standard error does fall below zero for some predictions with very low probabilities or above one for predictions with very high probabilities. The limit is arbitrary. For a more stringent definition of overlap, the boundaries for the confidence intervals could be 0.05

---

[25] To simplify the interpretation, only two alternatives are assumed for each model. It is further assumed that the results for those two alternatives would not be affected by the presence of other alternatives, that is, the independence of irrelevant alternatives (IIA).

[26] If p is the predicted probability, then $\log(p/(1-p))$ is the log odds. The log odds has a range from negative infinity to positive infinity compared to the range of the predicted probability of 0 to 1. Moreover, the distribution of log odds is more Gaussian or bell-shaped which makes it easier to visualize the overlap between distributions.

and 0.95 instead of 0 and 1. The boundaries of zero and one were chosen to emphasize the idea that, if there is 95 percent confidence that the probability "includes" zero of being FHA, then it is unlikely that FHA is competitive for that loan. The point is that confidence intervals "close" to or encompassing the endpoints (zero or one) indicate the extreme probabilities, either the loan is highly likely to be FHA or GSE. The intervals between the extremes indicate cases more evenly divided and less definite in the outcome. For example, the loans in the middle are nearly equally likely to go to either FHA or GSE. The confidence interval is used to trim off the cases with more definite outcomes leaving a set of overlap cases that have at least a modest chance of either outcome.

The boundaries are somewhat arbitrary, but the goal is to identify loans with probabilities evenly divided between the two possible outcomes. Those loans could have gone to either FHA or GSE which is the core idea of the overlap region. In this formulation, the boundaries chosen are that the lower limits of the confidence interval exceed 0 and the upper limits fall below 1. Other valid boundaries could have been chosen as a way to identify the loans with probabilities in the middle. In fact, the tolerance limit approach (explained below) is an example of a good alternative for defining the boundaries. The size of the overlap region is somewhat different, but the basic finding of a significant overlap region is robust to the method or boundaries chosen.

Note, the size of the overlap region is linked to the precision of the predictions and the goodness-of-fit of the origination model. A poor model fit, perhaps due to errors in the data, will lead to wide standard errors around the model predictions. All the predictions with confidence intervals reaching 0 or 1 are excluded from the overlap. Wide standard errors mean wide confidence intervals and more predictions being excluded from the overlap. Thus, a poor model fit is associated with a narrower overlap region. The "true" size of the overlap region may be larger, but limitations in the data or model specification can bias downward the estimate.

The diagram in Exhibit 17 shows overlapping distributions along the risk dimension. The GSE distribution falls mostly in the low-risk region and FHA distribution falls mostly in the higher-risk region. The portion of the distributions that overlaps corresponds to the range, shown in the lower panel, where the lower end of the confidence interval is greater than zero and the upper end of the interval is less than one.[27] As shown in the lower panel of Exhibit 17, the probability of a loan being FHA is aligned with higher default risk.

---

[27]   The lower panel of Exhibit 17 only shows the probability for FHA, whereas the upper panel shows the marginal distributions for both GSE and FHA. The probability of GSE would be 1 minus the probability of FHA and thus the mirror image of the FHA probability crossing at 0.5. The distinction to note is the upper panel shows the marginal distributions according to actual outcome, whereas the lower panel translates those outcomes into predicted probabilities in the same way that a logistic regression does. If there was perfect correspondence between actual outcomes and model predictions, then the CC' crossing point would correspond to the 50 percent probability.

Exhibit 17: FHA/GSE Overlap Based on the Confidence Interval Method
Distributions represented in percentage terms so equal in size.
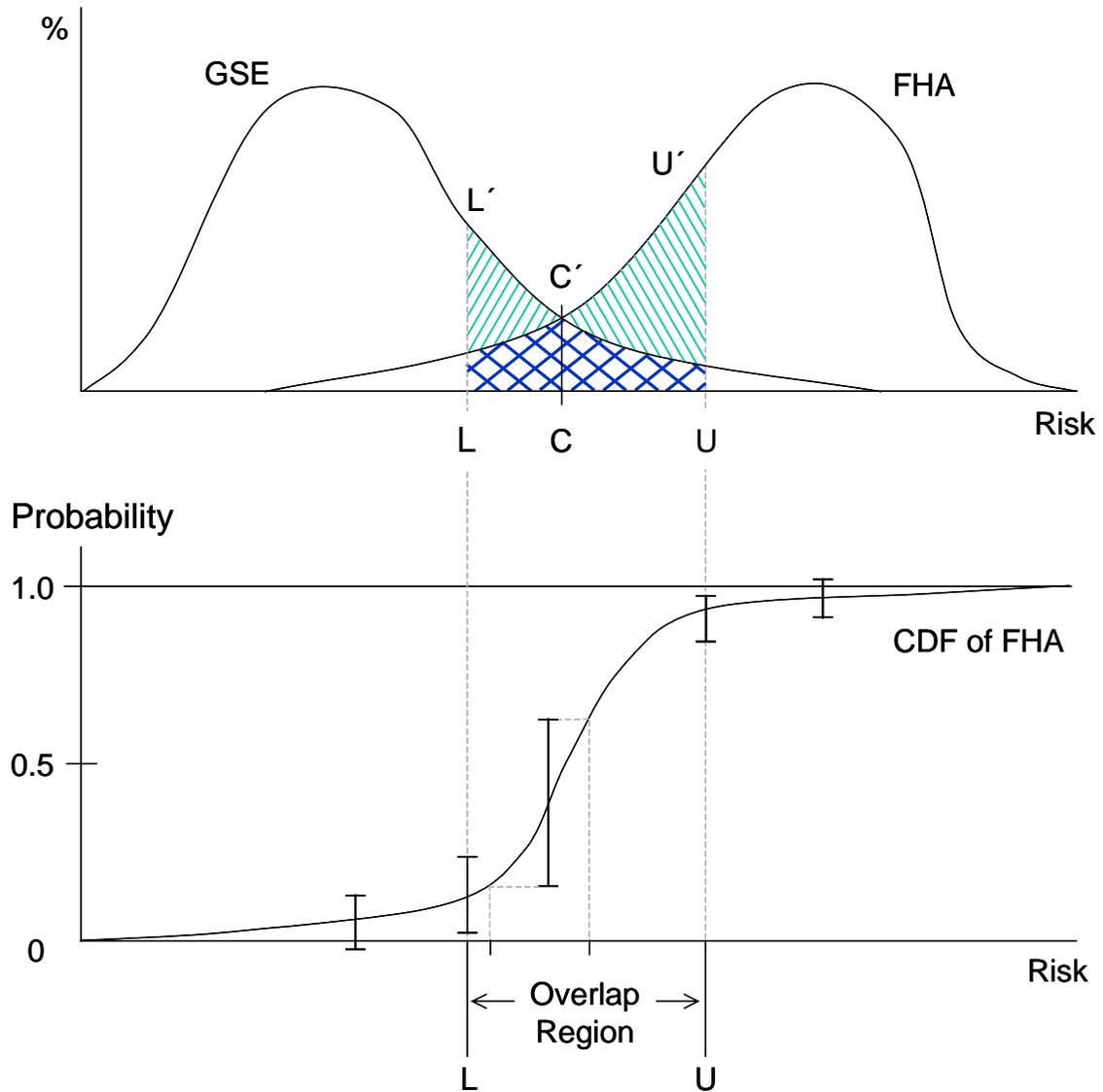
One disadvantage of the confidence interval method is that the width of the confidence interval is based on an arbitrarily chosen alpha value. In this case, an alpha of 0.05 was used, but that is based on the traditional preference of 95 percent confidence intervals used in a different context. Another issue is that a poor model with wide confidence intervals could have a very small overlap region, whereas greater uncertainty is expected to lead to greater overlap.[28] On the other hand, the tolerance limit methods (explained below) have the characteristic that poorly fitting models have greater

---

[28] As shown in Exhibit 19, the goodness-of-fit for the origination model for FHA vs. GSE is quite good (percent condcordant is 96.1, pseudo-R^2 is 0.64 and combined K-S statistic is 0.83).

overlap. In this respect, the confidence interval method is more conservative in its measure of overlap.[29]

A more important distinction between the methods is how the extreme values are handled. Consider again the overlapping distributions shown in Exhibit 17. The confidence interval method excludes the extreme high values of the FHA distribution (on the right) and the extreme low values of the GSE distribution (on the left). However, the high values of the GSE distribution and the low values of the FHA distribution may be close enough to the center of the combined sample that they are included in the overlap region. In bulk purchases, the GSEs do acquire some high-risk loans, and it is likely that some loans appear higher risk than they truly are due to errors in FICO score or LTV reported. For the purpose of determining which FHA loans could qualify for conventional underwriting, the overlap boundaries should not be distorted by the high-risk cases that do not represent typical GSE underwriting. Similarly, some low-risk FHA loans may be unusual cases or reporting errors. These outliers are unusual because they are not extremely high or low risks in the joint distribution, but rather extreme values relative to the marginal distributions, GSE or FHA. Therefore, a methodology is needed for trimming outliers from the marginal distributions before setting the overlap region boundaries. After the trimming, the overlap can be defined as the range between the FHA lower bound and the GSE upper bound.

Tolerance Limit Methods of Overlap

The concept of tolerance limits can help determine where the lower and upper bounds should be placed. Tolerance limits are statistical constructs designed to cover a fixed proportion of the population with a stated level of confidence using the sample data on hand. For example, a company that manufactures a particular engineering product can identify the tolerance limits such that, on average, a fixed proportion of the products may be expected to have a quality falling between the limits. Tolerance limits have also been used in the field of medicine. For instance, to identify the "normal" lower and upper limits for a particular physiological function (such as heart rate), a physician may base his/her estimates on a large sample of healthy subjects. Tolerance limits can be constructed so that a large proportion, say 90 percent, of the population will have a heart rate falling between the bounds with a stated level of confidence (such as 95%).

It is important to distinguish the concept of tolerance limits from *confidence intervals*. Confidence limits/intervals are statistical bounds within which a given population parameter, such as the mean, is expected to lie with a stated level of confidence. Tolerance limits, on the other hand, are statistical bounds within which a fixed proportion of the population is expected to lie with a stated level of confidence. In order to identify a set of typical cases, rather than the true value of a statistic, tolerance limits should be used. For example, the tolerance limits can be set so that 90 percent of the

---

[29] An alternative research approach has been suggested whereby mortgage markets are stratified according to FHA's competitive position. For example, if FHA garnered less than 5 percent of the loans, the market would be designated as low competitiveness for FHA, whereas if the FHA market share exceeded 25 percent then it could be designated as high competitiveness. Once the markets are defined as low, medium and high, then further analysis could be done to determine just which characteristics of the markets and borrowers are associated with each market segment. Moreover, the degree of potential FHA competitiveness can be determined by the predicted FHA probability. In a local market, if the actual market share deviates from the predicted share, this difference can lead to an investigation of the market in which FHA does particularly well or poorly. This approach is left for future research.

loans meet FHA underwriting requirements. By trimming the top and bottom 5 percent tails of the FHA distribution, the remaining loans represent the range of loans that meet the typical FHA underwriting requirements. The same process can be applied to the GSE distribution of loans to determine a 90 percent set of loans that meet the typical GSE underwriting requirements. Then the overlap is the set of loans that meet both the FHA and GSE underwriting requirements. Given that FHA underwriting allows for riskier loans than GSE underwriting, the overlap set is comprised of loans above the lower limit for FHA and below the upper limit for GSE.

There are two types of tolerance limits – parametric and non-parametric.[30] The parametric version assumes that the underlying distribution is normal, whereas the non-parametric makes no assumption about either the normality or symmetry of the distribution. The non-parametric version is preferred in the mortgage origination context because the distributions are neither normal or symmetric. Moreover, the non-parametric version efficiently trims off the extreme cases, which is important because the results should not be distorted by outliers or recording errors.

Parametric Tolerance Limits

The parametric tolerance limits require the strong assumption that the sample data were drawn from a normally distributed, i.e., symmetric, population. Below, the calculation formula for the limits is described. Readers interested in the derivation should consult the texts referenced in the footnote.

Suppose there are a series of measurements $Y_1, Y_2, ..., Y_N$. Let $\overline{Y}$ and S be the sample mean and sample standard deviation of the distribution. Then, the upper and lower tolerance limits that cover P percent of the population measurements with α level of confidence are:

$$Y_L = \overline{Y} - kS$$
$$Y_U = \overline{Y} + kS$$

where

$$k = \sqrt{\frac{(N-1)(1+\frac{1}{N})Z^2_{(1-p)/2}}{\chi^2_{\alpha,N-1}}}$$

$Z_{(1-p)/2}$ is the critical value if the standard normal distribution is exceeded with probability $(1-P)/2$ and $\chi^2_{\alpha,N-1}$ is the critical value of the chi-square distribution with N-1 degrees of freedom that is exceeded with probability α.

---

[30] This discussion is drawn heavily from the following sources: Howe, W. G. (1969). "Two-sided Tolerance Limits for Normal Populations - Some Improvements", *Journal of the American Statistical Association*, 64, pages 610-620. *Selected Techniques of Statistical Analysis for Scientific and Industrial Research and Production and Management Engineering*. Edited by Churchill Eisenhart et al. NY: McGraw-Hill Book (1947). *Sturdy Statistics: Nonparametric and Order Statistics*. Frederick Mosteller and Robert E.K. Rourke. MA: Addison-Wesley (1973).

Non-Parametric Tolerance Limits

Non-parametric tolerance limits relax the assumption that the sample data must come from a normally distributed population, particularly a symmetric distribution. The calculation is based on order statistics.[31] A parametric tolerance starts from the middle and goes outward plus or minus the same amount. A non-parametric tolerance starts at the extreme values and proceeds inward. Below, the derivation is skipped and only a sketch of the computation steps is provided.

Suppose there are a series of measurements $Y_1, Y_2, ..., Y_N$. Let Y_i and Y_{N-i+1} be the lower and upper limits that cover P percent of the population measurements with α level of confidence. According to the distribution theory of order statistics, the expected proportion of measurements in the interval (Y_i, Y_{N-i+1}) is $P = (N-2i+1)/N$, and the standard deviation of Y is $\sigma_Y = \sqrt{P(1-P)/(N+2)}$.

Let P* be the desired proportion and Z_α be the α-quantile from a standard normal distribution. To construct the tolerance limits Y_i and Y_{N-i+1}, the goal is to find the smallest P and the corresponding i such that:

$$\frac{p - p^*}{\sqrt{P(1-P)/(N+2)}} \geq |Z_\alpha| \tag{1}$$

and

$$P = (N-2i+1)/N \tag{2}$$

Equations (1) and (2) together form a system that is quadratic in P. Rather than solving this system of equations by formula, the optimal values of P and i are obtained by the method of iteration. Specifically, one starts with a very small value (.0001) for P and increases it by 0.0001 until the right-hand-side of equation (1) is greater than $|Z_\alpha|$. Once the optimal value of P is solved, the value of i can be obtained from equation (2).[32]

It is important to emphasize that the order statistics are applied to the predicted probabilities, which is a continuous distribution. The main purpose of the order statistics is to trim off the tails of the marginal distributions for FHA and GSE, or whichever two outcomes are competing. The remaining overlap of the trimmed distributions is based on "normal" underwriting conditions rather than the exceptional cases, which may dominate the tails. The method of non-parametric tolerance limits is robust in that it makes no assumptions about the underlying distributions for FHA or GSE. Just as with the confidence interval approach, the particular boundaries are arbitrary. Instead of the central 90 percent, the central 95 percent or the central 80 percent could have been chosen. Different order

---

[31] *Sturdy Statistics: Nonparametric and Order Statistics.* Frederick Mosteller and Robert E.K. Rourke. MA: Addison-Wesley (1973).

[32] Stata 8 is used to perform the iteration and search for the P and i values in this analysis. For the pooled sample of GSE and FHA loans that contain 114,780 observations, the optimal P and i values can be found in approximately 4 minutes of CPU time.

statistics would affect the size of the overlap. However, the main point is that a significant (although modest) overlap exists and the results are robust to different methodologies. Small changes in the definition of the boundaries would not eliminate the basic finding of an overlap region.

Origination Models and the Application of Overlap Methods

The first set of regression models form a bridge from the HUD 1995 Report by replicating the linear probability model using the original specification on 1998-2000 origination data. Exhibit 18 shows a progression of models. The first model replicates the specification used in the HUD 1995 Report (p. 6-24) on FHA-eligible loans with LTV between 80 and 95 percent. The R^2 in the replication (0.0427) is even lower than the original model (R^2=0.0927). This result is probably because even fewer FHA loans in 1998-2000 than in 1993 have loans with LTV below 95 percent. The coefficients in the replication model are somewhat different than the original model, but perhaps closer to expected. LTV is positive in the replication model, as opposed to the unexpected negative sign in the original model. The underserved indicator is still positive, but the component tract variables for income and percent minority are insignificant. The center city indicator is now negative. First-time homebuyer is not available in the new data. Otherwise, the signs and significance of the replication model are similar to the original model.

The next model in Exhibit 18 expands the range of loans to include all FHA-eligible loans with LTV greater than 80 percent. This change boosted the LTV coefficient more than tenfold and increased the model R^2 from 0.04 to 0.46. Switching to logistic regression brings few changes except the underserved indicator is no longer significant. Adding the credit score in the next model improves the model fit, as lower credit scores are strongly associated with FHA loans. Credit score also makes tract percent minority significant, attenuates the LTV coefficient and makes insignificant the ratio of loan to FHA loan limit. Incorporating the weights revives the tract income and percent minority variables and switches the sign on the high cost MSA indicator. The other changes and improvement in fit are modest.

Exhibit 18: Replication of FHA vs. PMI Model From HUD 1995 Report with Extensions to GSE Sector

| | | FHA and GSE loans that are FHA eligible and 80%<=LTV=<95%. Replication Model (Without FICO) | | Expand LTV range to LTV > 80% | | Logistic Regression Models | | | | | |
| | | | | | | Esimate using Logistic Regression | | Add Credit Score | | Incorporate Weights (obs_post_wght) | |
Parameter	Label	Parameter Estimate	Pr > |t|	Parameter Estimate	Pr > |t|	Estimate	Pr > ChiSq	Estimate	Pr > ChiSq	Estimate	Pr > ChiSq
Intercept	Intercept	-0.2216	<.0001	-3.7546	<.0001	-53.1352	<.0001	-46.2546	<.0001	-46.9196	<.0001
ex_fico	FICO Score							-0.0064	<.0001	-0.0059	<.0001
age_3549	Borrower Aged 35-49	-0.0071	0.0418	-0.0095	0.0024	-0.0534	0.0365	-0.0829	0.0015	-0.0794	<.0001
age_5064	Borrower Aged 50-64	-0.0169	0.0009	-0.0218	<.0001	-0.1273	0.0024	-0.1118	0.0089	-0.0786	<.0001
age_65	Borrower Aged 65 and up	-0.0544	<.0001	-0.0429	<.0001	-0.3703	<.0001	-0.2541	0.0048	-0.1637	<.0001
loantoFHA	Ratio of Borrower's Loan to FHA Loan Limit	-0.0903	<.0001	-0.0903	<.0001	-0.3929	<.0001	-0.1031	0.2116	-0.291	<.0001
relinc_c90t	Census Tract family income relative to MSA median family income in 1990	-0.0131	0.1129	-0.0102	0.2097	-0.0624	0.3638	-0.0525	0.455	-0.0754	0.0093
c90t_pminorty	Percent minority in tract in 1990	-0.00005286	0.5976	0.00008526	0.2798	-0.0005	0.4337	-0.0016	0.0225	-0.0013	<.0001
ccity_n	Center City tract indicator	-0.0325	<.0001	-0.0370	<.0001	-0.2427	<.0001	-0.2375	<.0001	-0.2264	<.0001
served_n	Tract Status, 1=Underserved, Omitted=Served	0.0103	0.0371	0.0159	0.0003	0.0493	0.1677	0.0286	0.4337	-0.0287	0.0653
EX_LTV	Loan to Value Ratio	0.0040	<.0001	0.046	<.0001	0.5589	<.0001	0.5317	<.0001	0.5364	<.0001
PTI_calc	Payment to Income Ratio- Calculated	0.1331	<.0001	0.0592	<.0001	0.3479	0.0765	0.3281	0.091	0.1694	0.0114
BO_RATIO	Borr Inc rel. to MSA median	-0.0099	0.0002	-0.0380	<.0001	-0.4224	<.0001	-0.4356	<.0001	-0.44	<.0001
race_black	African American Borrower	0.0944	<.0001	0.1333	<.0001	0.7849	<.0001	0.5307	<.0001	0.5548	<.0001
race_hisp	Hispanic Borrower	0.0580	<.0001	0.0978	<.0001	0.5929	<.0001	0.4809	<.0001	0.4423	<.0001
race_other	Other Borrower (Native American, Asian, or Other)	-0.0105	0.1101	-0.0257	0.0002	-0.2123	0.0004	-0.2663	<.0001	-0.3088	<.0001
female	Female Borrower	0.0179	<.0001	0.0152	<.0001	0.1209	<.0001	0.1120	<.0001	0.0916	<.0001
high_cost	High Cost City based on FHA Loan Limits	-0.0224	<.0001	-0.0505	<.0001	-0.2974	<.0001	-0.2491	<.0001	0.1132	<.0001
low_cost	Low Cost City based on FHA Loan Limits	0.0608	<.0001	0.0508	<.0001	0.388	<.0001	0.2225	<.0001	0.3298	<.0001
Adjusted R^2		0.0427		0.4621		pseudo R2=.5038		pseudo R2=.5239		pseudo R2=.5243	
Percent Concordant						94.2		94.2		94.2	
Percent Discordant						5.5		5.5		5.5	
Somers' D						0.887		0.887		0.888	

After considerable experimentation, the origination model presented in Exhibit 19 was selected. This origination model is the probability of FHA relative to GSE on the set of FHA-eligible loans with LTV between 80 and 100 percent. In weighted terms, 65.3 percent of the loans are actually FHA and the model predicted 96.1 percent correctly.[33] In terms of KS statistics, the value of 0.83 shows the model does a good job of distinguishing between the FHA and GSE distributions. The model uses marginal splines for more flexibility on the continuous variables. For example, higher credit scores reduce the odds of FHA by 3.3 percent (exp(-0.00328)=0.9967) for each 10 points increase in the FICO credit score. The odds ratios are reported in the next to last column. For scores above 620, there is an additional reduction of 5.1 percent for each 10 points. For high LTV loans, the effect of the splines is cumulative. LTV is positively related to the probability of FHA up to an LTV of 99 percent. The strong negative for the top spline offsets the cumulative effect of the other splines. For example, an increase in LTV from 99 to 100 percent would decrease the odds of FHA by 83 percent, which is calculated by summing the four LTV coefficients and exponentiating the total (exp(0.0441 + 1.2346 + 1.2502 - 4.3025) = 0.17 or an 83 percent reduction). This is certainly a large reduction in the odds, but the odds were 12.5 going from an LTV of 98 to 99 percent. Even after the 83 percent reduction, the odds of FHA are still 2.1 (12.5*0.17 = 2.1).

A graphical representation of the change in probability is shown in Exhibit 20. Looking at the solid line, the probability of FHA increases gradually as LTV increases from 80 to 95 percent, and then increases rapidly to reach a peak at 99 percent. The negative spline coefficient for LTV above 99 percent pulls down the probability of FHA slightly, but the overall impact of LTV is positive. The solid line assumes FICO and the other variables are held at their mean. Increasing FICO by one standard deviation shifts the curve lower (short dotted line). Conversely, decreasing FICO by one standard deviation (long dotted line) increases the probability of FHA, showing that FICO has a negative relation to the probability of FHA.

A disadvantage of splines is that the coefficient for one segment can be a large positive value while a neighboring segment is a nearly opposite negative value. The net effect may be small, even insignificant. The pattern is similar to a situation with highly correlated variables. For example, the PTI variable has knots at 20 and 28. The base variable, PTI_calc, has a large positive coefficient, 4.87. When combined with the marginal spline above 20 the total is 4.96, which almost exactly matches the negative coefficient for the top segment, -4.94. An increase in PTI from 28 to 29 would have a net effect of increasing the probability of being FHA by 2.6 percent.

---

[33] The high percent concordant indicates that the model does a good, but not perfect, job of predicting outcomes for the loans. A broad set of GSE and FHA loans have been included based on meeting the eligibility criteria for FHA loans. Most of the GSE loans have much lower risk than FHA loans based on LTV. The 1995 HUD study excluded the high LTV loans from the sample. Those loans are included in the current sample, not to boost the concordant rate of the model, but rather because there are sufficient numbers of GSE loans with high LTV. The percentage of such GSE loans is quite small, but the count of such loans is in the thousands. It no longer seems appropriate to exclude high LTV loans from the sample because some GSE loans have high LTV. The more important point is that a model that fits well is more reliable for identifying overlap loans, that is, loans with a high enough probability that they could have gone either way. The model does not suffer from omitted variables because most of the loans are correctly assigned. The overlap loans are the remaining loans that cannot be so easily assigned to one outcome or the other.

In general, the origination model has the expected results. The probability of FHA is increased by LTV, PTI, female, black, Hispanic, new construction, and low-cost MSA. Tract income has a positive coefficient, which is not expected. The probability of FHA is decreased by FICO, high loan amount relative to the FHA loan limit, age, other race, center city, high cost MSA and 5-year house price appreciation. In the presence of borrower race, the tract percent minority is not significant. It is somewhat surprising that new construction has a positive impact on FHA market share, but the new construction may be associated with starter homes, at least within the subset of FHA-eligible loans. New construction may also be a proxy for first time homebuyers. Most of the MSAs are high-cost, and the equation for high-cost areas in the 1995 HUD Report also had negative values for center city and tract income.

Exhibit 19: Origination Model Results on FHA vs. GSE Loans

FHA=1, GSE=0
On Matched, Conforming, Fixed-Rate FHA-Eligible FHA & GSE Loans with LTV 80-100%, Weighted
(n=114,780; weighted sample=674,238)

Parameter	Label	Full Model with Std Errs from SAS		Full Model with Robust Std Errs from STATA					
		Estimate	Pr > ChiSq	Robust Coefficient	Standard Error	P>z	Mean	Odds Ratio	Marginal Change (-SD/2 to +SD/2)
Intercept	Intercept	-1.9151	<.0001	-1.9151	0.5113	0.000		0.1473	
ex_fico	FICO Score	-0.0033	<.0001	-0.0033	0.0005	0.000	667.7813	0.9967	-0.0419
FICOs620	Marginal spline for FICO above 620	-0.0051	<.0001	-0.0051	0.0006	0.000	68.3338	0.9949	-0.0445
EX_LTV	Loan to Value Ratio	0.0441	<.0001	0.0441	0.0028	0.000	95.3128	1.0451	0.0373
LTVs95	Marginal spline for LTV above 95%	1.2346	<.0001	1.2346	0.0173	0.000	2.5360	3.4372	0.3577
LTVs98	Marginal spline for LTV above 98%	1.2502	<.0001	1.2503	0.1053	0.000	0.6877	3.4913	0.1392
LTVs99	Marginal spline for LTV above 99%	-4.3025	<.0001	-4.3026	0.1907	0.000	0.2339	0.0135	-0.2385
rounded loantoFHA	Ratio of Borrower's Loan to FHA Loan Limit	-1.9523	<.0001	-1.9523	0.0775	0.000	0.6449	0.1419	-0.0530
PTI_calc	Payment to Income Ratio- Calculated	4.8723	<.0001	4.8723	0.4503	0.000	0.2083	130.6243	0.0725
PTIs20	Marginal spline for PTI above .20	0.0896	<.0001	0.0896	0.8580	0.917	0.0317	1.0937	0.0012
PTIs28	Marginal spline for PTI above .28	-4.9366	0.7681	-4.9366	0.5855	0.000	0.0081	0.0072	-0.0566
age_3549	Borrower Aged 35-49	-0.0722	<.0001	-0.0722	0.0290	0.013	0.4921	0.9303	-0.0050
age_5064	Borrower Aged 50-64	-0.1266	<.0001	-0.1266	0.0430	0.003	0.1093	0.8811	-0.0055
age_65	Borrower Aged 65 and up	-0.3035	<.0001	-0.3035	0.0939	0.001	0.0199	0.7382	-0.0062
female	Female Borrower	0.1319	<.0001	0.1319	0.0297	0.000	0.3083	1.1410	0.0085
race_black	African American Borrower	0.5247	<.0001	0.5247	0.0495	0.000	0.1592	1.6899	0.0257
race_hisp	Hispanic Borrower	0.3679	<.0001	0.3679	0.0426	0.000	0.1932	1.4447	0.0193
race_other	Other Borrower (Native American, Asian, or Other)	-0.3050	<.0001	-0.3050	0.0627	0.000	0.0475	0.7371	-0.0085
new_constr	New Construction	0.3318	<.0001	0.3318	0.0489	0.000	0.0793	1.3935	0.0129
relinc_c90t	Census Tract family income relative to MSA median family income in 1990	0.1294	<.0001	0.1294	0.0504	0.010	1.0711	1.1382	0.0055
c90t_pminorty	Pct minority	0.0002	0.4130	0.0002	0.0008	0.778	27.9421	1.0002	0.0008
ccity_n	Center City tract indicator	-0.2728	<.0001	-0.2728	0.0327	0.000	0.2912	0.7613	-0.0177
high_cost	High Cost City based on FHA Loan Limits	-0.1311	<.0001	-0.1311	0.0565	0.020	0.1096	0.8771	-0.0052
low_cost	Low Cost City based on FHA Loan Limits	0.0724	0.0499	0.0724	0.0701	0.302	0.0292	1.0750	0.0025
hpiL5	OFHEO House Price 5-yr change lagged 1 yr	-0.9216	<.0001	-0.9216	0.3001	0.002	1.1169	0.3979	-0.0160
year98	Loan Originated in 1998	0.1623	<.0001	0.1623	0.0342	0.000	0.3120	1.1763	0.0097
year00	Loan Originated in 2000	-0.3526	<.0001	-0.3526	0.0333	0.000	0.3415	0.7029	-0.0242
Baltimore	Property in Baltimore	0.4400	<.0001	0.4400	0.0733	0.000	0.0874	1.5528	0.0133
Chicago	Property in Chicago	-0.2612	<.0001	-0.2612	0.0664	0.000	0.1756	0.7701	-0.0121
Cleveland	Property in Cleveland	-0.0651	0.0791	-0.0651	0.1014	0.521	0.0305	0.9370	-0.0018
Denver	Property in Denver	0.4179	<.0001	0.4179	0.0663	0.000	0.0999	1.5188	0.0218
LosAngeles	Property in Los Angeles	0.0188	0.5796	0.0118	0.0703	0.867	0.1518	1.0119	0.0004
Oakland	Property in Oakland	-0.2842	<.0001	-0.2842	0.0904	0.002	0.0336	0.7526	-0.0068
Philadelphia	Property in Philadelphia	-0.0603	0.0022	-0.0603	0.0576	0.295	0.1088	0.9415	-0.0026
Portland	Property in Portland	-0.2656	<.0001	-0.2656	0.1372	0.053	0.0470	0.7667	-0.0111
StLouis	Property in St Louis	0.1995	<.0001	0.1995	0.0847	0.018	0.0336	1.2208	0.0067
Tampa	Property in Tampa	0.1756	<.0001	0.1756	0.0692	0.011	0.0709	1.1920	0.0083

Percent Concordant	96.1	
Percent Discordant	3.8	
Somers' D	0.923	
Pseudo R2		0.6369
Log Likelihood- Intercept only	870753.95	
Log Likelihood- Intercept and Covariates	316165.03	
Combined K-S Statistic		0.8255 (P-Value = 0.000)

**Exhibit 20: Higher LTV Has Positive Effect on Probability of FHA
Higher FICO Shifts Probability of FHA Lower**

LTV
**Predicted Probability of FHA Relative to GSE Based on LTV, FICO & Splines
(Other Variables Held at Mean)**
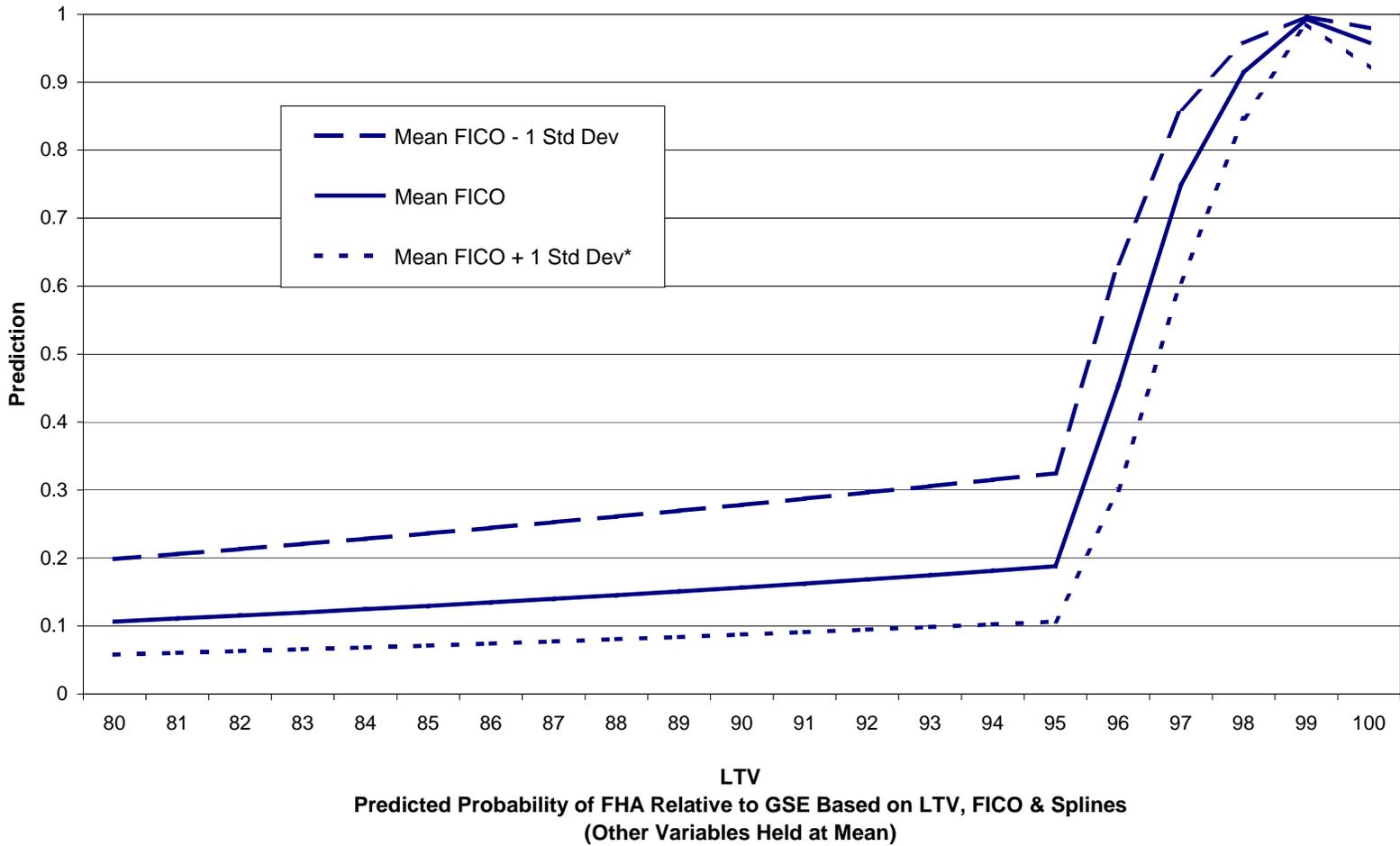
In Exhibit 19, the origination model results are shown for SAS and Stata. The only difference is that the Stata results show robust standard errors that are larger than the SAS standard errors, particularly for PTIs20. Curiously, PTIs28 becomes significant and low_cost becomes insignificant, but the coefficients are barely affected.

The far right column of Exhibit 19 presents the marginal changes in predicted probability from a standard deviation change for each variable holding the other variables at their means. Often marginal changes are based on a unit change, but the scale of the variables could mean the unit change is unusually large or small. To cope with variables of very different scales, the impact of one standard deviation change (from half a standard deviation below the mean to half a standard deviation above the mean of the selected variable) is presented. This approach makes it possible to identify which variables cause big changes in the predicted probability. For example, a standard deviation increase in FICO reduces the predicted probability of FHA by 4 percentage points. Given that most loans have FICO credit scores above the knot at 620, the negative effect of the spline should also be included, or a combined effect of –0.086. The LTV effect without the splines is somewhat smaller, a 3.7 percentage point increase in probability of FHA. However, most FHA loans have LTV above 95 percent. When the splines are included the predicted probability increases by 36 percentage points. This large jump in predicted probability corresponds to the steep slope above 95 percent LTV shown in Exhibit 20. The combined LTV effect clearly dominates the marginal changes from every other explanatory variable (or spline group). High LTV is a strong indicator of FHA loans relative to GSE loans and that is the primary reason the model fits the data so well.

After LTV and FICO, the next most important variable is PTI, which increases the probability of FHA by 7.3 percentage points. Most of that increase is offset above a payment-to-income ratio of 28 percent, but the net effect is still positive. Another large factor is borrower race. A standard deviation increase in black status raises the probability of FHA by 2.6 percentage points (1.9 percentage points for Hispanics). It may be more natural to use the odds ratio for indicator variables like race. A black borrower has a 69 percent increase in the odds of getting an FHA loan, and a Hispanic borrower has a 44 percent increase in the odds of getting an FHA loan. These results on low minority shares by the GSEs support the findings of Bunce (2000).

The main point is that the model fits the data quite well, and the predictions can be used for estimating the overlap. Exhibit 21 shows the pattern of predicted probability in the upper panel and the predicted log odds in the lower panel. The FHA distribution is distinct from the GSE distribution, but the left tail of the FHA distribution extends well into "GSE territory." Similarly, the right tail of the GSE distribution extends far into "FHA territory." Using the confidence interval method, the number of predictions with 95 percent confidence intervals that do not include zero (definitely GSE) or one (definitely FHA) is determined. Exhibit 22 shows that 20 percent of the loans from the pooled data is in the overlap region, split 11 percent for GSE and 9 percent for FHA. The breakdown of overlap loans by MSA is also shown in Exhibit 22, ranging from 7 percent overlap in St. Louis to 28 percent in Denver. In terms of just the FHA portfolio, the 9 percent of FHA in the overlap region of the pooled sample is equivalent to 60,681 loans or 14 percent of the FHA loans.[34]

---

[34] As showed in Exhibit 14 (column 4), there are 440,104 weighted number of FHA borrowers. The FHA overlap loans, 60,681, divided by total FHA loans, 440,104, equals 0.14 or 14 percent of FHA borrowers in the overlap region.

Exhibit 21

Distribution of Predicted Probability
of the GSE and FHA Loans
Origination Model Weighted

Distribution of Predicted Log Odds
of the GSE and FHA Loans
Origination Model Weighted
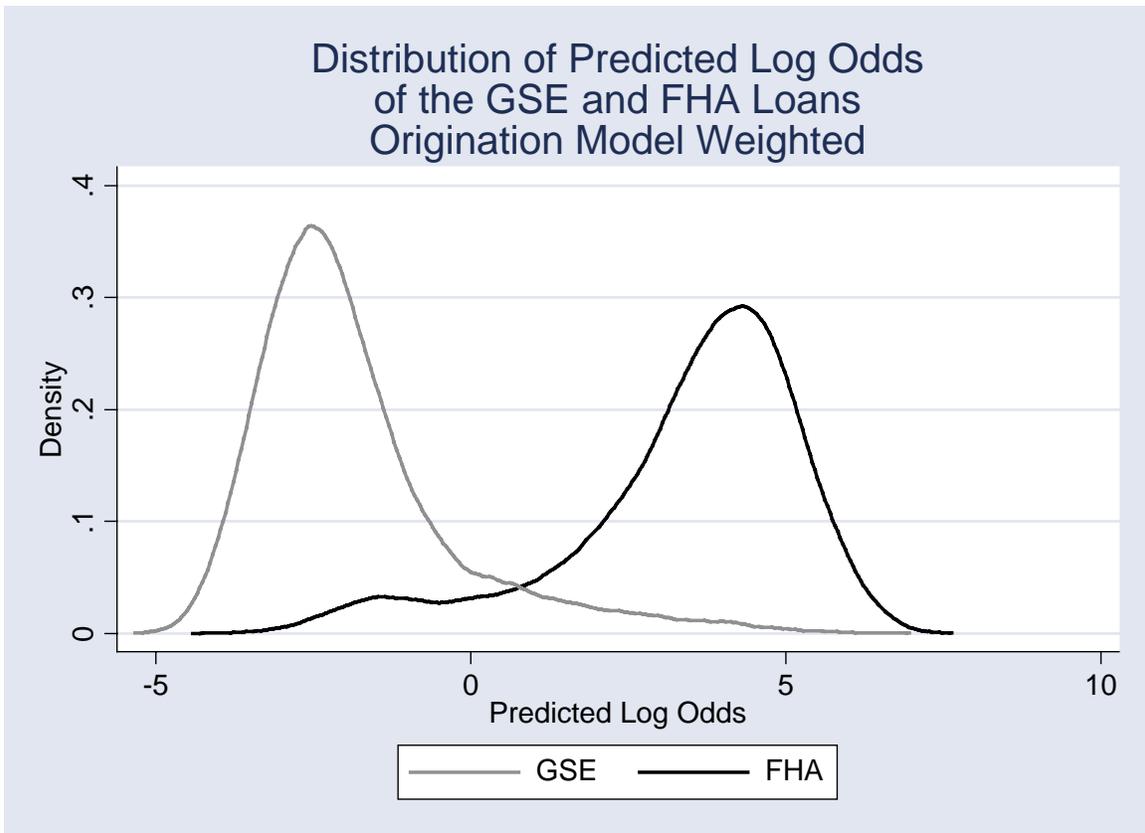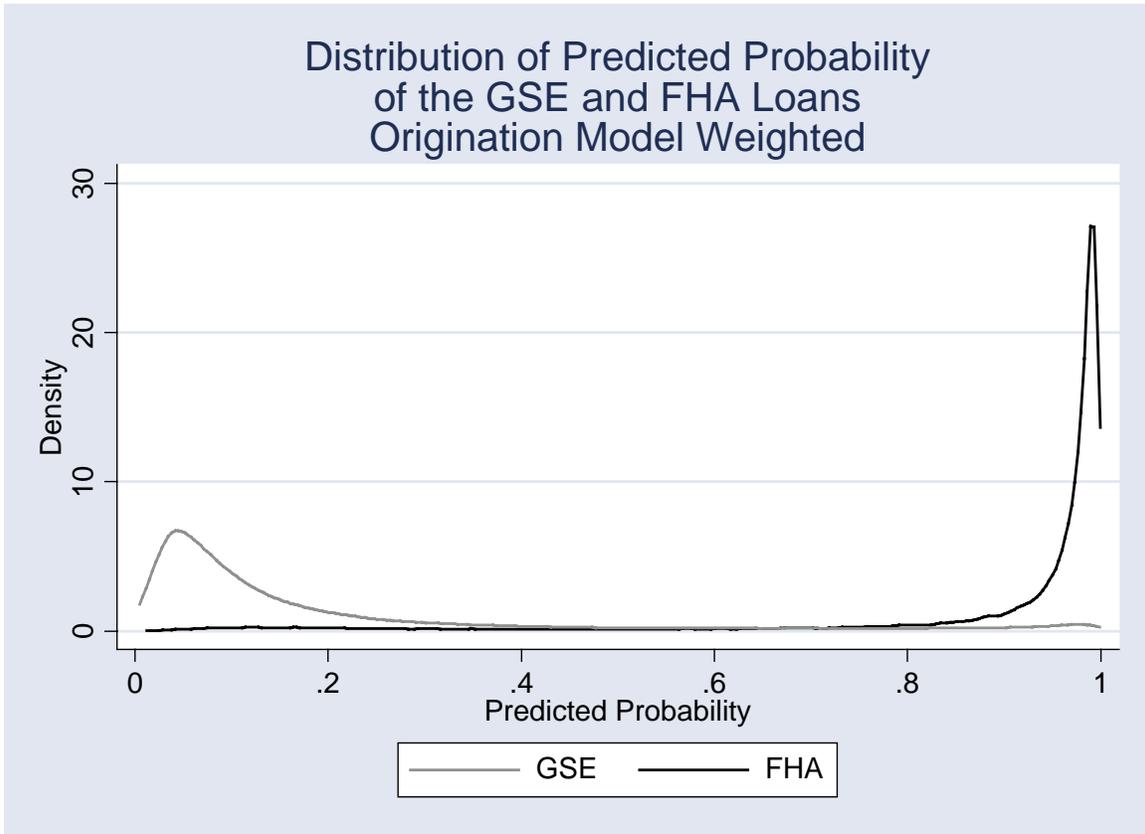
Exhibit 22: Percent of Loans in Overlap by Confidence Interval Method
FHA=1, GSE=0
On Matched, Conforming, Fixed-Rate FHA-Eligible FHA & GSE Loans with LTV 80-100%, Weighted
(n=114,780; weighted sample=674,238)

City	MSA	GSE (non-overlap)	GSE in Overlap	Combined Overlap	FHA in Overlap	FHA (non-overlap)
Baltimore	720	14%	7%	13%	6%	73%
Chicago	1600	38%	14%	24%	10%	38%
Cleveland	1680	37%	8%	13%	5%	50%
Denver	2080	17%	17%	28%	11%	55%
Los Angeles	4480	23%	10%	18%	9%	58%
Oakland	5775	43%	7%	15%	8%	42%
Philadelphia	6160	31%	12%	20%	8%	49%
Portland	6440	43%	12%	18%	6%	39%
St. Louis	7040	13%	5%	7%	2%	80%
Tampa	8280	21%	11%	19%	8%	60%
Washington DC	8840	18%	9%	19%	10%	63%
Pooled		23%	11%	20%	9%	57%

The comparison of characteristics for loans inside and outside the overlap region is presented in Exhibit 23. The values for the non-overlap regions clearly show lower risk for the GSE loans. For example, GSE borrowers have higher income, higher FICO, higher age and higher house price appreciation, whereas FHA borrowers have higher LTV and higher percentages in low-income and high minority areas. Nearly all those differences disappear (or reverse) in the overlap region. The median income for FHA overlap borrowers is actually $1,000 higher than GSE overlap borrowers. The income variable is sensitive to extreme values, so the median and trimmed mean are reported, as well as the overall mean. The higher FHA income and FICO may compensate for the slightly higher average LTV,[35] especially for the loans with LTV greater than 98 percent, or for the slightly lower house price appreciation.[36] Otherwise, the FHA overlap loans appear very similar to the GSE overlap loans.

The parametric tolerance interval method has a much smaller overlap region (2 percent) compared to the confidence interval method (20 percent), as shown in Exhibit 24. The reason is that the parametric tolerance interval assumes the distributions are normal and, as shown in Exhibit 21, the distributions are skewed. The comparison of characteristics in the overlap region shows few distinctions in Exhibit 25. The average and median income in FHA overlap is larger than in the GSE overlap. The loan amount in the GSE overlap is also larger than for FHA overlap, but the FICO and LTV are the same.

To avoid the problems created by skewed distributions, the non-parametric tolerance interval method is preferred. Exhibit 26 shows 11 percent overlap with 5 percent in GSE overlap and 7 percent in FHA overlap relative to the pooled sample. In terms of just the FHA distribution, 10 percent of FHA loans are in the overlap region. St. Louis still has the smallest number of overlap loans, but Washington, D.C., has replaced Denver as the MSA with the largest overlap. The non-parametric method uses order statistics to identify the central 90 percent for each distribution, which leaves 10 percent in the tails. In Exhibit 26, the Tails column is a combination of the left and right tails for both the FHA and GSE distributions. The characteristics of the overlap loans are shown in Exhibit 27. Again, the mean and median income in the GSE overlap is lower than in the FHA overlap. Moreover, the average FICO is also somewhat lower for GSE overlap (663 vs. 677). Even the loan amount and average LTV is lower for overlap GSE than overlap FHA, although the share of high LTV loans is larger for overlap FHA.[37] At least on average, it appears that the loans in the FHA overlap are as qualified for GSE purchase as the loans in the GSE overlap that were purchased by the GSEs.

---

[35]   The average LTV for FHA loans in the overlap region is 95 percent compared to 93 percent for GSE loans in the overlap region. While this difference in average LTV is small, the distribution of FHA loans clearly shows a higher share of high LTV loans. For example, 24 percent of GSE overlap loans compared to 61 percent of FHA overlap loans have LTV ratios above 96 percent.

[36]   The zero percent of GSE loans with LTV above 96 percent does not mean there are no GSE loans in that range, but rather the number of those loans are so small that it rounds down to zero when rounded to the nearest percentage unit.

[37]   In the overlap region, the average LTV for FHA loans is 96 percent compared to 95 percent for GSE loans. This near equality in average LTV does not highlight the substantial difference in share of high LTV loans. In this case, 54 percent of GSE loans compared to 75 percent of FHA loans have LTV ratios above 96 percent.

As a further drilldown, the probability of being FHA was estimated on the subset of non-parametric overlap loans. Given the similarities on average, it seemed possible that many of the explanatory variables would be insignificant. As it turned out, nearly all the coefficients had the same sign and significance, as shown in Exhibit 28. The first model is on the full set of FHA/GSE loans and the second model is on the subset of overlap loans. One interesting exception is that the base LTV variable became negative, though the splines retained the same sign with more extreme values. The other change is that tract percent minority became negative and significant. Not surprisingly, the model fit, in terms of percent concordant, weakens from 96.1 percent for the full model to 71.3 percent for the overlap loans.

Exhibit 23: Characteristics of Loans In and Out of Overlap by Confidence Interval Method
FHA=1, GSE=0
On Matched, Conforming, Fixed-Rate FHA-Eligible FHA & GSE Loans with LTV 80-100%, Weighted
(n=114,780; weighted sample=674,238)

	GSE (non-overlap)	GSE in Overlap	Combined Overlap	FHA in Overlap	FHA (non-overlap)
Share of Loans	23%	11%	20%	9%	57%
Borrower Characteristics					
Unweighted Number of Borrowers	26,016	13,180	22,883	9,703	65,875
Weighted Number of Borrowers	155,256	76,777	137,419	60,641	381,508
Average Annual Income	$65,621	$49,574	$51,207	$53,273	$47,525
Median Annual Income	$60,000	$46,000	$46,000	$47,000	$44,000
Average Annual Income (Trimmed Top 1%)	$63,508	$48,669	$49,185	$49,964	$46,162
% Estimated Income Information	1%	1%	1%	1%	1%
Average FICO	742	672	673	675	636
% With FICO <620	3%	22%	24%	25%	41%
% With FICO 620-680	10%	26%	25%	22%	25%
% With FICO =>680	87%	51%	52%	53%	34%
% Missing FICO Information	0%	0%	0%	0%	0%
% White	76%	65%	65%	65%	47%
% Black	3%	11%	11%	11%	23%
% Hispanic	7%	17%	17%	17%	25%
% Other	10%	4%	4%	4%	3%
% Missing Race Information	3%	3%	3%	3%	2%
% Female	24%	33%	34%	35%	32%
% Age 19-34	33%	38%	39%	39%	40%
% Age 35-49	49%	51%	50%	49%	49%
% Age 50-64	15%	10%	10%	11%	10%
% Age >65	4%	1%	1%	1%	2%
% Missing Age Information	0%	0%	0%	0%	0%
Loan Characteristics					
Average Loan Amount	$132,432	$113,570	$113,247	$112,837	$111,564
Average LTV %	87	93	94	95	99
% With LTV<=80	25%	4%	2%	0%	0%
% With LTV 80-90	38%	18%	15%	11%	0%
% With LTV 90-96	37%	54%	42%	27%	0%
% With LTV 96-98	0%	24%	39%	57%	11%
% With LTV>98	0%	0%	2%	4%	89%
% Missing LTV Information	0%	0%	0%	0%	0%
Average Ratio of Loan Amount to FHA Loan Limit	73%	62%	62%	62%	62%
% With LoantoFHA Ratio <=.5	13%	28%	29%	31%	28%
% With LoantoFHA Ratio of .5 - 1	87%	72%	71%	69%	72%
% With LoantoFHA Ratio of 1 - 1.2	0%	0%	0%	0%	0%
% With LoantoFHA Ratio of >1.2	0%	0%	0%	0%	0%
Average PTI	0.19	0.21	0.21	0.20	0.21
% Originated in 1998	33%	33%	33%	34%	30%
% Originated in 1999	31%	36%	34%	31%	37%
% Originated in 2000	37%	31%	33%	35%	34%

Exhibit 23 (cont.): Characteristics of Loans In and Out of Overlap by Confidence Interval Method

FHA=1, GSE=0
On Matched, Conforming, Fixed-Rate FHA-Eligible FHA & GSE Loans with LTV 80-100%, Weighted
(n=114,780; weighted sample=674,238)

	GSE (non-overlap)	GSE in Overlap	Combined Overlap	FHA in Overlap	FHA (non-overlap)
Mortgaged Property Characteristics					
% New Construction	8%	8%	8%	8%	8%
% Unit Size 1	94%	92%	93%	94%	96%
% Unit Size 2	2%	2%	2%	2%	2%
% Unit Size 3	3%	3%	3%	2%	2%
% Unit Size 4	1%	2%	2%	2%	1%
Borrower Neighborhood Characteristics (1990 Census)					
% In Underserved Tracts	27%	36%	36%	35%	51%
% In High Cost MSA	12%	9%	10%	10%	11%
% In Medium Cost MSA	85%	88%	87%	87%	86%
% In Low Cost MSA	3%	3%	3%	3%	3%
% In Center City	30%	26%	26%	25%	30%
Average 5-yr House Price Appreciation, Lagged 1 year	114%	113%	112%	111%	111%
% In Area with Depreciation	11%	11%	11%	12%	12%
% In Area with Appreciation up to 20%	71%	78%	78%	79%	74%
% In Area with Appreciation over 20%	18%	11%	10%	9%	14%
% In Tracts with Income <90% of MSA Income	20%	25%	25%	24%	35%
% In Tracts with Income 90 - 120% of MSA Income	37%	38%	38%	39%	41%
% In Tracts with Income =>120% of MSA Income	44%	37%	37%	37%	24%
% In <10% Minority Tracts	44%	38%	36%	32%	30%
% In 10-30% Minority Tracts	34%	35%	37%	40%	31%
% In =>30% Minority Tracts	21%	27%	27%	28%	39%

Exhibit 24: Percent of Loans in Overlap by Parametric Tolerance Interval Method
FHA=1, GSE=0
On Matched, Conforming, Fixed-Rate FHA-Eligible FHA & GSE Loans with LTV 80-100%, Weighted
(n=114,780; weighted sample=674,238)

City	MSA	GSE (non-overlap)	GSE in Overlap	Combined Overlap	FHA in Overlap	FHA (non-overlap)	Tails
Baltimore	720	20%	0.4%	3%	2%	69%	8%
Chicago	1600	47%	2%	3%	2%	42%	8%
Cleveland	1680	41%	1%	2%	1%	50%	8%
Denver	2080	32%	1%	2%	1%	58%	8%
Los Angeles	4480	30%	1%	2%	1%	60%	8%
Oakland	5775	44%	2%	4%	2%	44%	8%
Philadelphia	6160	40%	0%	0%	0%	51%	8%
Portland	6440	45%	3%	4%	1%	43%	8%
St. Louis	7040	17%	0%	0%	0%	78%	5%
Tampa	8280	29%	1%	2%	1%	60%	8%
Washington DC	8840	23%	3%	11%	9%	57%	8%
Pooled		31%	1%	2%	1%	58%	9%

Exhibit 25: Characteristics of Loans In and Out of Overlap by Parametric Tolerance Interval Method

FHA=1, GSE=0
On Matched, Conforming, Fixed-Rate FHA-Eligible FHA & GSE Loans with LTV 80-100%, Weighted
(n=114,780; weighted sample=674,238)

	GSE (non-overlap)	GSE in Overlap	Combined Overlap	FHA in Overlap	FHA (non-overlap)
Share of Loans (9% in Tails)	31%	1%	2%	1%	58%
Borrower Characteristics					
Unweighted Number of Borrowers	35,413	847	2,072	1,225	67,366
Weighted Number of Borrowers	210,835	4,914	12,986	8,072	390,553
Average Annual Income	$60,695	$51,657	$61,433	$67,385	$47,848
Median Annual Income	$56,000	$48,000	$49,000	$50,000	$44,000
Average Annual Income (Trimmed Top 1%)	$58,945	$50,363	$51,974	$53,098	$46,474
% Estimated Income Information	1%	1%	1%	1%	1%
Average FICO	721	693	694	694	640
% With FICO <620	9%	15%	17%	18%	39%
% With FICO 620-680	15%	18%	18%	19%	25%
% With FICO =>680	76%	67%	65%	63%	36%
% Missing FICO Information	0%	0%	0%	0%	0%
% White	73%	63%	64%	65%	48%
% Black	5%	13%	11%	10%	22%
% Hispanic	10%	16%	15%	15%	25%
% Other	8%	6%	6%	6%	3%
% Missing Race Information	3%	2%	3%	3%	2%
% Female	27%	35%	36%	37%	32%
% Age 19-34	34%	39%	39%	39%	40%
% Age 35-49	49%	50%	48%	47%	49%
% Age 50-64	14%	10%	12%	13%	10%
% Age >65	3%	2%	2%	2%	2%
% Missing Age Information	0%	0%	0%	0%	0%
Loan Characteristics					
Average Loan Amount	$126,796	$117,069	$116,253	$115,755	$111,798
Average LTV %	89	97	97	97	99
% With LTV<=80	19%	0%	0%	0%	0%
% With LTV 80-90	33%	0%	0%	0%	0%
% With LTV 90-96	44%	12%	12%	12%	0%
% With LTV 96-98	4%	87%	87%	87%	15%
% With LTV>98	0%	0%	0%	1%	85%
% Missing LTV Information	0%	0%	0%	0%	0%
Average Ratio of Loan Amount to FHA Loan Limit	70%	63%	62%	62%	62%
% With LoantoFHA Ratio <=.5	18%	27%	29%	30%	28%
% With LoantoFHA Ratio of .5 - 1	82%	73%	71%	70%	72%
% With LoantoFHA Ratio of 1 - 1.2	0%	0%	0%	0%	0%
% With LoantoFHA Ratio of >1.2	0%	0%	0%	0%	0%

Exhibit 25 (cont.): Characteristics of Loans In and Out of Overlap by Parametric Tolerance Interval Method

FHA=1, GSE=0
On Matched, Conforming, Fixed-Rate FHA-Eligible FHA & GSE Loans with LTV 80-100%, Weighted
(n=114,780; weighted sample=674,238)

	GSE (non-overlap)	GSE in Overlap	Combined Overlap	FHA in Overlap	FHA (non-overlap)
Average PTI	0.20	0.21	0.20	0.20	0.21
% Originated in 1998	33%	28%	32%	35%	30%
% Originated in 1999	33%	33%	31%	29%	36%
% Originated in 2000	34%	39%	37%	36%	34%
Mortgaged Property Characteristics					
% New Construction	8%	8%	9%	10%	8%
% Unit Size 1	93%	93%	92%	91%	96%
% Unit Size 2	2%	2%	2%	2%	2%
% Unit Size 3	3%	2%	3%	3%	2%
% Unit Size 4	2%	2%	3%	3%	1%
Borrower Neighborhood Characteristics (1990 Census)					
% In Underserved Tracts	29%	38%	37%	36%	50%
% In High Cost MSA	11%	12%	12%	12%	11%
% In Medium Cost MSA	86%	85%	86%	86%	86%
% In Low Cost MSA	3%	3%	2%	2%	3%
% In Center City	29%	24%	27%	29%	30%
Average 5-yr House Price Appreciation, Lagged 1 year	114%	114%	112%	111%	111%
% In Area with Depreciation	11%	9%	12%	14%	12%
% In Area with Appreciation up to 20%	73%	74%	76%	77%	74%
% In Area with Appreciation over 20%	16%	16%	12%	9%	14%
% In Tracts with Income <90% of MSA Income	21%	25%	24%	24%	34%
% In Tracts with Income 90 - 120% of MSA Income	37%	40%	38%	37%	40%
% In Tracts with Income =>120% of MSA Income	42%	36%	38%	39%	25%
% In <10% Minority Tracts	43%	33%	30%	29%	30%
% In 10-30% Minority Tracts	34%	38%	40%	41%	32%
% In =>30% Minority Tracts	23%	30%	30%	30%	38%

Exhibit 26: Percent of Loans in Overlap by Non-Parametric Tolerance Interval Method
FHA=1, GSE=0
On Matched, Conforming, Fixed-Rate FHA-Eligible FHA & GSE Loans with LTV 80-100%, Weighted
(n=114,780; weighted sample=674,238)

City	MSA	GSE (non-overlap)	GSE in Overlap	Combined Overlap	FHA in Overlap	FHA (non-overlap)	Tails
Baltimore	720	19%	2%	26%	24%	46%	10%
Chicago	1600	36%	10%	16%	6%	38%	10%
Cleveland	1680	33%	7%	12%	5%	46%	9%
Denver	2080	27%	5%	11%	6%	53%	9%
Los Angeles	4480	27%	3%	11%	8%	53%	9%
Oakland	5775	36%	9%	16%	7%	39%	9%
Philadelphia	6160	35%	4%	7%	4%	47%	11%
Portland	6440	33%	14%	18%	4%	40%	9%
St. Louis	7040	16%	0%	6%	5%	69%	10%
Tampa	8280	24%	5%	14%	9%	51%	10%
Washington DC	8840	20%	6%	27%	21%	44%	9%
Pooled		27%	5%	11%	7%	52%	10%

Exhibit 27: Characteristics of Loans In and Out of Overlap by Non-Parametric Tolerance Interval Method
FHA=1, GSE=0
On Matched, Conforming, Fixed-Rate FHA-Eligible FHA & GSE Loans with LTV 80-100%, Weighted
(n=114,780; weighted sample=674,238)

	GSE (non-overlap)	GSE in Overlap	Combined Overlap	FHA in Overlap	FHA (non-overlap)
Share of Loans (10% in Tails)	27%	5%	11%	7%	52%
Borrower Characteristics					
Unweighted Number of Borrowers	29,839	5,624	12,492	6,868	61,133
Weighted Number of Borrowers	180,184	32,173	76,438	44,265	349,001
Average Annual Income	$60,375	$48,907	$52,443	$55,014	$48,012
Median Annual Income	$56,000	$46,000	$46,000	$47,000	$44,000
Average Annual Income (Trimmed Top 1%)	$58,746	$47,931	$49,513	$50,769	$46,676
% Estimated Income Information	1%	0%	1%	4%	0%
Average FICO	726	663	671	677	642
% With FICO <620	6%	29%	27%	25%	38%
% With FICO 620-680	14%	23%	21%	20%	26%
% With FICO =>680	79%	48%	52%	55%	37%
% Missing FICO Information	0%	0%	0%	0%	0%
% White	75%	58%	61%	62%	49%
% Black	4%	15%	13%	12%	21%
% Hispanic	10%	19%	18%	17%	24%
% Other	8%	5%	5%	5%	3%
% Missing Race Information	4%	2%	3%	3%	2%
% Female	26%	35%	36%	36%	31%
% Age 19-34	34%	37%	38%	38%	40%
% Age 35-49	49%	51%	49%	48%	49%
% Age 50-64	14%	11%	11%	12%	10%
% Age >65	3%	1%	2%	2%	2%
% Missing Age Information	0%	0%	0%	0%	0%
Loan Characteristics					
Average Loan Amount	$127,204	$112,223	$112,789	$113,200	$111,718
Average LTV %	88	95	96	96	99
% With LTV<=80	19%	1%	1%	0%	0%
% With LTV 80-90	35%	8%	7%	6%	0%
% With LTV 90-96	45%	37%	27%	19%	0%
% With LTV 96-98	0%	53%	63%	71%	11%
% With LTV>98	0%	1%	2%	4%	89%
% Missing LTV Information	0%	0%	0%	0%	0%
Average Ratio of Loan Amount to FHA Loan Limit	70%	61%	61%	61%	62%
% With LoantoFHA Ratio <=.5	17%	31%	32%	32%	28%
% With LoantoFHA Ratio of .5 - 1	83%	69%	68%	68%	72%
% With LoantoFHA Ratio of 1 - 1.2	0%	0%	0%	0%	0%
% With LoantoFHA Ratio of >1.2	0%	0%	0%	0%	0%

Exhibit 27 (cont.): Characteristics of Loans In and Out of Overlap by Non-Parametric Tolerance Interval Method
FHA=1, GSE=0
On Matched, Conforming, Fixed-Rate FHA-Eligible FHA & GSE Loans with LTV 80-100%, Weighted
(n=114,780; weighted sample=674,238)

	GSE (non-overlap)	GSE in Overlap	Combined Overlap	FHA in Overlap	FHA (non-overlap)
Average PTI	0.20	0.21	0.20	0.20	0.21
% Originated in 1998	33%	32%	34%	35%	29%
% Originated in 1999	33%	34%	32%	30%	36%
% Originated in 2000	34%	34%	34%	35%	34%
Mortgaged Property Characteristics					
% New Construction	8%	9%	9%	9%	7%
% Unit Size 1	94%	93%	93%	93%	96%
% Unit Size 2	2%	2%	2%	1%	2%
% Unit Size 3	3%	3%	3%	2%	2%
% Unit Size 4	1%	2%	3%	3%	1%
Borrower Neighborhood Characteristics (1990 Census)					
% In Underserved Tracts	28%	41%	39%	37%	50%
% In High Cost MSA	11%	11%	12%	12%	10%
% In Medium Cost MSA	86%	86%	86%	85%	86%
% In Low Cost MSA	3%	3%	3%	2%	3%
% In Center City	28%	29%	27%	27%	31%
Average 5-yr House Price Appreciation, Lagged 1 year	113%	113%	112%	111%	111%
% In Area with Depreciation	11%	12%	13%	14%	11%
% In Area with Appreciation up to 20%	74%	73%	75%	76%	74%
% In Area with Appreciation over 20%	15%	15%	12%	9%	14%
% In Tracts with Income <90% of MSA Income	20%	29%	27%	25%	35%
% In Tracts with Income 90 - 120% of MSA Income	37%	37%	38%	38%	40%
% In Tracts with Income =>120% of MSA Income	43%	33%	35%	36%	25%
% In <10% Minority Tracts	44%	32%	30%	29%	31%
% In 10-30% Minority Tracts	34%	36%	38%	40%	32%
% In =>30% Minority Tracts	22%	32%	31%	31%	37%

Exhibit 28: Comparison of Origination Model Results for Loans in FHA/GSE Sectors and Subset of Loans in Non-Parametric Overlap

FHA=1, GSE=0
On Matched, Conforming, Fixed-Rate FHA-Eligible FHA & GSE Loans with LTV 80-100%, Weighted (n=114,780; weighted sample=674,238)

Parameter	Label	Full Model on Loans in All Sectors, with Std Errs from SAS		Full Model on Loans in Non-Parametric Overlap, with Std Errs from SAS		Difference Between Overlap and All-Sector Coefficients
		Estimate	Pr > ChiSq	Estimate	Pr > ChiSq	
Intercept	Intercept	-1.9151	<.0001	6.9622	<.0001	8.8773
ex_fico	FICO Score	-0.0033	<.0001	-0.0041	<.0001	-0.0008
FICO_620	Marginal spline for FICO above 620	-0.0051	<.0001	-0.00231	<.0001	0.0028
EX_LTV	Loan to Value Ratio	0.0441	<.0001	-0.0422	<.0001	-0.0863
LTVs95	Marginal spline for LTV above 95%	1.2346	<.0001	1.2662	<.0001	0.0316
LTVs98	Marginal spline for LTV above 98%	1.2502	<.0001	4.8222	<.0001	3.572
LTVs99	Marginal spline for LTV above 99%	-4.3025	<.0001	-10.8555	<.0001	-6.553
rounded loantoFHA	Ratio of Borrower's Loan to FHA Loan Limit	-1.9523	<.0001	-1.8969	<.0001	0.0554
PTI_calc	Payment to Income Ratio- Calculated	4.8723	<.0001	1.9326	<.0001	-2.9397
PTIs20	Marginal spline for PTI above .20	0.0896	<.0001	2.7266	<.0001	2.637
PTIs28	Marginal spline for PTI above .28	-4.9366	0.7681	-3.8012	<.0001	1.1354
age_3549	Borrower Aged 35-49	-0.0722	<.0001	-0.1018	<.0001	-0.0296
age_5064	Borrower Aged 50-64	-0.1266	<.0001	-0.0041	0.8813	0.1225
age_65	Borrower Aged 65 and up	-0.3035	<.0001	0.0354	0.5943	0.3389
female	Female Borrower	0.1319	<.0001	0.1878	<.0001	0.0559
race_black	African American Borrower	0.5247	<.0001	0.3132	<.0001	-0.2115
race_hisp	Hispanic Borrower	0.3679	<.0001	0.1279	<.0001	-0.24
race_other	Other Borrower (Native American, Asian, or Other)	-0.3050	<.0001	-0.3783	<.0001	-0.0733
new_constr	New Construction	0.3318	<.0001	0.366	<.0001	0.0342
relinc_c90t	Census Tract family income relative to MSA median family income in 1990	0.1294	<.0001	0.3453	<.0001	0.2159
c90t_pminorty	Pct minority	0.0002	0.413	-0.0024	<.0001	-0.0027
ccity_n	Center City tract indicator	-0.2728	<.0001	-0.1701	<.0001	0.1027
high_cost	High Cost City based on FHA Loan Limits	-0.1311	<.0001	-0.0484	0.1321	0.0827
low_cost	Low Cost City based on FHA Loan Limits	0.0724	0.0499	0.0908	0.1603	0.0184
hpiL5	OFHEO House Price 5-yr change lagged 1 yr	-0.9216	<.0001	-0.859	<.0001	0.0626
year98	Loan Originated in 1998	0.1623	<.0001	0.3843	<.0001	0.222
year00	Loan Originated in 2000	-0.3526	<.0001	-0.227	<.0001	0.1256
Baltimore	Property in Baltimore	0.4400	<.0001	0.2318	<.0001	-0.2082
Chicago	Property in Chicago	-0.2612	<.0001	-0.395	<.0001	-0.1338
Cleveland	Property in Cleveland	-0.0651	0.0791	-0.8052	<.0001	-0.7401
Denver	Property in Denver	0.4179	<.0001	0.2313	<.0001	-0.1866
LosAngeles	Property in Los Angeles	0.0188	0.5796	0.1008	0.0029	0.082
Oakland	Property in Oakland	-0.2842	<.0001	-0.3959	<.0001	-0.1117
Philadelphia	Property in Philadelphia	-0.0603	0.0022	-0.5975	<.0001	-0.5372
Portland	Property in Portland	-0.2656	<.0001	-1.2343	<.0001	-0.9687
StLouis	Property in St Louis	0.1995	<.0001	-0.4984	<.0001	-0.6979
Tampa	Property in Tampa	0.1756	<.0001	-0.2685	<.0001	-0.4441

Percent Concordant	96.1	71.3
Percent Discordant	3.8	28.5
Somers' D	0.923	0.428
Log Likelihood- Intercept only	870753.95	104044.08
Log Likelihood- Intercept and Covariates	316165.03	93220.287

To make the overlap loans so similar in overall risk, loans with low FICO scores are expected to have compensating low LTV or maybe high house price appreciation. Exhibit 29 shows three tabulations in which the overlap loans are divided in approximately thirds and the median value is reported for each subsection. In fact, there is a positive and significant correlation between FICO and LTV within the subset of overlap loans (correlation of 0.51 for FHA and 0.45 for GSE). Low FICO scores correspond to low LTV, and high FICO scores correspond to high LTV. The relation is far less obvious for house price appreciation over the last 5 years lagged one year (HPIL5), correlation of 0.11 for FHA and 0.16 for GSE. One likely problem may be that the house prices are measured at the MSA level, so there is relatively little variation even though there are 12,492 overlap loans. A better test would be to use Census 1990-2000 data at the tract level to get neighborhood house price changes.

Exhibit 29: Compensating Risk Factors for Overlap Loan by Non-Parametric Tolerance Interval Method
On Matched, Conforming, Fixed-Rate FHA-Eligible FHA & GSE Loans with LTV 80-100%, Weighted (n=12,492; weighted sample=76,438)

By FICO			FHA Loans		GSE Loans	
	Range	%	Median LTV	Median HPIL5	Median LTV	Median HPIL5
High FICO	>719	35%	97.16	1.09	96.91	1.14
Medium FICO	634 - 719	33%	96.80	1.14	96.55	1.14
Low FICO	<=634	32%	95.51	1.14	94.93	1.14

By LTV			FHA Loans		GSE Loans	
	Range		Median HPIL5	Median FICO	Median HPIL5	Median FICO
High LTV	>97	30%	1.11	741	1.14	734
Medium LTV	96 - 97	36%	1.14	691	1.14	717
Low LTV	<=96	34%	1.09	596	1.14	611

By HPIL5			FHA Loans		GSE Loans	
	Range		Median LTV	Median FICO	Median LTV	Median FICO
High HPIL5	>1.18	30%	96.83	684	96.61	680
Medium HPIL5	1.06 - 1.18	32%	96.84	686	96.47	670
Low HPIL5	<=1.06	37%	96.67	704	95.41	666

FHA vs. PMI within GSE

As shown in Exhibit 4, many GSE loans with LTV above 80 percent have private mortgage insurance, but not all. The same set of FHA loans are retained as in the previous section, but the GSE loans without PMI are excluded in order to get a more direct comparison between the two mortgage insurance sectors. In other words, PMI loans within GSE refers to GSE purchases of loans covered by private mortgage insurance. In the new data set of 94,135 loans, 76 percent of the loans are FHA. An origination model is estimated on the probability of a loan being insured by FHA, as shown in Exhibit 30. For most variables the coefficients are the same sign and significance. LTV is an exception, switching to negative and significant. However, the positive marginal splines for LTV at 95 and 98 are more positive, which offsets the negative coefficient on the base LTV variable. The tract percent minority variable also switched sign, but is only significant at the 9 percent level in terms of the robust standard errors. The goodness-of-fit is slightly lower. The percent concordant is the same at 96 percent and the pseudo-R^2 is 0.61 compared to 0.64 for FHA vs. GSE.

Exhibit 30: Origination Model Results on FHA vs. PMI within GSE

FHA=1, PMI=0
On Matched, Conforming, Fixed-Rate FHA-Eligible FHA or GSE+PMI Loans with LTV 80-100%, Weighted
(n=94,135; weighted sample=548,993)

| Parameter | Label | Full Model with Std Errs from SAS | | Full Model with Robust Std Errs from STATA | | | | | |
		Estimate	Pr > ChiSq	Robust Coefficient	Standard Error	P>z	Mean	Odds Ratio	Marginal Change (-SD/2 to +SD/2)
Intercept	Intercept	8.7061	<.0001	8.7061	0.6348	0.000		6039.3520	
ex_fico	FICO Score	-0.00289	<.0001	-0.0029	0.0006	0.000	655.8821	0.9971	-0.0091
FICOs620	Marginal spline for FICO above 620	-0.00673	<.0001	-0.0067	0.0008	0.000	59.8017	0.9933	-0.0139
EX_LTV	Loan to Value Ratio	-0.0604	<.0001	-0.0604	0.0042	0.000	97.1791	0.9414	-0.0082
LTVs95	Marginal spline for LTV above 95%	1.4682	<.0001	1.4682	0.0212	0.000	3.0345	4.3412	0.1166
LTVs98	Marginal spline for LTV above 98%	1.6900	<.0001	1.6900	0.1623	0.000	0.8297	5.4196	0.0491
LTVs99	Marginal spline for LTV above 99%	-4.8835	<.0001	-4.8835	0.3079	0.000	0.2810	0.0076	-0.0785
rounded loantoFHA	Ratio of Borrower's Loan to FHA Loan Limit	-2.0636	<.0001	-2.0636	0.0905	0.000	0.6336	0.1270	-0.0138
PTI_calc	Payment to Income Ratio- Calculated	1.5486	<.0001	1.5486	0.5365	0.004	0.2109	4.7048	0.0057
PTIs20	Marginal spline for PTI above .20	4.4408	<.0001	4.4418	1.0627	0.000	0.0324	84.9301	0.0144
PTIs28	Marginal spline for PTI above .28	-3.6186	<.0001	-3.6186	1.1547	0.002	0.0079	0.0268	-0.0104
age_3549	Borrower Aged 35-49	-0.1404	<.0001	-0.1404	0.0342	0.000	0.4931	0.8690	-0.0024
age_5064	Borrower Aged 50-64	-0.1615	<.0001	-0.1615	0.0503	0.001	0.1023	0.8508	-0.0017
age_65	Borrower Aged 65 and up	-0.2597	<.0001	-0.2597	0.1070	0.015	0.0173	0.7713	-0.0012
female	Female Borrower	0.1215	<.0001	0.1215	0.0350	0.001	0.3163	1.1292	0.0020
race_black	African American Borrower	0.4108	<.0001	0.4108	0.0601	0.000	0.1837	1.5080	0.0053
race_hisp	Hispanic Borrower	0.2927	<.0001	0.2927	0.0514	0.000	0.2160	1.3400	0.0040
race_other	Other Borrower (Native American, Asian, or Other)	-0.1521	<.0001	-0.1521	0.0758	0.045	0.0378	0.8589	-0.0010
new_constr	New Construction	0.5203	<.0001	0.5203	0.0570	0.000	0.0767	1.6825	0.0049
relinc_c90t	Census Tract family income relative to MSA median family income in 1990	0.3040	<.0001	0.3040	0.0606	0.000	1.0482	1.3552	0.0031
c90t_pminorty	Pct minority	-0.00162	<.0001	-0.0016	0.0010	0.090	29.6042	0.9984	-0.0015
ccity_n	Center City tract indicator	-0.3137	<.0001	-0.3137	0.0387	0.000	0.2972	0.7308	-0.0051
high_cost	High Cost City based on FHA Loan Limits	0.0581	0.0138	0.0581	0.0686	0.397	0.1060	1.0598	0.0006
low_cost	Low Cost City based on FHA Loan Limits	-0.0200	0.6467	-0.0200	0.0834	0.810	0.0303	0.9802	-0.0002
hpiL5	OFHEO House Price 5-yr change lagged 1 yr	-0.7483	<.0001	-0.7483	0.3530	0.034	1.1142	0.4732	-0.0031
year98	Loan Originated in 1998	0.2591	<.0001	0.2591	0.0404	0.000	0.3103	1.2957	0.0038
year00	Loan Originated in 2000	-0.2837	<.0001	-0.2837	0.0396	0.000	0.3375	0.7530	-0.0048
Baltimore	Property in Baltimore	0.1867	<.0001	0.1867	0.0920	0.042	0.0949	1.2053	0.0015
Chicago	Property in Chicago	-0.6718	<.0001	-0.6718	0.0780	0.000	0.1642	0.5108	-0.0074
Cleveland	Property in Cleveland	-0.6099	<.0001	-0.6099	0.1187	0.000	0.0298	0.5434	-0.0042
Denver	Property in Denver	0.0335	0.2535	0.0335	0.0789	0.671	0.1010	1.0340	0.0004
LosAngeles	Property in Los Angeles	-0.2984	<.0001	-0.2984	0.0841	0.000	0.1543	0.7420	-0.0027
Oakland	Property in Oakland	-0.4813	<.0001	-0.4813	0.1074	0.000	0.0290	0.6180	-0.0027
Philadelphia	Property in Philadelphia	-0.4717	<.0001	-0.4717	0.0699	0.000	0.1070	0.6239	-0.0050
Portland	Property in Portland	-0.8000	<.0001	-0.8000	0.1616	0.000	0.0429	0.4493	-0.0080
StLouis	Property in St Louis	-0.3654	<.0001	-0.3654	0.1027	0.000	0.0376	0.6939	-0.0032
Tampa	Property in Tampa	-0.2853	<.0001	-0.2853	0.0830	0.001	0.0739	0.7518	-0.0034

Percent Concordant	96		
Percent Discordant	3.8		
Somers' D	0.923		
Pseudo R2		0.6124	
Log Likelihood- Intercept only	546903.42		
Log Likelihood- Intercept and Covariates	211967.26		
Combined K-S Statistic		0.8250 (P-Value = 0.000)	

The overlap is pictured in Exhibit 31. The PMI distribution is rather flat for predicted probability in the top panel, showing that the model is less certain in the PMI predictions than the FHA predictions. The characteristics for the overlap and non-overlap loans are presented in Exhibit 32. The income, FICO and LTV are slightly higher for FHA, but most characteristics are remarkably similar. Traditionally the story has been that FHA loans are too high-risk to be insured by private mortgage insurance. The results from this section show that a sizable portion of the FHA loans (7 percent of the pooled and 9 percent of the FHA sample) carries the same risk as loans already getting PMI coverage.

FHA vs. All PMI

Many PMI-insured loans are not sold to the GSEs, but rather held by depository lenders or other investors. In this section, all the fixed-rate PMI loans are included that have loan amounts below the FHA loan limits. Exhibit 13 compares the 40,435 PMI loans with the 79,175 FHA loans from the matched Experian/HMDA data. While 60 percent of the PMI loans are sold to the GSEs, there are still 27 percent held by depository lenders and 13 percent purchased by other investors. Based on the characteristics of PMI loans in Exhibit 13, it appears that the differences between the GSE and depository loans are modest, but in most cases the GSE loans are lower risk. For example, 12 percent of the GSE loans with PMI have FICO scores below 620 compared to 14 percent of depository loans with PMI. Moreover, 12 percent of GSE loans with PMI have LTV above 96 percent compared to 22 percent of depository loans with PMI. As a third example, only 22 percent of GSE loans with PMI are in low-income tracts compared to 31 percent of depository loans with PMI. In general, depository PMI loans have lower income and FICO scores, but higher LTV, percent minority, and percent underserved. The investor PMI loans have nearly identical characteristics to the GSE PMI loans.

Estimates from the origination model between FHA and all PMI are shown in Exhibit 33. The logit model estimates the probability of a loan being insured by FHA relative to PMI for fixed rate, FHA eligible loans. For the most part, the coefficients are similar to the previous model estimated for PMI loans purchased by the GSEs: FICO is negative, LTV positive, and minority positive. In other words, FHA is less likely to insure loans with high FICO score but more likely to insure loans with high LTV or percent minority (either based on borrower race or neighborhood composition). The goodness-of-fit statistics show a slightly weaker fit (for example, pseudo R^2 is 0.57 vs. 0.61 within GSE and the combined K-S statistic is 0.79 vs. 0.83 within GSE). This may be due to the heterogeneity between the investor and depository loans or the greater similarity between depository and FHA loans.

The predicted distributions of FHA and all PMI loans are shown in Exhibit 34 with the log odds distributions in the lower panel. The overlap between distributions is 18 percent based on the non-parametric tolerance interval method, which is larger than the 10 percent overlap between FHA and the PMI loans purchased by the GSEs. The characteristics of the overlap loans are presented in Exhibit 35. In terms of the FHA distribution, 15 percent of FHA loans have characteristics very similar to conventional loans with PMI insurance. It appears that up to 15 percent of the FHA loans could qualify as conventional loans. Between FHA and PMI, FHA loans have lower income, FICO and loan amounts, but higher LTV and higher tract variables, percent minority, and percent underserved.

Exhibit 31

Distribution of Predicted Probability
of the GSE with PMI and FHA Loans
Origination Model Weighted

Distribution of Predicted Log Odds
of the GSE with PMI and FHA Loans
⁻MSA Origination Model Weighted
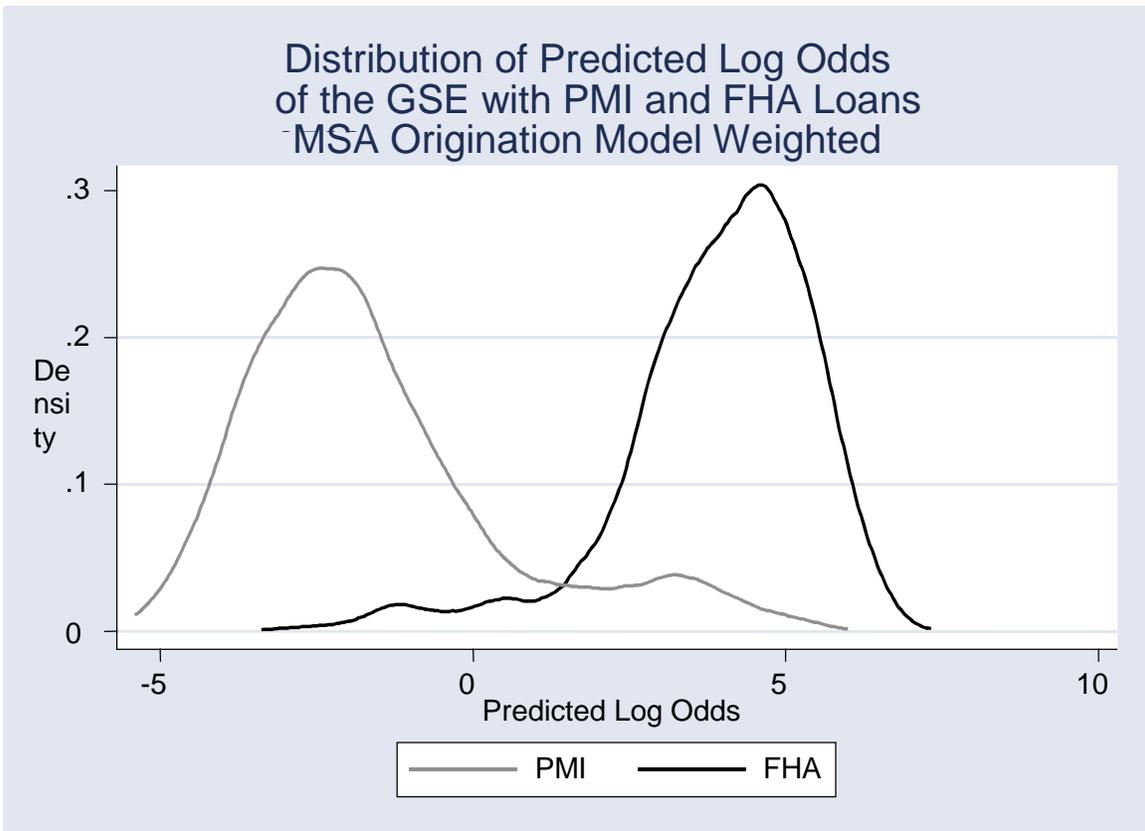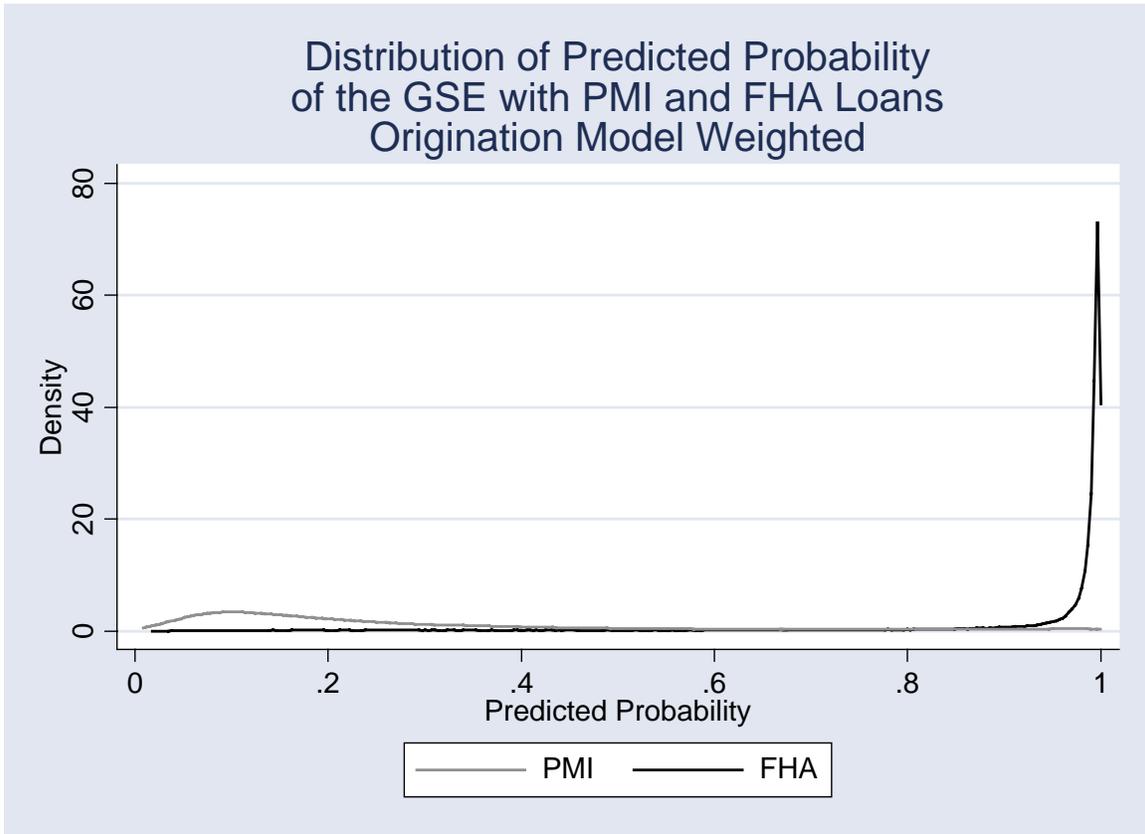
Exhibit 32: Characteristics of FHA and PMI Loans In and Out of Overlap by Non-Parametric Tolerance Interval Method

FHA=1, PMI=0
On Matched, Conforming, Fixed-Rate FHA-Eligible FHA or GSE+PMI Loans with LTV 80-100%, Weighted
(n=94,135; weighted sample=548,993)

	PMI (non-overlap)	PMI in Overlap	Combined Overlap	FHA in Overlap	FHA (non-overlap)
Share of Loans (10% in Tails)	15%	3%	10%	7%	64%
Borrower Characteristics					
Unweighted Number of Borrowers	13,938	2,924	9,119	6,195	61,805
Weighted Number of Borrowers	82,521	16,916	56,416	39,499	352,608
Average Annual Income	$57,153	$47,645	$51,323	$52,898	$48,307
Median Annual Income	$54,000	$45,000	$45,000	$45,000	$45,000
Average Annual Income (Trimmed Top 1%)	$56,196	$46,994	$48,132	$48,726	$46,869
% Estimated Income Information	2%	1%	1%	1%	1%
Average FICO	721	661	668	670	643
% With FICO <620	7%	30%	28%	28%	37%
% With FICO 620-680	16%	23%	21%	21%	25%
% With FICO =>680	77%	48%	50%	51%	37%
% Missing FICO Information	0%	0%	0%	0%	0%
% White	73%	61%	61%	60%	49%
% Black	5%	14%	14%	14%	21%
% Hispanic	12%	18%	19%	19%	24%
% Other	6%	5%	5%	5%	3%
% Missing Race Information	4%	2%	2%	2%	2%
% Female	26%	35%	37%	37%	32%
% Age 19-34	35%	37%	38%	38%	39%
% Age 35-49	50%	50%	49%	49%	49%
% Age 50-64	13%	11%	11%	11%	10%
% Age >65	2%	2%	2%	2%	2%
% Missing Age Information	0%	0%	0%	0%	0%
Loan Characteristics					
Average Loan Amount	$124,495	$111,948	$111,749	$111,664	$111,528
Average LTV %	92	94	95	95	99
% With LTV<=80	3%	3%	1%	1%	0%
% With LTV 80-90	32%	17%	14%	12%	0%
% With LTV 90-96	63%	25%	20%	18%	0%
% With LTV 96-98	2%	55%	64%	68%	12%
% With LTV>98	0%	0%	1%	1%	88%
% Missing LTV Information	0%	0%	0%	0%	0%
Average Ratio of Loan Amount to FHA Loan Limit	69%	61%	60%	60%	62%
% With LoantoFHA Ratio <=.5	18%	31%	33%	33%	28%
% With LoantoFHA Ratio of .5 - 1	82%	69%	67%	67%	72%
% With LoantoFHA Ratio of 1 - 1.2	0%	0%	0%	0%	0%
% With LoantoFHA Ratio of >1.2	0%	0%	0%	0%	0%

Exhibit 32 (cont.): Characteristics of FHA and PMI Loans In and Out of Overlap by Non-Parametric Tolerance Interval Method

FHA=1, PMI=0
On Matched, Conforming, Fixed-Rate FHA-Eligible FHA or GSE+PMI Loans with LTV 80-100%, Weighted
(n=94,135; weighted sample=548,993)

	PMI (non-overlap)	PMI in Overlap	Combined Overlap	FHA in Overlap	FHA (non-overlap)
Average PTI	0.20	0.21	0.21	0.21	0.21
% Originated in 1998	34%	33%	34%	34%	30%
% Originated in 1999	33%	33%	32%	31%	36%
% Originated in 2000	33%	34%	34%	35%	34%
Mortgaged Property Characteristics					
% New Construction	6%	10%	9%	9%	7%
% Unit Size 1	94%	93%	93%	93%	96%
% Unit Size 2	2%	2%	2%	1%	2%
% Unit Size 3	3%	3%	3%	3%	2%
% Unit Size 4	1%	2%	3%	3%	1%
Borrower Neighborhood Characteristics (1990 Census)					
% In Underserved Tracts	31%	39%	39%	39%	50%
% In High Cost MSA	10%	12%	12%	12%	10%
% In Medium Cost MSA	87%	85%	85%	85%	86%
% In Low Cost MSA	3%	3%	3%	3%	3%
% In Center City	30%	27%	28%	28%	31%
Average 5-yr House Price Appreciation, Lagged 1 year	114%	113%	111%	111%	111%
% In Area with Depreciation	11%	12%	14%	15%	12%
% In Area with Appreciation up to 20%	73%	73%	75%	76%	74%
% In Area with Appreciation over 20%	16%	15%	11%	9%	15%
% In Tracts with Income <90% of MSA Income	23%	27%	27%	27%	35%
% In Tracts with Income 90 - 120% of MSA Income	38%	36%	37%	38%	40%
% In Tracts with Income =>120% of MSA Income	39%	36%	36%	35%	25%
% In <10% Minority Tracts	44%	34%	30%	29%	31%
% In 10-30% Minority Tracts	33%	36%	38%	39%	32%
% In =>30% Minority Tracts	23%	31%	32%	32%	37%

Exhibit 33: Origination Model Results on FHA vs. PMI

FHA=1, PMI=0
On Matched, Conforming, Fixed-Rate FHA-Eligible FHA or PMI Loans, Weighted
(n=119,610; weighted sample=692,316)

Parameter	Label	Full Model with Std Errs from SAS		Full Model with Robust Std Errs from STATA					
		Estimate	Pr > ChiSq	Robust Coefficient	Standard Error	P>z	Mean	Odds Ratio	Marginal Change (-SD/2 to +SD/2)
Intercept	Intercept	7.0679	<.0001	7.0679	0.4010	0.000			
ex_fico	FICO Score	-0.00213	<.0001	-0.0021	0.0004	0.000	664.6334	0.997876	-0.0247
FICOs620	Marginal spline for FICO above 620	-0.00798	<.0001	-0.0080	0.0005	0.000	65.9295	0.992057	-0.0628
EX_LTV	Loan to Value Ratio	-0.0408	<.0001	-0.0408	0.0013	0.000	93.9578	0.959997	-0.0483
LTVs95	Marginal spline for LTV above 95%	1.3078	<.0001	1.3078	0.0150	0.000	2.4825	3.698006	0.3546
LTVs98	Marginal spline for LTV above 98%	1.9471	<.0001	1.9474	0.1160	0.000	0.6759	7.010756	0.1987
LTVs99	Marginal spline for LTV above 99%	-5.5626	<.0001	-5.5633	0.2140	0.000	0.2280	0.003836	-0.2820
rounded loantoFHA	Ratio of Borrower's Loan to FHA Loan Limit	-1.6226	<.0001	-1.6226	0.0686	0.000	0.6309	0.197388	-0.0402
PTI_calc	Payment to Income Ratio- Calculated	1.1848	<.0001	1.1848	0.3918	0.002	0.2071	3.269936	0.0149
PTIs20	Marginal spline for PTI above .20	4.0967	<.0001	4.0967	0.7803	0.000	0.0306	60.13895	0.0443
PTIs28	Marginal spline for PTI above .28	-3.1442	<.0001	-3.1442	0.8166	0.000	0.0070	0.043103	-0.0297
age_3549	Borrower Aged 35-49	-0.1803	<.0001	-0.1803	0.0255	0.000	0.4852	0.835057	-0.0114
age_5064	Borrower Aged 50-64	-0.2779	<.0001	-0.2779	0.0374	0.000	0.1096	0.757385	-0.0110
age_65	Borrower Aged 65 and up	-0.441	<.0001	-0.4410	0.0774	0.000	0.0217	0.643398	-0.0081
female	Female Borrower	0.0591	<.0001	0.0591	0.0262	0.024	0.3134	1.060933	0.0035
race_black	African American Borrower	0.5684	<.0001	0.5684	0.0437	0.000	0.1501	1.765353	0.0257
race_hisp	Hispanic Borrower	0.2528	<.0001	0.2528	0.0392	0.000	0.1726	1.287577	0.0121
race_other	Other Borrower (Native American, Asian, or Other)	-0.1379	<.0001	-0.1379	0.0590	0.019	0.0353	0.871212	-0.0032
new_constr	New Construction	0.2268	<.0001	0.2268	0.0416	0.000	0.0812	1.254553	0.0079
relinc_c90t	Census Tract family income relative to MSA median family income in 1990	0.1691	<.0001	0.1691	0.0455	0.000	1.0403	1.184232	0.0066
c90t_pminorty	Pct minority	0.00131	<.0001	0.0013	0.0007	0.059	24.5297	1.001316	0.0045
ccity_n	Center City tract indicator	-0.3282	<.0001	-0.3282	0.0290	0.000	0.3232	0.720183	-0.0195
high_cost	High Cost City based on FHA Loan Limits	0.0325	0.0724	0.0325	0.0540	0.547	0.0838	1.033063	0.0011
low_cost	Low Cost City based on FHA Loan Limits	0.0752	0.0249	0.0752	0.0627	0.230	0.0599	1.078135	0.0023
hpiL5	OFHEO House Price 5-yr change lagged 1 yr	-1.977	<.0001	-1.9770	0.2860	0.000	1.1626	0.138489	-0.0304
year98	Loan Sold in 1998	0.3822	<.0001	0.3822	0.0303	0.000	0.2499	1.46549	0.0210
year00	Loan Sold in 2000	-0.0357	0.002	-0.0357	0.0290	0.218	0.3936	0.964917	-0.0022
Baltimore	Property in Baltimore	0.575	<.0001	0.5750	0.0673	0.000	0.0472	1.777046	0.0155
Chicago	Property in Chicago	-0.2029	<.0001	-0.2029	0.0600	0.001	0.1317	0.816349	-0.0087
Cleveland	Property in Cleveland	-0.3056	<.0001	-0.3056	0.0911	0.001	0.0436	0.736652	-0.0079
Denver	Property in Denver	0.3687	<.0001	0.3687	0.0617	0.000	0.1530	1.445866	0.0168
LosAngeles	Property in Los Angeles	-0.3514	<.0001	-0.3514	0.0648	0.000	0.0684	0.703696	-0.0112
Oakland	Property in Oakland	-0.7221	<.0001	-0.7221	0.0891	0.000	0.0269	0.485738	-0.0148
Philadelphia	Property in Philadelphia	-0.2823	<.0001	-0.2823	0.0528	0.000	0.1009	0.754065	-0.0108
Portland	Property in Portland	-0.1643	0.0003	-0.1643	0.1298	0.206	0.0931	0.848465	-0.0061
StLouis	Property in St Louis	-0.4288	<.0001	-0.4289	0.0680	0.000	0.0981	0.651257	-0.0162
Tampa	Property in Tampa	-0.0733	0.0056	-0.0733	0.0631	0.245	0.1298	0.929343	-0.0031

Percent Concordant	95	
Percent Discordant	4.9	
Somers' D	0.900	
Pseudo R2		0.5674
Log Likelihood- Intercept only	876301.61	
Log Likelihood- Intercept and Covariates	379107.55	
Combined K-S Statistic		0.7850 (P-Value = 0.000)

Exhibit 34

Distribution of Predicted Log Odds
of all PMI and FHA Loans
Origination Model Weighted

Distribution of Predicted Probability
of all PMI and FHA Loans
Origination Model Weighted
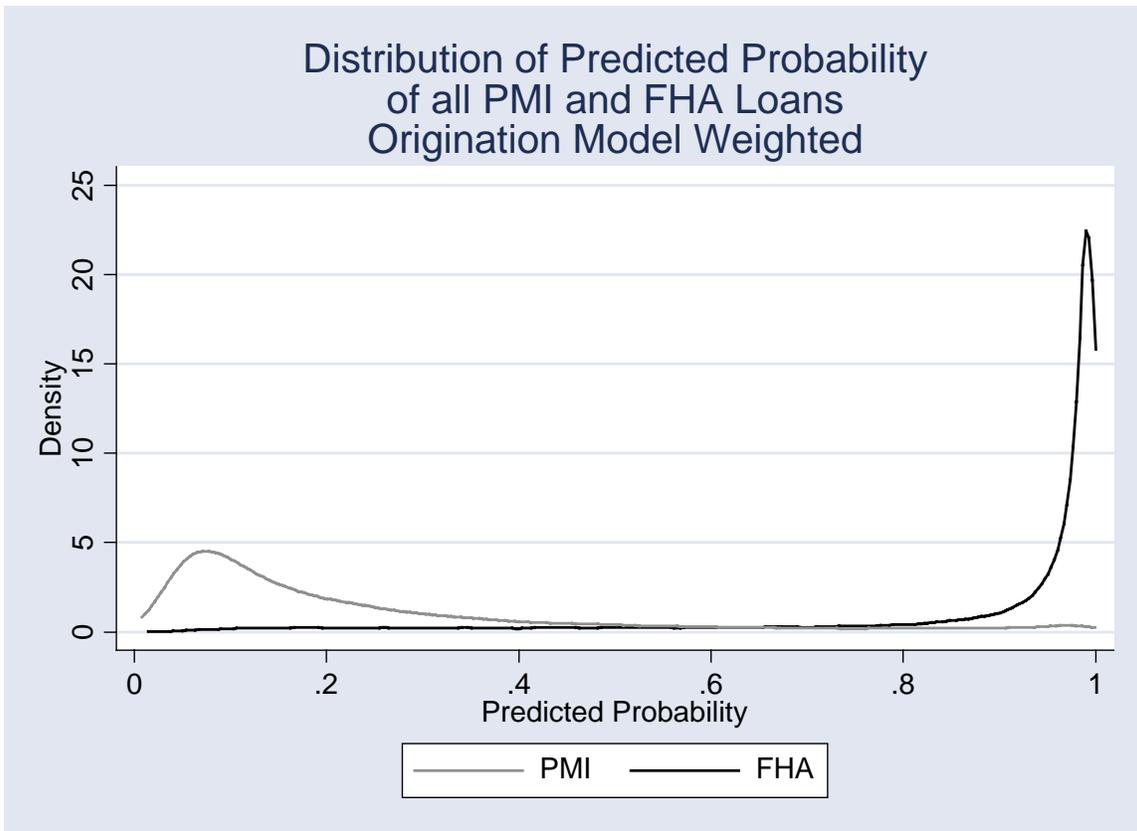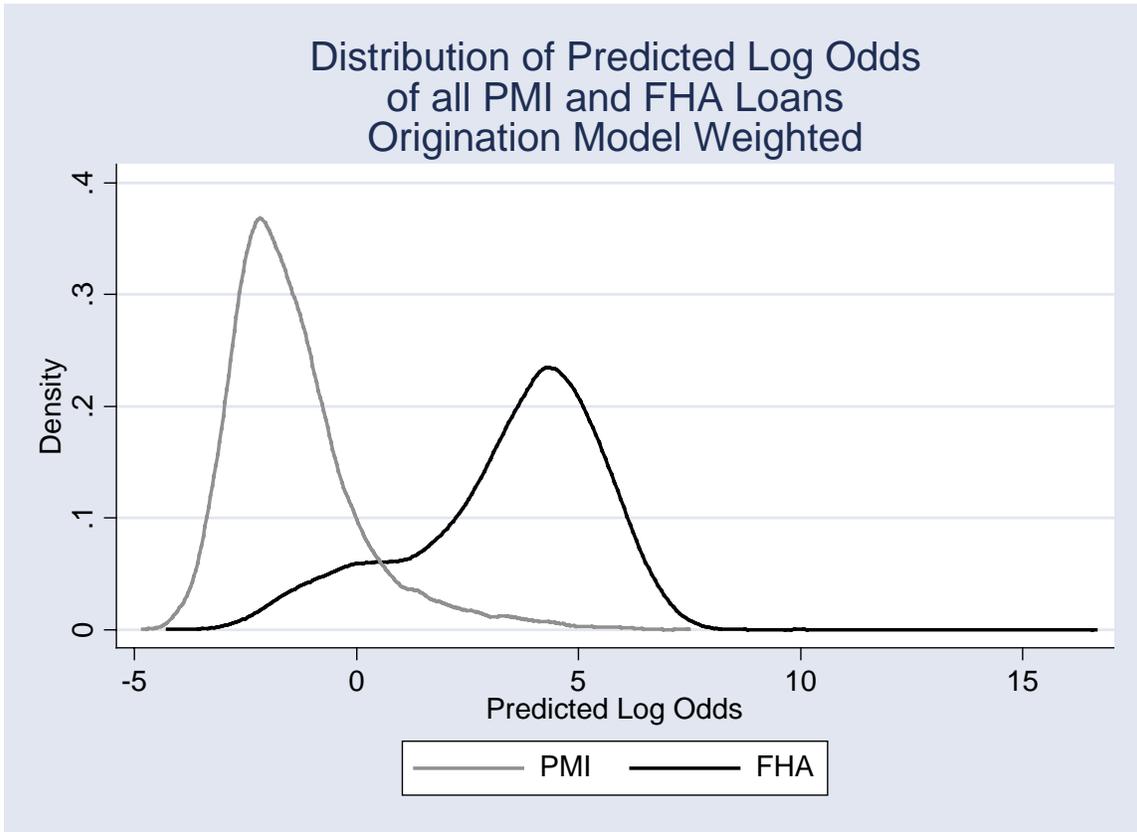
Exhibit 35: Characteristics of FHA and PMI Loans In and Out of Overlap by Non-Parametric Tolerance Interval Method
FHA=1, PMI=0
On Matched, Conforming, Fixed-Rate FHA-Eligible FHA or PMI Loans, Weighted
(n=119,610; weighted sample=692,316)

	PMI (non-overlap)	PMI in Overlap	Combined Overlap	FHA in Overlap	FHA (non-overlap)
Share of Loans (10% in Tails)	22%	8%	18%	10%	50%
Borrower Characteristics					
Unweighted Number of Borrowers	26,976	9,499	19,967	10,468	60,881
Weighted Number of Borrowers	150,975	55,509	122,802	67,293	347,498
Average Annual Income	$57,398	$47,581	$49,037	$50,238	$48,204
Median Annual Income	$54,000	$44,000	$44,000	$44,000	$44,000
Average Annual Income (Trimmed Top 1%)	$56,350	$46,677	$46,904	$47,199	$46,797
% Estimated Income Information	3%	3%	2%	1%	1%
Average FICO	733	657	657	657	641
% With FICO <620	4%	31%	32%	33%	38%
% With FICO 620-680	12%	26%	24%	23%	25%
% With FICO =>680	84%	43%	44%	44%	37%
% Missing FICO Information	0%	0%	0%	0%	0%
% White	77%	58%	56%	55%	49%
% Black	4%	15%	17%	18%	21%
% Hispanic	10%	19%	20%	20%	25%
% Other	6%	4%	4%	4%	3%
% Missing Race Information	3%	3%	3%	2%	2%
% Female	25%	35%	36%	37%	32%
% Age 19-34	34%	36%	37%	37%	39%
% Age 35-49	49%	50%	50%	49%	49%
% Age 50-64	14%	12%	12%	11%	10%
% Age >65	3%	2%	2%	2%	2%
% Missing Age Information	0%	0%	0%	0%	0%
Loan Characteristics					
Average Loan Amount	$123,401	$107,947	$107,806	$107,690	$111,662
Average LTV %	88	87	88	88	99
% With LTV<=80	24%	28%	28%	29%	0%
% With LTV 80-90	25%	13%	11%	9%	0%
% With LTV 90-96	50%	26%	20%	14%	0%
% With LTV 96-98	1%	32%	40%	46%	12%
% With LTV>98	0%	0%	2%	3%	88%
% Missing LTV Information	0%	0%	0%	0%	0%
Average Ratio of Loan Amount to FHA Loan Limit	69%	59%	59%	58%	62%
% With LoantoFHA Ratio <=.5	18%	35%	36%	37%	28%
% With LoantoFHA Ratio of .5 - 1	82%	65%	64%	63%	72%
% With LoantoFHA Ratio of 1 - 1.2	0%	0%	0%	0%	0%
% With LoantoFHA Ratio of >1.2	0%	0%	0%	0%	0%

**Exhibit 35 (cont.): Characteristics of FHA and PMI Loans In and Out of Overlap by Non-Parametric Tolerance Interval Method**

FHA=1, PMI=0
On Matched, Conforming, Fixed-Rate FHA-Eligible FHA or PMI Loans, Weighted
(n=119,610; weighted sample=692,316)

	PMI (non-overlap)	PMI in Overlap	Combined Overlap	FHA in Overlap	FHA (non-overlap)
Average PTI	0.19	0.21	0.21	0.21	0.21
% Originated in 1998	33%	34%	34%	34%	29%
% Originated in 1999	35%	32%	31%	31%	37%
% Originated in 2000	32%	34%	35%	35%	34%
Mortgaged Property Characteristics					
% New Construction	7%	9%	9%	8%	7%
% Unit Size 1	94%	93%	93%	93%	96%
% Unit Size 2	2%	2%	2%	2%	2%
% Unit Size 3	3%	3%	3%	3%	2%
% Unit Size 4	1%	2%	2%	2%	1%
Borrower Neighborhood Characteristics (1990 Census)					
% In Underserved Tracts	29%	42%	43%	44%	50%
% In High Cost MSA	10%	11%	10%	10%	11%
% In Medium Cost MSA	88%	87%	87%	87%	86%
% In Low Cost MSA	3%	3%	3%	2%	3%
% In Center City	30%	30%	30%	30%	31%
Average 5-yr House Price Appreciation, Lagged 1 year	115%	113%	112%	111%	111%
% In Area with Depreciation	9%	11%	11%	12%	12%
% In Area with Appreciation up to 20%	71%	76%	78%	79%	74%
% In Area with Appreciation over 20%	20%	13%	11%	9%	14%
% In Tracts with Income <90% of MSA Income	21%	30%	31%	31%	35%
% In Tracts with Income 90 - 120% of MSA Income	36%	35%	36%	37%	40%
% In Tracts with Income =>120% of MSA Income	42%	34%	33%	32%	25%
% In <10% Minority Tracts	48%	35%	32%	29%	31%
% In 10-30% Minority Tracts	31%	32%	34%	35%	32%
% In =>30% Minority Tracts	20%	32%	35%	36%	37%

When comparing these results to the PMI loans purchased by the GSEs in Exhibit 32, it is worth remembering that the "all PMI" sample includes low LTV loans (below 80 percent), which lower LTV and perhaps FICO scores too. It appears that the depositories are more willing to hold the higher risk loans (PMI coverage is less than 100 percent) relative to the PMI loans sold to the GSEs as long as the low FICO scores are balanced by low LTVs. For example, comparing the PMI overlap loans between Exhibits 32 and 35, there are 26 percent more loans in the "all PMI" sample than the "PMI within GSE" sample, but over three times as many loans in the overlap with FHA. The average FICO score for the overlap PMIs is lower for the "all PMI" sample (657 vs. 661), but so is the average LTV (87 vs. 94) compared to overlap PMI loans purchased by the GSEs.

FHA vs. Subprime

Replacing PMI with subprime loans, a data set is created with 80 percent FHA and 20 percent subprime. The subprime designation is based on the name of the lender, rather than the specific characteristics of either the borrower or the loan. Also note, the loans are home purchase mortgages only, so there are no refinances where subprimes are more common. Despite that limitation, the fit of the model is only a little worse than for the GSE or PMI models ($R^2 = 0.57$ and concordant is 95.5 percent), see Exhibit 36.

The coefficient on FICO is now positive and significant – higher FICO scores increase the probability of a loan being FHA. The impact of LTV is captured by a set of marginal splines, which have a cumulative effect relative to the continuous base variable. The positive coefficient for the base variable and first marginal spline are partially offset in the second marginal spline and turns negative for the high LTV. Frankly, the negative coefficient for high LTV loans seems unusual because so few subprime loans have high LTV and so many FHA loans have an LTV ratio above 98 percent. The large difference in LTV between FHA and subprime loans enables the model to distinguish the two types of loans and keep the overlap modest, even though in other respects FHA and subprime loans are relatively similar. The PTI coefficient has become extremely positive, though still offset and matched by the marginal spline above 0.28. The black borrower indicator is negative and significant showing that African Americans are less likely to have an FHA loan than a subprime loan. The tract percent minority is also negative and significant. Hispanic borrowers, however, are more likely to get an FHA loan, whereas borrowers of other races are more likely to get a subprime loan. The negative center city coefficient shows that FHA loans are less likely than subprime loans to finance properties in the center city.

The overlap graphs in Exhibit 37 look quite similar to the FHA/PMI graph with a bimodal shape to the FHA distribution. Exhibit 38 shows the overlap based on the non-parametric tolerance interval method is 13 percent with more FHA loans (9 percent) than subprime (4 percent). In terms of just the FHA distribution, 13 percent of FHA loans overlap with the subprime market. The most obvious difference in loan characteristics is that FHA loans have much higher LTVs. The model can readily distinguish subprime from FHA loans in most cases based on the difference in LTV. Subprime loans with lower LTV may qualify for FHA to a much higher degree than implied by the 13 percent overlap. Overlap FHA borrowers have higher mean income, but nearly equal median incomes. A high proportion of the subprime incomes are imputed, which suggests the overlap results may be sensitive to the assumptions used to impute for missing data. A much higher share of subprime loans are missing race information as well. The non-overlap FHA incomes are definitely lower than subprime incomes.

Exhibit 36: Origination Model Results on FHA vs. Subprime

FHA=1, Subprime=0
On All FHA and Subprime Matched, Conforming Loans, Weighted
(n=111,942; weighted sample= 637,982)

Parameter	Label	Full Model with Std Errs from SAS		Full Model with Robust Std Errs from STATA					
		Estimate	Pr > ChiSq	Robust Coefficient	Standard Error	P>z	Mean	Odds Ratio	Marginal Change (-SD/2 to +SD/2)
Intercept	Intercept	-5.2327	<.0001	-5.2327	0.4293	0.000		0.0053	
ex_fico	FICO Score	0.0055	<.0001	0.0055	0.0004	0.000	642.2536	1.0056	0.0121
FICOs620	Marginal spline for FICO above 620	-0.00798	<.0001	-0.0080	0.0006	0.000	50.8206	0.9921	-0.0107
EX_LTV	Loan to Value Ratio	0.0169	<.0001	0.0169	0.0016	0.000	94.3422	1.0171	0.004
LTVs95	Marginal spline for LTV above 95%	1.7315	<.0001	1.7315	0.0342	0.000	2.9573	5.6491	0.1112
LTVs98	Marginal spline for LTV above 98%	-0.8583	<.0001	-0.8583	0.1851	0.000	0.8147	0.4239	-0.0164
LTVs99	Marginal spline for LTV above 99%	-2.8368	<.0001	-2.8368	0.2980	0.000	0.2729	0.0586	-0.0285
rounded loantoFHA	Ratio of Borrower's Loan to FHA Loan Limit	-1.2698	<.0001	-1.2698	0.0741	0.000	0.6317	0.2809	-0.0066
PTI_calc	Payment to Income Ratio- Calculated	11.0672	<.0001	11.0672	0.4376	0.000	0.2095	64036.5953	0.032
PTIs20	Marginal spline for PTI above .20	-0.5922	0.071	-0.5922	0.8994	0.510	0.0331	0.5531	-0.0014
PTIs28	Marginal spline for PTI above .28	-10.8532	<.0001	-10.8532	0.6304	0.000	0.0086	0.0000	-0.0245
age_3549	Borrower Aged 35-49	-0.4015	<.0001	-0.4015	0.0322	0.000	0.4989	0.6693	-0.0048
age_5064	Borrower Aged 50-64	-0.5049	<.0001	-0.5049	0.0469	0.000	0.1040	0.6036	-0.0037
age_65	Borrower Aged 65 and up	-0.7557	<.0001	-0.7557	0.0924	0.000	0.0179	0.4697	-0.0025
female	Female Borrower	-0.0949	<.0001	-0.0949	0.0316	0.003	0.3253	0.9094	-0.0011
race_black	African American Borrower	-0.1157	<.0001	-0.1157	0.0449	0.010	0.2135	0.8907	-0.0011
race_hisp	Hispanic Borrower	0.5393	<.0001	0.5393	0.0491	0.000	0.2116	1.7148	0.005
race_other	Other Borrower (Native American, Asian, or Other)	-0.4726	<.0001	-0.4726	0.0722	0.000	0.0372	0.6234	-0.002
new_constr	New Construction	0.2135	<.0001	0.2135	0.0513	0.000	0.0820	1.2380	0.0014
relinc_c90t	Census Tract family income relative to MSA median family income in 1990	0.0266	0.2092	0.0266	0.0559	0.634	1.0437	1.0270	0.0002
c90t_pminorty	Pct minority	-0.00272	<.0001	-0.0027	0.0007	0.000	31.4086	0.9973	-0.0018
ccity_n	Center City tract indicator	-0.3130	<.0001	-0.3130	0.0346	0.000	0.3007	0.7312	-0.0035
high_cost	High Cost City based on FHA Loan Limits	-0.1242	<.0001	-0.1242	0.0619	0.045	0.1172	0.8832	-0.0009
low_cost	Low Cost City based on FHA Loan Limits	0.2045	<.0001	0.2045	0.0742	0.006	0.0281	1.2269	0.0011
hpiL5	OFHEO House Price 5-yr change lagged 1 yr	0.3463	0.0015	0.3463	0.3237	0.285	1.1166	1.4138	0.001
year98	Loan Originated in 1998	0.5930	<.0001	0.5930	0.0386	0.000	0.2922	1.8094	0.0059
year00	Loan Originated in 2000	-0.2937	<.0001	-0.2937	0.0355	0.000	0.3484	0.7455	-0.0034
Baltimore	Property in Baltimore	0.0366	0.2074	0.0366	0.0899	0.684	0.0932	1.0373	0.0002
Chicago	Property in Chicago	-0.2671	<.0001	-0.2671	0.0762	0.000	0.1537	0.7656	-0.002
Cleveland	Property in Cleveland	-1.1141	<.0001	-1.1141	0.1050	0.000	0.0370	0.3282	-0.0061
Denver	Property in Denver	-1.0339	<.0001	-1.0339	0.0783	0.000	0.1040	0.3556	-0.0093
LosAngeles	Property in Los Angeles	-1.2943	<.0001	-1.2943	0.0772	0.000	0.1576	0.2741	-0.008
Oakland	Property in Oakland	-2.0659	<.0001	-2.0659	0.0968	0.000	0.0346	0.1267	-0.0084
Philadelphia	Property in Philadelphia	-0.7039	<.0001	-0.7039	0.0690	0.000	0.0968	0.4947	-0.005
Portland	Property in Portland	-1.8963	<.0001	-1.8963	0.1453	0.000	0.0423	0.1501	-0.013
StLouis	Property in St Louis	-1.2702	<.0001	-1.2702	0.0827	0.000	0.0499	0.2808	-0.0085
Tampa	Property in Tampa	-1.2636	<.0001	-1.2636	0.0799	0.000	0.0694	0.2826	-0.01

Percent Concordant	95.5
Percent Discordant	4.3
Somers' D	0.912
Pseudo R2	0.5677
Log Likelihood- Intercept only	539811.99
Log Likelihood- Intercept and Covariates	233335.14
Combined K-S Statistic	0.8192 (P-Value = 0.000)

Exhibit 37

Distribution of Predicted Probability
of the Subprime and FHA Loans
Origination Model Weighted

Distribution of Predicted Log Odds
of the Subprime and FHA Loans
Origination Model Weighted
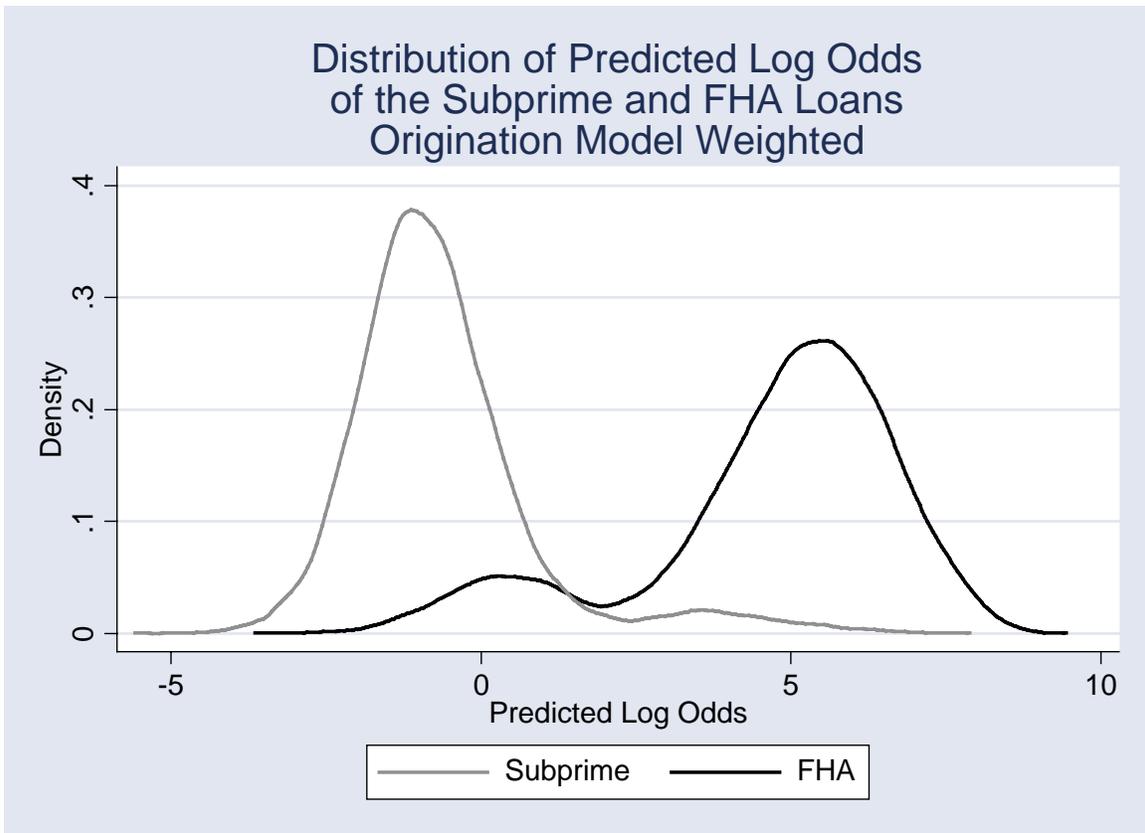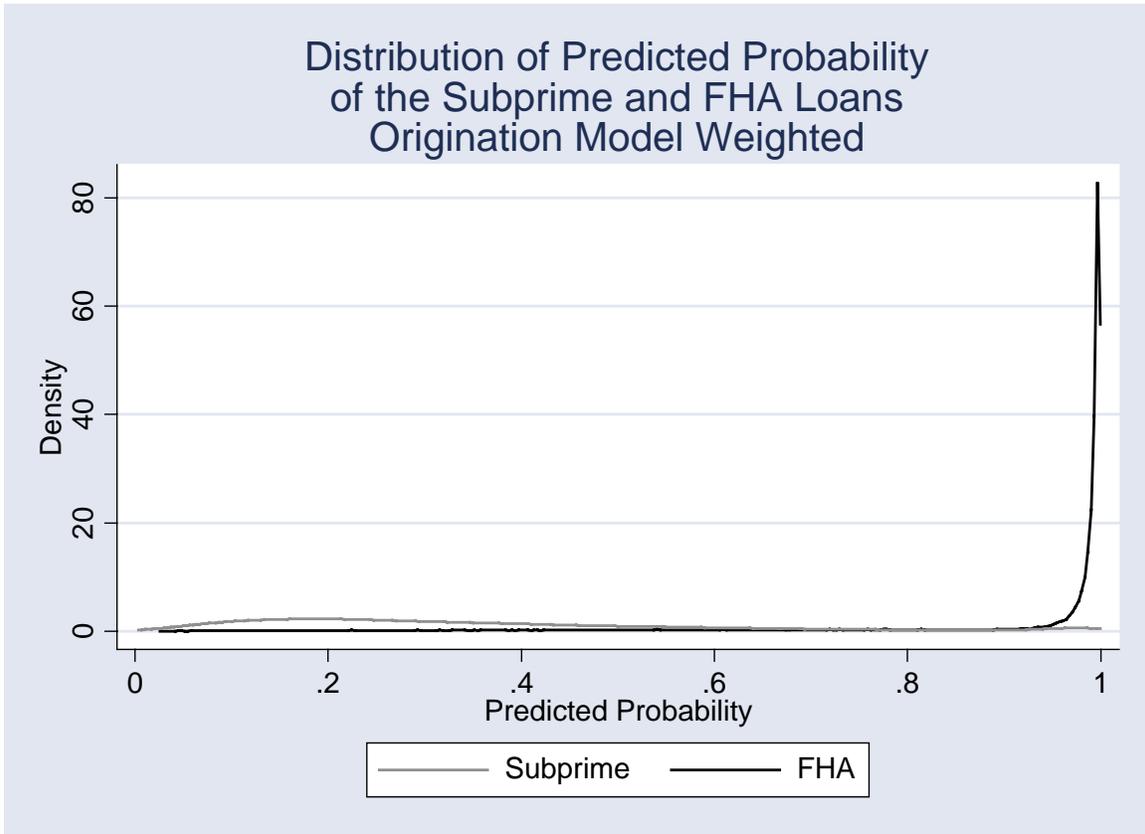
Exhibit 38: Characteristics of FHA and Subprime Loans In and Out of Overlap by Non-Parametric Tolerance Interval Method

FHA=1, Subprime=0
On All FHA and Subprime Matched, Conforming Loans, Weighted
(n=111,942; weighted sample= 637,982)

	Subprime (non-overlap)	Subprime in Overlap	Combined Overlap	FHA in Overlap	FHA (non-overlap)
Share of Loans (10% in Tails)	10%	4%	13%	9%	66%
Borrower Characteristics					
Unweighted Number of Borrowers	12,944	3,613	12,641	9,028	75,322
Weighted Number of Borrowers	64,129	23,050	80,717	57,668	423,111
Average Annual Income	$64,891	$46,431	$49,639	$50,921	$49,309
Median Annual Income	$59,000	$42,500	$43,000	$43,000	$45,000
Average Annual Income (Trimmed Top 1%)	$62,326	$45,505	$46,623	$47,150	$47,962
% Estimated Income Information	10%	15%	5%	1%	1%
Average FICO	634	648	647	647	642
% With FICO <620	45%	37%	36%	36%	38%
% With FICO 620-680	19%	26%	26%	26%	24%
% With FICO =>680	36%	38%	38%	39%	38%
% Missing FICO Information	0%	0%	0%	0%	0%
% White	54%	47%	51%	53%	51%
% Black	22%	21%	20%	20%	22%
% Hispanic	8%	21%	21%	21%	22%
% Other	7%	4%	4%	4%	3%
% Missing Race Information	9%	7%	4%	2%	2%
% Female	33%	36%	36%	36%	32%
% Age 19-34	23%	38%	38%	38%	39%
% Age 35-49	58%	49%	49%	49%	50%
% Age 50-64	16%	11%	11%	11%	10%
% Age >65	3%	2%	2%	2%	2%
% Missing Age Information	0%	0%	0%	0%	0%
Loan Characteristics					
Average Loan Amount	$114,475	$111,049	$111,487	$111,663	$114,598
Average LTV %	79	84	86	86	99
% With LTV<=80	64%	42%	35%	32%	0%
% With LTV 80-90	25%	28%	18%	14%	0%
% With LTV 90-96	12%	27%	32%	35%	0%
% With LTV 96-98	0%	3%	12%	16%	18%
% With LTV>98	0%	1%	3%	4%	82%
% Missing LTV Information	0%	0%	0%	0%	0%
Average Ratio of Loan Amount to FHA Loan Limit	64%	61%	61%	61%	64%
% With LoantoFHA Ratio <=.5	37%	38%	35%	33%	27%
% With LoantoFHA Ratio of .5 - 1	50%	54%	60%	62%	71%
% With LoantoFHA Ratio of 1 - 1.2	10%	6%	5%	4%	2%
% With LoantoFHA Ratio of >1.2	3%	2%	1%	0%	0%

Exhibit 38 (cont.): Characteristics of FHA and Subprime Loans In and Out of Overlap by Non-Parametric Tolerance Interval Method

FHA=1, Subprime=0
On All FHA and Subprime Matched, Conforming Loans, Weighted
(n=111,942; weighted sample= 637,982)

	Subprime (non-overlap)	Subprime in Overlap	Combined Overlap	FHA in Overlap	FHA (non-overlap)
Average PTI	0.17	0.22	0.22	0.22	0.21
% Originated in 1998	25%	34%	34%	34%	28%
% Originated in 1999	36%	31%	31%	31%	37%
% Originated in 2000	40%	35%	35%	35%	35%
Mortgaged Property Characteristics					
% New Construction	7%	9%	9%	9%	8%
% Unit Size 1	96%	93%	94%	94%	96%
% Unit Size 2	1%	2%	2%	2%	2%
% Unit Size 3	2%	3%	3%	2%	2%
% Unit Size 4	1%	2%	2%	2%	1%
Borrower Neighborhood Characteristics (1990 Census)					
% In Underserved Tracts	48%	43%	44%	44%	49%
% In High Cost MSA	20%	8%	9%	9%	11%
% In Medium Cost MSA	76%	90%	89%	89%	86%
% In Low Cost MSA	4%	2%	2%	2%	3%
% In Center City	39%	27%	29%	29%	30%
Average 5-yr House Price Appreciation, Lagged 1 year	114%	112%	112%	112%	111%
% In Area with Depreciation	17%	8%	9%	10%	12%
% In Area with Appreciation up to 20%	56%	83%	82%	81%	73%
% In Area with Appreciation over 20%	27%	8%	9%	9%	15%
% In Tracts with Income <90% of MSA Income	35%	30%	31%	31%	33%
% In Tracts with Income 90 - 120% of MSA Income	31%	36%	36%	35%	40%
% In Tracts with Income =>120% of MSA Income	34%	34%	33%	33%	27%
% In <10% Minority Tracts	30%	34%	32%	32%	31%
% In 10-30% Minority Tracts	30%	33%	32%	32%	32%
% In =>30% Minority Tracts	40%	33%	35%	36%	37%

The FICO distributions in the overlap region are very similar between FHA and subprime. In the non-overlap regions, mean FICO scores are higher for FHA (642 vs. 634) with more subprime loans having FICO scores below 620.

Within the overlap region, the distribution by race is quite similar, though subprime has fewer whites and more missing. Between the non-overlap regions, FHA has a much higher share of Hispanic borrowers (22 vs. 8 percent). The age distribution within the overlap is very similar, but for non-overlap loans FHA serves more young borrowers.

Compensating for the lower FICO scores of subprime loans is the lower LTV distribution. Within the overlap, the average LTV for subprime is 84 percent, modestly lower than FHA's 86 percent. Outside the overlap the subprime average LTV is 79 percent, much lower than the FHA average of 99 percent. Another compensating factor for the non-overlapping regions is PTI, which tends not to vary much, but subprime average PTI is 0.17 compared to FHA average PTI of 0.21.

Considering the neighborhood characteristics, more non-overlapping subprime loans are in high cost MSAs than overlap or FHA loans. Also, 39 percent of non-overlapping subprime loans are in center cities compared to 30 percent for FHA. In terms of house price appreciation, the overlap loans are quite similar, but the non-overlap subprime loans (27 percent) are more likely to be in high appreciation MSAs than the non-overlap FHA loans (15 percent).

Given the problems with identifying subprime loans, any conclusions are tenuous. Nevertheless, this analysis shows that 23,050 subprime loans out of 101,464 total subprime loans, or 23 percent of subprime loans have very similar risk profiles to FHA loans.

GSE vs. Depository Lenders

The last mortgage market interface is between GSE-purchased loans and conforming loans held by depository lenders. The GSEs often voice concern over adverse selection by which depository lenders sell the high-risk loans to the GSEs and keep the low risk loans for their own portfolio. However, that selection is done on non-observable characteristics, at least on characteristics not available to this analysis. So, in terms of observable characteristics, the GSE and depository loans are expected to be quite similar. Owing to large sample sizes, most explanatory variables are significant, but the model goodness-of-fit is paltry (pseudo-R^2 is 0.09, concordant is 68.3 percent, KS statistic is 0.268). The same specification is used as with the other origination models. Some customization for conventional loans might have improved the fit. The sample is evenly divided between GSE and depository loans.

The origination model estimates the probability of a loan being GSE, as shown in Exhibit 39. Loans with higher FICO scores, higher LTV, higher loan amount, older borrowers and higher neighborhood income are more likely to be GSE loans. On the other hand, loans with higher payment burdens (PTI), female and minority borrowers, new construction, minority and center city neighborhoods, high cost and high house price appreciation are less likely to be purchased by the GSEs.

Exhibit 39: Origination Model Results on GSE vs. Depository Lender

GSE=1, Depository=0
On All Matched, Conforming GSE or Depository Loans
(n=229,808, weighted sample=1,309,153)

Parameter	Label	Full Model with Std Errs from SAS		Full Model with Robust Std Errs from STATA					
		Estimate	Pr > ChiSq	Robust Coefficient	Standard Error	P>z	Mean	Odds Ratio	Marginal Change (-SD/2 to +SD/2)
Intercept	Intercept	-3.6070	<.0001	-3.6070	0.1908	0.000		0.0271	
ex_fico	FICO Score	0.0038	<.0001	0.0038	0.0002	0.000	712.5595	1.0038	0.0802
FICOs620	Marginal spline for FICO above 620	-0.00175	<.0001	-0.0018	0.0003	0.000	102.0566	0.9982	-0.0283
EX_LTV	Loan to Value Ratio	0.0056	<.0001	0.0056	0.0004	0.000	80.6390	1.0056	0.0215
LTVs95	Marginal spline for LTV above 95%	-0.3698	<.0001	-0.3697	0.0111	0.000	0.4866	0.6909	-0.116
LTVs98	Marginal spline for LTV above 98%	-1.5596	<.0001	-1.5603	0.0874	0.000	0.1152	0.2101	-0.1631
LTVs99	Marginal spline for LTV above 99%	3.0889	<.0001	3.0903	0.1597	0.000	0.0435	21.9827	0.1483
rounded loantoFHA	Ratio of Borrower's Loan to FHA Loan Limit	0.4290	<.0001	0.4290	0.0198	0.000	0.7528	1.5357	0.0349
PTI_calc	Payment to Income Ratio- Calculated	2.6145	<.0001	2.6145	0.1486	0.000	0.1920	13.6605	0.0746
PTIs20	Marginal spline for PTI above .20	-4.7816	<.0001	-4.7816	0.3198	0.000	0.0260	0.0084	-0.1152
PTIs28	Marginal spline for PTI above .28	2.3066	<.0001	2.3066	0.2466	0.000	0.0073	10.0406	0.0505
age_3549	Borrower Aged 35-49	0.0312	<.0001	0.0312	0.0120	0.009	0.5035	1.0317	0.0039
age_5064	Borrower Aged 50-64	0.0567	<.0001	0.0567	0.0161	0.000	0.1579	1.0583	0.0052
age_65	Borrower Aged 65 and up	-0.0177	0.0782	-0.0177	0.0270	0.511	0.0394	0.9824	-0.0009
female	Female Borrower	-0.0724	<.0001	-0.0724	0.0123	0.000	0.2666	0.9301	-0.008
race_black	African American Borrower	-0.3075	<.0001	-0.3075	0.0239	0.000	0.0742	0.7353	-0.0196
race_hisp	Hispanic Borrower	-0.1714	<.0001	-0.1714	0.0215	0.000	0.0885	0.8425	-0.0111
race_other	Other Borrower (Native American, Asian, or Other)	-0.1668	<.0001	-0.1668	0.0210	0.000	0.0813	0.8464	-0.0103
new_constr	New Construction	-0.0449	<.0001	-0.0449	0.0163	0.006	0.1109	0.9561	-0.0036
relinc_c90t	Census Tract family income relative to MSA median family income in 1990	0.1744	<.0001	0.1744	0.0178	0.000	1.2021	1.1905	0.0161
c90t_pminorty	Pct minority	-0.00296	<.0001	-0.0030	0.0003	0.000	20.4354	0.9970	-0.016
ccity_n	Center City tract indicator	-0.0798	<.0001	-0.0798	0.0131	0.000	0.2618	0.9233	-0.0089
high_cost	High Cost City based on FHA Loan Limits	-0.2356	<.0001	-0.2356	0.0239	0.000	0.1172	0.7901	-0.018
low_cost	Low Cost City based on FHA Loan Limits	-0.0141	0.3169	-0.0141	0.0262	0.592	0.0301	0.9860	-0.0008
hpiL5	OFHEO House Price 5-yr change lagged 1 yr	-0.3209	<.0001	-0.3209	0.1240	0.010	1.1347	0.7255	-0.0098
year98	Loan Originated in 1998	0.3073	<.0001	0.3073	0.0132	0.000	0.3222	1.3597	0.0337
year00	Loan Originated in 2000	0.2175	<.0001	0.2175	0.0128	0.000	0.3399	1.2429	0.0265
Baltimore	Property in Baltimore	-0.0608	<.0001	-0.0608	0.0323	0.060	0.0541	0.9410	-0.0025
Chicago	Property in Chicago	0.2881	<.0001	0.2881	0.0263	0.000	0.2159	1.3339	0.0256
Cleveland	Property in Cleveland	-0.4069	<.0001	-0.4069	0.0366	0.000	0.0545	0.6657	-0.0273
Denver	Property in Denver	0.8019	<.0001	0.8019	0.0286	0.000	0.0694	2.2298	0.0609
LosAngeles	Property in Los Angeles	0.2366	<.0001	0.2366	0.0286	0.000	0.1242	1.2669	0.0134
Oakland	Property in Oakland	0.3603	<.0001	0.3603	0.0351	0.000	0.0532	1.4338	0.0182
Philadelphia	Property in Philadelphia	0.1741	<.0001	0.1741	0.0230	0.000	0.1040	1.1902	0.0123
Portland	Property in Portland	0.5774	<.0001	0.5774	0.0548	0.000	0.0547	1.7815	0.0453
StLouis	Property in St Louis	0.1285	<.0001	0.1285	0.0293	0.000	0.0694	1.1371	0.0105
Tampa	Property in Tampa	0.0739	<.0001	0.0739	0.0272	0.007	0.0774	1.0767	0.0064

Percent Concordant	68.3		
Percent Discordant	31.3		
Somers' D	0.370		
Pseudo R2		0.0912	
Log Likelihood- Intercept only	1814694.9		
Log Likelihood- Intercept and Covariates	1649140.0		
Combined K-S Statistic		0.2682 (P-Value = 0.000)	

Using this origination model to predict probability distributions (see Exhibit 40) leads to nearly coincident distributions. The major deviation is that the depository distribution has a bulge in its left tail, presumably representing loans that clearly do not meet GSE underwriting standards. The overlap of these distributions is 78 percent split between 42 percent GSE loans and 37 percent depository loans. This large overlap leaves only 4 percent for non-overlapping GSE and 9 percent for the non-overlapping depository loans, which are distinctive along a number of dimensions. As shown in Exhibit 41, borrower income in the overlap region is slightly higher for GSE loans ($61,000 vs. $59,000 medians), but 65 percent higher for non-overlapping regions ($71,000 for GSE vs. $43,000 for depository). In the overlap, GSE loans have higher FICO (728 vs. 718), but the gap is much larger in the non-overlap (761 vs. 638).

Considering demographics, in the overlap the minority share of borrowers is fairly even, but in the non-overlap the GSEs have only 6 percent minority compared to 49 percent by the depositories. The same pattern holds for percent female with about 26 percent in the overlap, but only 14 percent for GSEs compared to 38 percent female for depositories in the non-overlap. Depositories have a higher share of young borrowers. With such a large overlap region and tails excluded, the remaining non-overlap regions are relatively small (only 4 percent for GSE and 9 percent for depositories). Given small portions at opposite ends of the distribution, it is not surprising to have relatively large differences in characteristics.

In terms of loan characteristics, average loan amounts for the GSEs are larger in both the overlap and especially in the non-overlap ($172,378 vs. $97,752). The LTVs are similar in the overlap, but the depositories have much higher LTVs in the non-overlap (81 percent for GSEs vs. 91 percent for depositories). Although the ratio of loan amount to FHA loan limit may not seem like a useful benchmark for conventional loans, most of the overlap loans are below the FHA loan limit (about 76 percent). The non-overlap GSE loans are much larger (105 percent) compared to the non-overlap depositories (54 percent).

The coefficient on new construction is negative in the origination model, meaning new houses are less likely to have GSE loans. Yet, when the non-overlap regions are compared, the GSEs have a higher share of new construction loans, 16 percent, than the depositories, 8 percent. Even in the overlap, the depositories have more loans in underserved areas (22 percent for GSEs vs. 27 percent for depositories). Much larger gaps appear in the non-overlap regions for underserved (8 percent for GSEs vs. 56 percent for depositories) and for percent in center cities (13 percent for GSEs vs. 38 percent for depositories). Similarly, the depositories provide more loans to low-income borrowers and high minority tracts.

The conclusion is that most loans held by depositories could have been sold to the GSEs, and thus the 78 percent overlap. However, in the non-overlap loans the depositories have smaller, riskier loans from young, low-income and minority homebuyers.

Exhibit 40

Distribution of Predicted Probability
of the GSE and Depository Loans
Origination Model Weighted

Distribution of Predicted Log Odds
of the GSE and Depository Loans
Origination Model Weighted
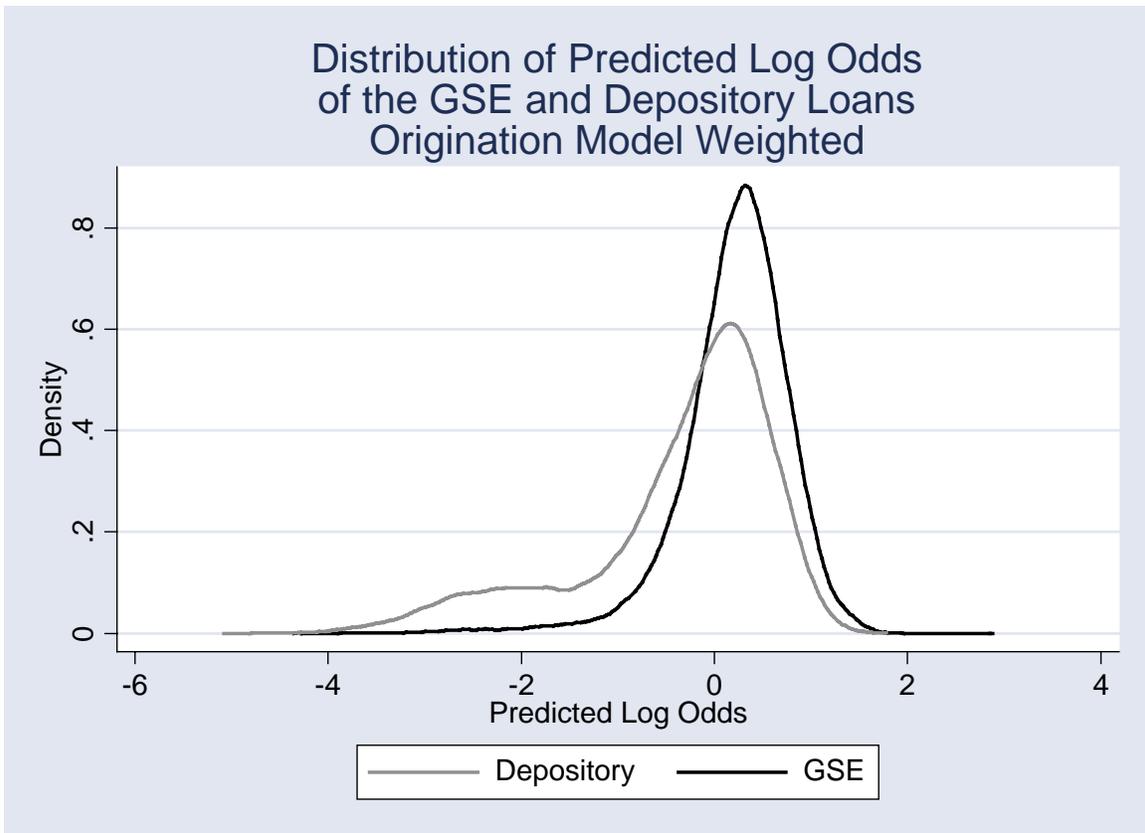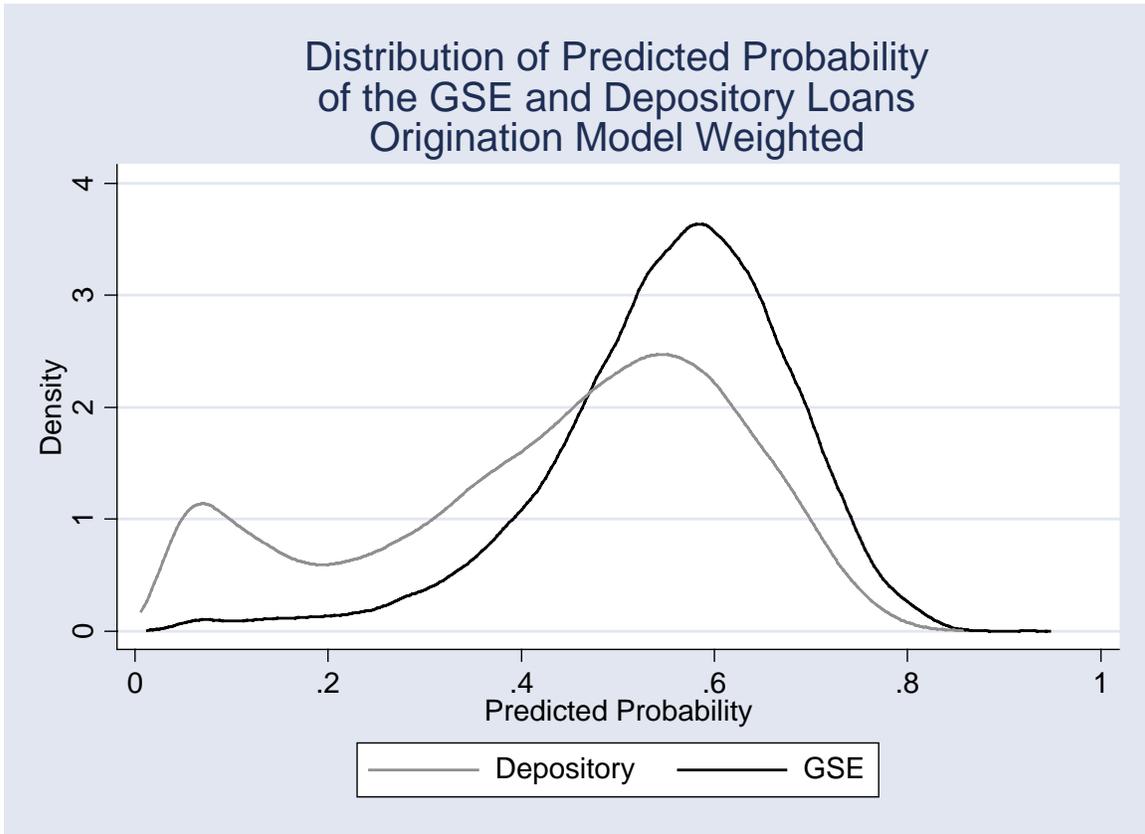
Exhibit 41: Characteristics of GSE and Depositories In and Out of Overlap by Non-Parametric Tolerance Interval Method

GSE=1, Depository=0
On All Matched, Conforming GSE or Depository Loans
(n=229,808, weighted sample=1,309,153)

	GSE (non-overlap)	GSE in Overlap	Combined Overlap	Depository in Overlap	Depository (non-overlap)
Share of Loans (10% in Tails)	4%	42%	78%	37%	9%
Borrower Characteristics					
Unweighted Number of Borrowers	9,336	93,655	175,266	81,611	22,486
Weighted Number of Borrowers	46,054	547,145	1,025,509	478,365	120,412
Average Annual Income	$75,635	$68,719	$68,798	$68,889	$49,836
Median Annual Income	$71,000	$61,000	$60,000	$59,000	$43,000
Average Annual Income (Trimmed Top 1%)	$74,240	$66,173	$65,852	$65,536	$47,316
% Estimated Income Information	1%	1%	1%	3%	2%
Average FICO	761	728	723	718	638
% With FICO <620	1%	8%	10%	12%	41%
% With FICO 620-680	6%	13%	14%	15%	21%
% With FICO =>680	93%	79%	76%	73%	38%
% Missing FICO Information	0%	0%	0%	0%	0%
% White	91%	77%	75%	74%	48%
% Black	0%	4%	4%	5%	27%
% Hispanic	2%	7%	8%	8%	17%
% Other	4%	9%	9%	10%	4%
% Missing Race Information	3%	3%	3%	3%	3%
% Female	14%	25%	26%	27%	38%
% Age 19-34	27%	29%	29%	29%	35%
% Age 35-49	53%	50%	50%	50%	50%
% Age 50-64	17%	16%	16%	16%	12%
% Age >65	2%	4%	4%	5%	3%
% Missing Age Information	0%	0%	0%	0%	0%
Loan Characteristics					
Average Loan Amount	$172,378	$136,882	$133,759	$130,187	$97,752
Average LTV %	81	79	78	78	91
% With LTV<=80	54%	59%	59%	59%	19%
% With LTV 80-90	21%	16%	17%	18%	5%
% With LTV 90-96	25%	21%	20%	18%	5%
% With LTV 96-98	0%	3%	4%	5%	22%
% With LTV>98	0%	0%	0%	0%	49%
% Missing LTV Information	0%	0%	0%	0%	0%
Average Ratio of Loan Amount to FHA Loan Limit	105%	78%	76%	73%	54%
% With LoantoFHA Ratio <=.5	2%	18%	21%	24%	46%
% With LoantoFHA Ratio of .5 - 1	42%	59%	58%	57%	52%
% With LoantoFHA Ratio of 1 - 1.2	26%	15%	14%	13%	2%
% With LoantoFHA Ratio of >1.2	30%	7%	7%	6%	1%

Exhibit 41 (cont.): Characteristics of GSE and Depositories In and Out of Overlap by Non-Parametric Tolerance Interval Method
GSE=1, Depository=0
On All Matched, Conforming GSE or Depository Loans
(n=229,808, weighted sample=1,309,153)

	GSE (non-overlap)	GSE in Overlap	Combined Overlap	Depository in Overlap	Depository (non-overlap)
Average PTI	0.21	0.19	0.19	0.19	0.19
% Originated in 1998	64%	33%	31%	29%	24%
% Originated in 1999	12%	33%	35%	37%	41%
% Originated in 2000	23%	34%	34%	35%	35%
Mortgaged Property Characteristics					
% New Construction	16%	11%	11%	11%	8%
% Unit Size 1	98%	95%	95%	95%	95%
% Unit Size 2	1%	2%	2%	2%	2%
% Unit Size 3	1%	2%	2%	2%	2%
% Unit Size 4	0%	1%	1%	1%	1%
Borrower Neighborhood Characteristics (1990 Census)					
% In Underserved Tracts	8%	22%	24%	27%	56%
% In High Cost MSA	3%	10%	12%	13%	15%
% In Medium Cost MSA	97%	87%	85%	83%	82%
% In Low Cost MSA	1%	3%	3%	3%	3%
% In Center City	13%	25%	25%	26%	38%
Average 5-yr House Price Appreciation, Lagged 1 year	117%	113%	113%	113%	112%
% In Area with Depreciation	13%	10%	10%	10%	8%
% In Area with Appreciation up to 20%	69%	72%	71%	70%	68%
% In Area with Appreciation over 20%	18%	19%	19%	20%	24%
% In Tracts with Income <90% of MSA Income	6%	15%	18%	20%	43%
% In Tracts with Income 90 - 120% of MSA Income	20%	33%	33%	33%	35%
% In Tracts with Income =>120% of MSA Income	74%	52%	49%	47%	21%
% In <10% Minority Tracts	63%	48%	48%	47%	33%
% In 10-30% Minority Tracts	32%	33%	32%	32%	27%
% In =>30% Minority Tracts	5%	18%	20%	21%	41%

Section 4: Default Model and Market Sector Overlap

The origination model predicts loan choice based on characteristics about the borrower, loan and property. Implicitly the underwriter is using that information to determine the default risk. An alternative approach is to use performance data (FHA claims) to determine the risk for combinations of borrower/loan/property combinations and assign a risk score to each loan. If risk were the only factor in underwriting, then the risk score might dominate loan choice in an origination model. However, if non-credit factors, such as borrower race or neighborhood income, are important considerations by lenders, then many non-credit variables would still be significant alongside the risk score in the origination model.

Two key assumptions are made in the default analysis. The first is that a default model estimated on FHA performance data is adequate for projecting the risk of conventional loans. It would have been far better to use conventional performance data to estimate default risk for conventional loans, but those data were not available. A specific issue is that the FHA data underlying the default model showed little variation in LTV ratios (most FHA LTVs are above 95 percent), which suggests that the LTV coefficient in the default model may be understated for measuring the effects of LTV on the default risk for conventional loans (see discussion below). A second important assumption is the credit measures included in the default model (FICO score, LTV and payment-to-income ratio) are sufficient to capture loan level risk. If the credit measures are not sufficient, then it is possible that the non-credit variables are serving as proxies for risk factors. For example, the neighborhood income or center city location may proxy for the expected rate of house price appreciation. More complete data, especially conventional performance data, were required to test these assumptions.

This section first uses a default model, estimated by Unicon on FHA claims data, to project a risk score and overlap; then, the risk score replaces the FICO score in the origination model. The predicted log odds of a claim become the risk score for each loan. The risk score is significant, but so are many non-credit variables. Without those non-credit variables, the overlap percentage is very high (60 percent). Just like with the FICO score (see Exhibit 15), there is a high degree of overlap in risk scores among FHA-eligible borrowers. But, in a full origination model with risk score and non-credit variables, the overlap is much smaller (11 percent). In fact, the model and overlap is very similar to the FHA/GSE origination model with FICO score.

Default Model on FHA Performance Data

While performance data for conventional loans are not available, they are available for FHA loans (along with credit scores from origination). The cohorts of FHA loans began in 1992, 1994 and 1996. The dependent variable used to measure performance is an indicator of claim four years after origination. By using the same covariates in the estimation as available in the Experian/HMDA matched data, the coefficients from the Default Model can be applied to the loans (both FHA and GSE) in the matched data to predict the risk of these loans. If there are many low-risk FHA loans that are below the upper risk limit of GSE loans, then this indicates many FHA loans probably could qualify as conventional loans.

Bob Cotterman at Unicon did the estimation of the Default Model using FHA performance data. He experimented with a number of specifications, which are shown as the final three estimated equations in Exhibit 42. The full model includes neighborhood characteristics and demographics. Claim is a

Exhibit 42: Default Model Estimated on FHA Data Originated in 1992, 1994 and 1996

Parameter	Label	Full Default Model		Drop Tract Chars.		Drop Demographics	
		Estimate	Pr > ChiSq	Estimate	Pr > ChiSq	Estimate	Pr > ChiSq
yr1992	Indicator for application year 1992	-0.396	0.000	-0.390	0.000	-0.401	0.000
yr1994	Indicator for application year 1994	-0.242	0.000	-0.249	0.000	-0.259	0.000
fico	Borrower Beacon (coborrower Beacon used when no borrower beacon available)	-0.007	0.000	-0.007	0.000	-0.007	0.000
fico620	Marginal spline of FICO at 620	-0.006	0.000	-0.006	0.000	-0.006	0.000
ltvratio	Ratio of loan to value (percentage)	0.034	0.000	0.034	0.000	0.034	0.000
payTOinc	Payment to income ratio	6.280	0.000	6.279	0.000	6.454	0.000
payTOinc20	Marginal spline of payment to income ratio at .20	-5.680	0.000	-5.537	0.000	-5.636	0.000
msaRelHP	House price to SMSA median (OFHEO SMSA house price index and 1990 Census SMSA median house price)	-0.898	0.000	-1.132	0.000	-1.205	0.000
origTO4yrs	SMSA level house price growth at 4 years relative to origination (as a percentage)	-0.038	0.000	-0.038	0.000	-0.038	0.000
condo	Indicator for condominium	-0.164	0.024	-0.248	0.001	-0.266	0.000
FRM30_4yrs	30 FRM mortgage interest rate 4 years after orig less the orig interest rate	-0.083	0.003	-0.087	0.002	-0.090	0.001
age36_49	Indicator for borrower between 36 and 49 years old	0.051	0.121	0.058	0.073	0.079	0.015
age50_64	Indicator for borrower between 50 and 64 years old	0.072	0.253	0.088	0.159	0.113	0.070
age65_99	Indicator for borrower between 65 and 99 years old	0.702	0.000	0.721	0.000	0.749	0.000
unmarried	Indicator for unmarried borrower	-0.070	0.021	-0.068	0.024	-0.061	0.043
CA	Indicator for California	0.893	0.000	0.881	0.000	0.897	0.000
CA92	Indicator for California and 1992 application year	-0.564	0.000	-0.544	0.000	-0.558	0.000
black	Indicator for black race/ethnicity	0.122	0.009	0.243	0.000		
hispanic	Indicator for hispanic race/ethnicity	0.095	0.042	0.123	0.007		
other	Indicator for non-white, non-black, non-hispanic, and non-asian race/ethnicity	0.196	0.056	0.230	0.025		
ccity	Indicator for central city	-0.075	0.015				
pctBLK	1990 Census tract percent black in tract	0.003	0.000				
tractRELmsa	1990 Census tract income to 1990 Census MSA income	-0.229	0.006				
povrt	1990 Census tract poverty rate (as a percentage)	0.006	0.051				
_cons	Intercept	-1.231	0.017	-1.182	0.019	-0.932	0.063

Pseudo R2		0.1040		0.1025		0.1015	
Log Likelihood-Intercept Only		-21075.667		-21075.667		-21075.667	
Log Likelihood-Intercept and Covariates		-18884.087		-18915.872		-18936.834	

relatively rare event, so it is not surprising that the goodness-of-fit for the logit model is modest (pseudo-R^2 of 0.10). The inclusion of neighborhood characteristics or demographics seems to have little impact on the fit. FICO has a negative impact on claims, whereas LTV and PTI have a positive impact.

It was important to control for unusual changes in interest rates and house prices that might have affected the claim rates for the specific cohorts used in the estimation. Therefore, the Default Model includes changes in house prices and interest rates during the 4-year period under examination, but then excludes or neutralizes those coefficients when the model is applied to the matched Experian/HMDA data. Similarly, indicators for California and the California recession of the early 1990s are included in the estimation model, but not applied in the subsequent risk scoring.

The predicted risks (either as predicted probability of claim or log odds of a claim) are shown in Exhibit 43 for the pooled MSA data. It is clearer in the log odds (lower panel) that there is more overlap in the default models than in the origination model. It also looks like the distributions are more "normal." Actually, the distributions still fail the normality test (sktest), but it is expected that there would be less difference between the overlap measured by the parametric method vs. the non-parametric method. The KS statistic is 0.48, which is in between the FICO credit score KS of 0.37 and the origination model KS of 0.83.

The combined overlap using the parametric tolerance on the pooled MSA data is 56 percent divided between GSE (19 percent) and FHA (36 percent). See Exhibit 44. There is little variation by MSA ranging from 48 percent overlap in Philadelphia to 59 percent in Tampa and Washington, D.C. The characteristics of the loans are shown in Exhibit 45. On theoretical grounds, the non-parametric overlap, which is slightly larger (60 percent vs. 56 percent as shown in Exhibit 46), is preferred. In terms of just the FHA distribution, 62 percent of FHA loans overlap in risk with the GSE risk distribution. Although this overlap suggests that many FHA loans have the same risk as conventional loans, as noted below, this overlap could be due to the default model not adequately measuring the effect of the LTV ratio on default risk for conventional loans. The characteristics for the non-parametric overlap are quite similar to the characteristics in the parametric overlap, so the non-parametric results shown in Exhibit 47 are focused on here.

With such a large risk overlap range, it is not surprising that there is more variation between the GSE and FHA loans in the overlap. Nevertheless, the characteristics of FHA loans in the overlap are relatively similar to the GSE loans. The average predicted likelihood of default for non-overlap GSE loans is 0.5 percent compared to 14 percent for non-overlap FHA loans. Within the overlap range, the comparison is 2.5 percent for GSE loans and 3.7 percent for FHA loans. In the non-overlap sectors, GSE median income is 55 percent greater than FHA ($62,000 vs. $40,000) even though all loans are FHA-eligible. In the overlap region, median incomes for GSE borrowers are only 8.5 percent larger than incomes for FHA borrowers ($51,000 vs. $47,000). The same pattern is repeated with FICO score, percent minority and loan amount. However, the difference in LTV within the overlap is nearly as large (GSE = 91 percent vs. FHA=98 percent) as the difference in the non-overlap sectors (GSE=87 percent vs. FHA=99 percent). This result reflects how concentrated FHA is in high LTV loans. The lack of variation in LTV by FHA loans suggests low LTVs are rarely used to offset

Exhibit 43

Distribution of Predicted Probability
of the GSE and FHA Loans
Default Model Weighted

Distribution of Predicted Log Odds
of the GSE and FHA Loans
Default Model Weighted
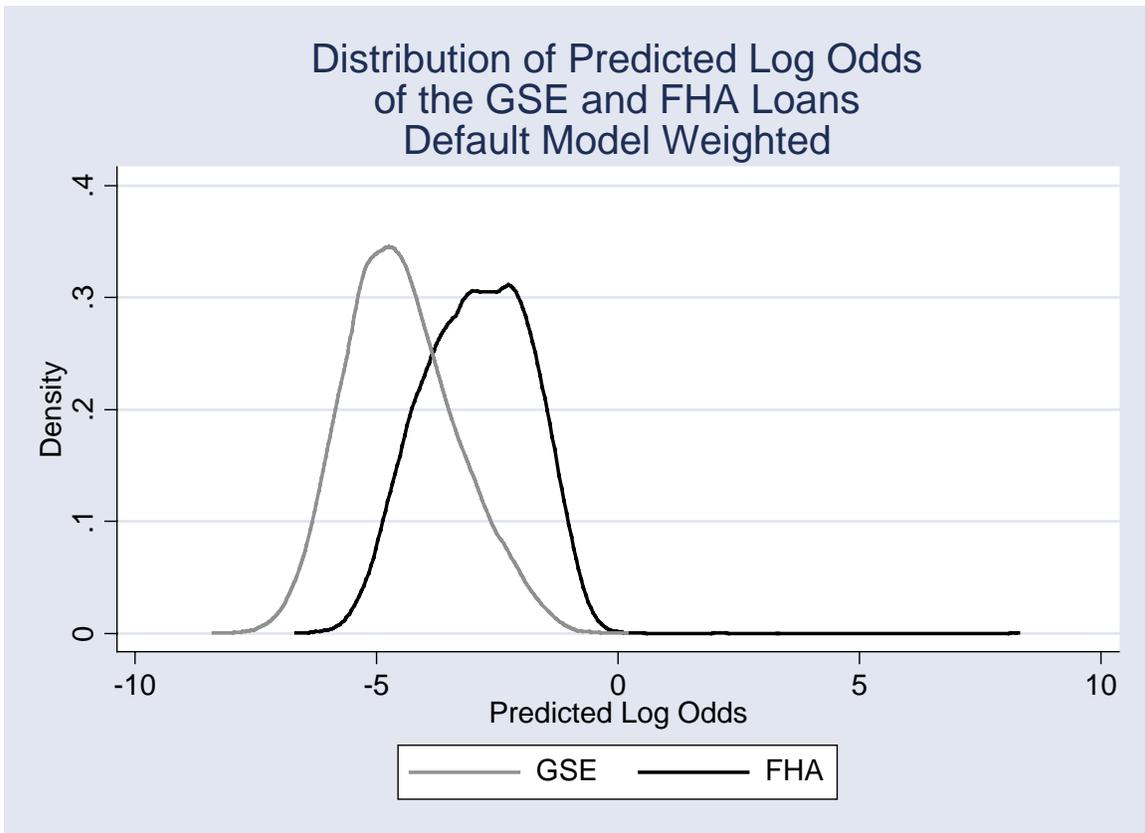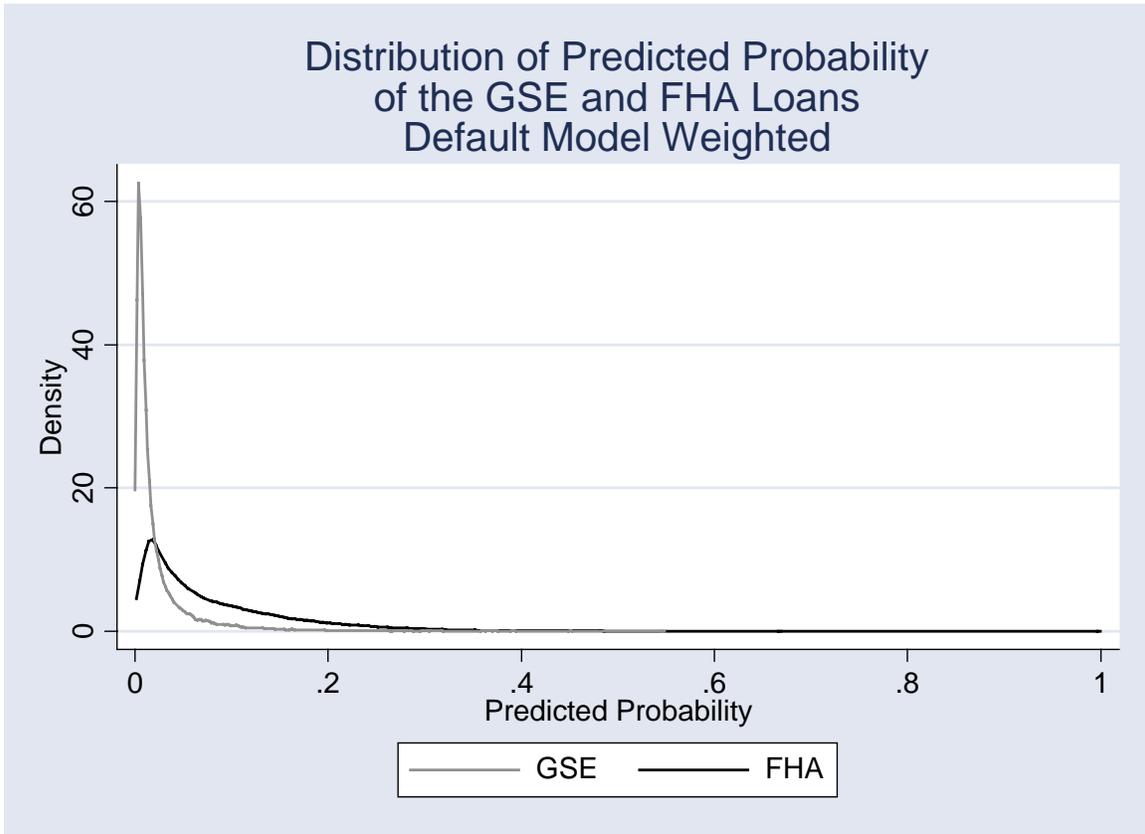
Exhibit 44: Percent of Loans in Default Model Market Sectors by Parametric Tolerance Intervals
FHA=1, GSE=0
On Matched, Conforming, Fixed-Rate FHA-Eligible FHA & GSE Loans with LTV 80-100%, Weighted
(n=114,780; weighted sample=674,238)

City	MSA	GSE (non-overlap)	GSE in Overlap	Combined Overlap	FHA in Overlap	FHA (non-overlap)	Tails
Baltimore	720	11%	10%	54%	44%	28%	8%
Chicago	1600	24%	23%	52%	29%	15%	9%
Cleveland	1680	20%	21%	58%	37%	14%	8%
Denver	2080	16%	15%	56%	41%	18%	9%
Los Angeles	4480	14%	16%	57%	41%	21%	8%
Oakland	5775	22%	23%	56%	33%	13%	9%
Philadelphia	6160	24%	16%	48%	32%	18%	9%
Portland	6440	21%	27%	57%	31%	13%	9%
St. Louis	7040	9%	8%	56%	48%	27%	9%
Tampa	8280	13%	17%	59%	42%	19%	9%
Washington DC	8840	12%	14%	59%	45%	21%	8%
Pooled		12%	19%	56%	36%	23%	9%

Exhibit 45: Analysis of Loans In Default Model Market Sectors (Parametric Tolerance Intervals)

FHA=1, GSE=0
On Matched, Conforming, Fixed-Rate FHA-Eligible FHA & GSE Loans with LTV 80-100%, Weighted
(n=114,780; weighted sample=674,238)

	GSE (non-overlap)	GSE in Overlap	Combined Overlap	FHA in Overlap	FHA (non-overlap)
Model Results					
Share of Loans (9% in Tails)	12%	19%	56%	36%	23%
Average Likelihood of Default (prediction)	0.005	0.023	0.030	0.033	0.128
Average Likelihood of Default (log odds)	-5.47	-3.94	-3.68	-3.56	-1.97
Borrower Characteristics					
Unweighted Number of Borrowers	14,752	20,653	64,551	43,898	24,618
Weighted Number of Borrowers	81,412	129,416	374,258	244,842	155,703
Average Annual Income	$67,465	$55,579	$53,018	$51,664	$43,180
Median Annual Income	$62,000	$51,000	$48,000	$47,000	$41,000
Average Annual Income (Trimmed Top 1%)	$65,457	$54,215	$51,243	$49,689	$42,388
% Estimated Income Information	1%	1%	1%	1%	1%
Average FICO	770	698	694	691	570
% With FICO <620	0%	9%	10%	10%	78%
% With FICO 620-680	1%	24%	28%	30%	21%
% With FICO =>680	99%	67%	62%	59%	1%
% Missing FICO Information	0%	0%	0%	0%	0%
% White	86%	67%	62%	60%	35%
% Black	2%	6%	11%	13%	31%
% Hispanic	4%	14%	19%	22%	29%
% Other	5%	9%	5%	3%	3%
% Missing Race Information	3%	4%	3%	2%	2%
% Female	24%	29%	28%	28%	38%
% Age 19-34	37%	34%	39%	42%	36%
% Age 35-49	47%	50%	48%	47%	53%
% Age 50-64	14%	13%	11%	10%	10%
% Age >65	2%	3%	2%	1%	2%
% Missing Age Information	0%	0%	0%	0%	0%
Loan Characteristics					
Average Loan Amount	$133,464	$121,948	$118,482	$116,650	$105,596
Average LTV %	87	91	96	98	99
% With LTV<=80	28%	12%	4%	0%	0%
% With LTV 80-90	38%	26%	11%	3%	1%
% With LTV 90-96	30%	47%	20%	6%	4%
% With LTV 96-98	4%	12%	15%	17%	15%
% With LTV>98	1%	4%	50%	74%	80%
% Missing LTV Information	0%	0%	0%	0%	0%
Average Ratio of Loan Amount to FHA Loan Limit	75%	66%	66%	65%	58%
% With LoantoFHA Ratio <=.5	10%	22%	22%	21%	36%
% With LoantoFHA Ratio of .5 - 1	90%	78%	78%	79%	64%
% With LoantoFHA Ratio of 1 - 1.2	0%	0%	0%	0%	0%
% With LoantoFHA Ratio of >1.2	0%	0%	0%	0%	0%

Exhibit 45 (cont.): Analysis of Loans In Default Model Market Sectors (Parametric Tolerance Intervals)
FHA=1, GSE=0
On Matched, Conforming, Fixed-Rate FHA-Eligible FHA & GSE Loans with LTV 80-100%, Weighted
(n=114,780; weighted sample=674,238)

	GSE (non-overlap)	GSE in Overlap	Combined Overlap	FHA in Overlap	FHA (non-overlap)
Average PTI	0.19	0.20	0.21	0.21	0.22
% Originated in 1998	35%	32%	31%	31%	29%
% Originated in 1999	34%	31%	35%	36%	35%
% Originated in 2000	31%	37%	34%	33%	36%
Mortgaged Property Characteristics					
% New Construction	10%	7%	8%	9%	6%
% Unit Size 1	94%	94%	95%	95%	96%
% Unit Size 2	2%	2%	2%	2%	2%
% Unit Size 3	3%	3%	2%	2%	2%
% Unit Size 4	1%	2%	1%	1%	1%
Borrower Neighborhood Characteristics (1990 Census)					
% In Underserved Tracts	18%	35%	38%	40%	61%
% In High Cost MSA	7%	13%	11%	9%	13%
% In Medium Cost MSA	91%	84%	86%	88%	84%
% In Low Cost MSA	2%	3%	3%	3%	3%
% In Center City	24%	30%	27%	26%	34%
Average 5-yr House Price Appreciation, Lagged 1 year	116%	112%	112%	112%	109%
% In Area with Depreciation	4%	14%	11%	10%	16%
% In Area with Appreciation up to 20%	76%	73%	74%	75%	74%
% In Area with Appreciation over 20%	20%	13%	14%	15%	10%
% In Tracts with Income <90% of MSA Income	14%	24%	26%	27%	41%
% In Tracts with Income 90 - 120% of MSA Income	34%	40%	41%	42%	40%
% In Tracts with Income =>120% of MSA Income	52%	36%	33%	31%	19%
% In <10% Minority Tracts	54%	36%	35%	35%	22%
% In 10-30% Minority Tracts	34%	36%	37%	37%	28%
% In =>30% Minority Tracts	11%	28%	28%	28%	49%

Exhibit 46: Percent of Loans in Default Model Market Sectors by Non-Parametric Tolerance Intervals
FHA=1, GSE=0
On Matched, Conforming, Fixed-Rate FHA-Eligible FHA & GSE Loans with LTV 80-100%, Weighted
(n=114,780; weighted sample=674,238)

City	MSA	GSE (non-overlap)	GSE in Overlap	Combined Overlap	FHA in Overlap	FHA (non-overlap)	Tails
Baltimore	720	10%	10.2%	62%	51%	19%	9%
Chicago	1600	22%	25%	57%	33%	12%	9%
Cleveland	1680	19%	21%	63%	42%	9%	9%
Denver	2080	15%	16%	61%	45%	14%	10%
Los Angeles	4480	12%	18%	62%	44%	16%	9%
Oakland	5775	20%	25%	61%	36%	10%	9%
Philadelphia	6160	20%	19%	57%	39%	13%	10%
Portland	6440	19%	28%	61%	33%	10%	9%
St. Louis	7040	8%	8%	61%	53%	21%	10%
Tampa	8280	12%	18%	64%	47%	14%	9%
Washington DC	8840	11%	15%	66%	51%	14%	9%
Pooled		11%	20%	60%	40%	19%	10%

Exhibit 47: Analysis of Loans In Default Model Market Sectors (Non-Parametric Tolerance Intervals)
FHA=1, GSE=0
On Matched, Conforming, Fixed-Rate FHA-Eligible FHA & GSE Loans with LTV 80-100%, Weighted
(n=114,780; weighted sample=674,238)

	GSE (non-overlap)	GSE in Overlap	Combined Overlap	FHA in Overlap	FHA (non-overlap)
Share of Loans (9% in Tails)	11%	20%	60%	40%	19%
Average Likelihood of Default (prediction)	0.005	0.025	0.033	0.037	0.137
Average Likelihood of Default (log odds)	-5.44	-3.90	-3.59	-3.46	-1.88
Borrower Characteristics					
Unweighted Number of Borrowers	13,962	21,500	69,984	48,484	19,518
Weighted Number of Borrowers	77,534	134,930	406,789	271,859	125,097
Average Annual Income	$67,007	$55,381	$52,660	$51,309	$42,439
Median Annual Income	$62,000	$51,000	$48,000	$47,000	$40,000
Average Annual Income (Trimmed Top 1%)	$65,095	$54,026	$50,921	$49,382	$41,758
% Estimated Income Information	1%	1%	1%	1%	1%
Average FICO	769	695	688	684	563
% With FICO <620	0%	11%	13%	14%	83%
% With FICO 620-680	1%	25%	29%	32%	17%
% With FICO =>680	99%	65%	58%	54%	1%
% Missing FICO Information	0%	0%	0%	0%	0%
% White	86%	66%	61%	58%	33%
% Black	2%	7%	11%	14%	33%
% Hispanic	4%	14%	20%	22%	29%
% Other	5%	9%	5%	3%	3%
% Missing Race Information	3%	4%	3%	2%	2%
% Female	24%	29%	29%	29%	39%
% Age 19-34	37%	34%	39%	41%	35%
% Age 35-49	47%	50%	48%	47%	53%
% Age 50-64	14%	13%	11%	10%	10%
% Age >65	2%	3%	2%	1%	2%
% Missing Age Information	0%	0%	0%	0%	0%
Loan Characteristics					
Average Loan Amount	$133,084	$121,612	$117,930	$116,103	$104,836
Average LTV %	87	91	96	98	99
% With LTV<=80	27%	11%	4%	0%	0%
% With LTV 80-90	38%	25%	10%	3%	1%
% With LTV 90-96	30%	47%	20%	6%	3%
% With LTV 96-98	4%	12%	15%	17%	15%
% With LTV>98	1%	4%	51%	74%	81%
% Missing LTV Information	0%	0%	0%	0%	0%
Average Ratio of Loan Amount to FHA Loan Limit	74%	66%	65%	65%	57%
% With LoantoFHA Ratio <=.5	10%	23%	22%	22%	37%
% With LoantoFHA Ratio of .5 - 1	90%	77%	78%	78%	63%
% With LoantoFHA Ratio of 1 - 1.2	0%	0%	0%	0%	0%
% With LoantoFHA Ratio of >1.2	0%	0%	0%	0%	0%

Exhibit 47 (cont.): Analysis of Loans In Default Model Market Sectors (Non-Parametric Tolerance Intervals)
FHA=1, GSE=0
On Matched, Conforming, Fixed-Rate FHA-Eligible FHA & GSE Loans with LTV 80-100%, Weighted
(n=114,780; weighted sample=674,238)

	GSE (non-overlap)	GSE in Overlap	Combined Overlap	FHA in Overlap	FHA (non-overlap)
Average PTI	0.19	0.20	0.21	0.21	0.22
% Originated in 1998	35%	32%	31%	30%	29%
% Originated in 1999	34%	31%	35%	36%	35%
% Originated in 2000	31%	37%	35%	33%	35%
Mortgaged Property Characteristics					
% New Construction	10%	7%	8%	8%	6%
% Unit Size 1	94%	94%	95%	95%	96%
% Unit Size 2	2%	2%	2%	2%	2%
% Unit Size 3	3%	3%	2%	2%	2%
% Unit Size 4	1%	2%	1%	1%	1%
Borrower Neighborhood Characteristics (1990 Census)					
% In Underserved Tracts	18%	36%	39%	41%	63%
% In High Cost MSA	7%	13%	11%	10%	13%
% In Medium Cost MSA	91%	84%	86%	87%	83%
% In Low Cost MSA	2%	3%	3%	3%	3%
% In Center City	25%	30%	27%	26%	35%
Average 5-yr House Price Appreciation, Lagged 1 year	116%	112%	112%	112%	108%
% In Area with Depreciation	5%	14%	11%	10%	16%
% In Area with Appreciation up to 20%	76%	73%	74%	75%	74%
% In Area with Appreciation over 20%	19%	13%	14%	15%	10%
% In Tracts with Income <90% of MSA Income	14%	25%	26%	27%	42%
% In Tracts with Income 90 - 120% of MSA Income	34%	40%	41%	42%	40%
% In Tracts with Income =>120% of MSA Income	51%	36%	32%	31%	18%
% In <10% Minority Tracts	54%	35%	35%	35%	21%
% In 10-30% Minority Tracts	35%	36%	36%	37%	28%
% In =>30% Minority Tracts	11%	29%	29%	29%	51%

poor credit history or high payment-to-income ratios. It is possible that the FHA-benchmarked default model is not as sensitive to variation in LTV as a model estimated on conventional performance data. It is also possible that low LTV is a stronger signal of a conventional loan in the origination model even if it does not fully capture the differences in risk estimated in the FHA-based default model.

The risk score is designed to capture all the risk factors that would affect the underwriting decision. However, does the choice of FHA vs. GSE loan depend only on risk or are non-credit factors also important. In Exhibit 48, the first column shows the FHA vs. GSE origination model (same as Exhibit 19). The middle columns show the results of a logistic regression on just the risk score. The coefficient is positive and significant, FHA loans are more risky, but the overall fit of the model is much worse than the full origination model (R^2 is 0.24 compared to R^2 of 0.64). The right columns show the result of replacing FICO with the risk score, but otherwise including the full specification from the origination model. The risk score is highly correlated with the FICO credit score (correlation is –0.87). Such high correlation indicates that the risk score captures many of the same effects as the credit score, and there would be a problem of multicollinearity if both risk score and credit score were included in the specification. The risk score coefficient is still positive and significant, and the pseudo R^2 has returned to 0.64. Apparently, the risk score is a good replacement of the FICO credit score. The tract percent minority has become negative and significant. The high-cost MSA indicator has become insignificant. Otherwise, the other coefficients are similar.

Given the similarity in fit, it is expected that the overlap is also similar, as shown in Exhibit 49. In fact, the overlap is 11 percent and the characteristics (Exhibit 50) are virtually identical to the characteristics in Exhibit 27 for the non-parametric overlap using FICO scores. In terms of just the FHA distribution, 10 percent of FHA loans are in the overlap region. The overlap FHA incomes and FICO scores are higher than the overlap GSE incomes and FICO scores. Those lower risk characteristics for FHA overlap loans may just offset younger borrowers and slightly higher average LTV. The main point is that the overlap loans, FHA and GSE, are very similar in overall risk.

The risk score showed a high degree of overlap, but that may have been because a default model based on FHA performance is not a good proxy for the default risk on non-FHA loans. In effect, the default model was not very accurate for assigning risk. When the risk score was used in the origination model, the risk score alone could not explain the loan choice, FHA vs. GSE. It appears that the risk score is a good proxy for the FICO credit score, which is an important element of the risk score, but there are other factors important to the loan choice or underwriting decision (FHA vs. GSE). The results confirm that about 10 percent of FHA loans have characteristics, including risk profile, comparable to conventional borrowers. However, conventional performance data were needed to more precisely measure the overlap between FHA default risk and GSE default risk.

Exhibit 48: Origination Model with Predicted Risk of Default Replacing FICO

FHA=1, GSE=0
On Matched, Conforming, Fixed-Rate FHA-Eligible FHA & GSE Loans with LTV 80-100%, Weighted
(n=114,780; weighted sample=674,238)

Parameter	Label	Full Origination Model		Model Origination On Predicted Risk Only		Full Origination Model Replacing FICO with Predicted Risk	
		Robust Coefficient	P>z	Robust Coefficient	P>z	Robust Coefficient	P>z
Intercept	Intercept	-1.9151	0.000	4.6464	0.000	0.4335	0.329
Risk Score	Log Odds of Risk of Default			1.0984	0.000	0.6419	0.000
ex_fico	FICO Score	-0.0033	0.000				
FICO_620	Marginal spline for FICO above 620	-0.0051	0.000				
EX_LTV	Loan to Value Ratio	0.0441	0.000			0.0156	0.000
LTVs95	Marginal spline for LTV above 95%	1.2346	0.000			1.2333	0.000
LTVs98	Marginal spline for LTV above 98%	1.2503	0.000			1.2618	0.000
LTVs99	Marginal spline for LTV above 99%	-4.3026	0.000			-4.3246	0.000
rounded loantoFHA	Ratio of Borrower's Loan to FHA Loan Limit	-1.9523	0.000			-1.1058	0.000
PTI_calc	Payment to Income Ratio- Calculated	4.8723	0.000			0.9527	0.038
PTIs20	Marginal spline for PTI above .20	0.0896	0.917			3.6363	0.000
PTIs28	Marginal spline for PTI above .28	-4.9366	0.000			-4.9444	0.000
age_3549	Borrower Aged 35-49	-0.0722	0.013			-0.1018	0.000
age_5064	Borrower Aged 50-64	-0.1266	0.003			-0.1762	0.000
age_65	Borrower Aged 65 and up	-0.3035	0.001			-0.7615	0.000
female	Female Borrower	0.1319	0.000			0.1191	0.000
race_black	African American Borrower	0.5247	0.000			0.4224	0.000
race_hisp	Hispanic Borrower	0.3679	0.000			0.3306	0.000
race_other	Other Borrower (Native American, Asian, or Other)	-0.3050	0.000			-0.4188	0.000
new_constr	New Construction	0.3318	0.000			0.3359	0.000
relinc_c90t	Census Tract family income relative to MSA median family income in 1990	0.1294	0.010			0.3574	0.000
c90t_pminorty	Pct minority	0.0002	0.778			-0.0016	0.043
ccity_n	Center City tract indicator	-0.2728	0.000			-0.2308	0.000
high_cost	High Cost City based on FHA Loan Limits	-0.1311	0.020			-0.0774	0.170
low_cost	Low Cost City based on FHA Loan Limits	0.0724	0.302			0.0603	0.389
hpiL5	OFHEO House Price 5-yr change lagged 1 yr	-0.9216	0.002			-1.0051	0.001
year98	Loan Originated in 1998	0.1623	0.000			0.0766	0.025
year00	Loan Originated in 2000	-0.3526	0.000			-0.3105	0.000
Baltimore	Property in Baltimore	0.4400	0.000			0.6438	0.000
Chicago	Property in Chicago	-0.2612	0.000			-0.0434	0.513
Cleveland	Property in Cleveland	-0.0651	0.521			0.2813	0.006
Denver	Property in Denver	0.4179	0.000			0.8029	0.000
LosAngeles	Property in Los Angeles	0.0118	0.867			-0.1054	0.132
Oakland	Property in Oakland	-0.2842	0.002			-0.4278	0.000
Philadelphia	Property in Philadelphia	-0.0603	0.295			0.0633	0.270
Portland	Property in Portland	-0.2656	0.053			0.2801	0.040
StLouis	Property in St Louis	0.1995	0.018			0.4347	0.000
Tampa	Property in Tampa	0.1756	0.011			0.3265	0.000
Pseudo R2		0.6369		0.2416		0.6372	
Combined K-S Statistic		0.8255(P-Value=0.000)				0.8255(P-Value=0.000)	

Exhibit 49

Distribution of Predicted Probability
of the GSE and FHA Loans
Origination Model with Risk, Weighted

Distribution of Predicted Log Odds
of the GSE and FHA Loans
Origination Model with Risk, Weighted
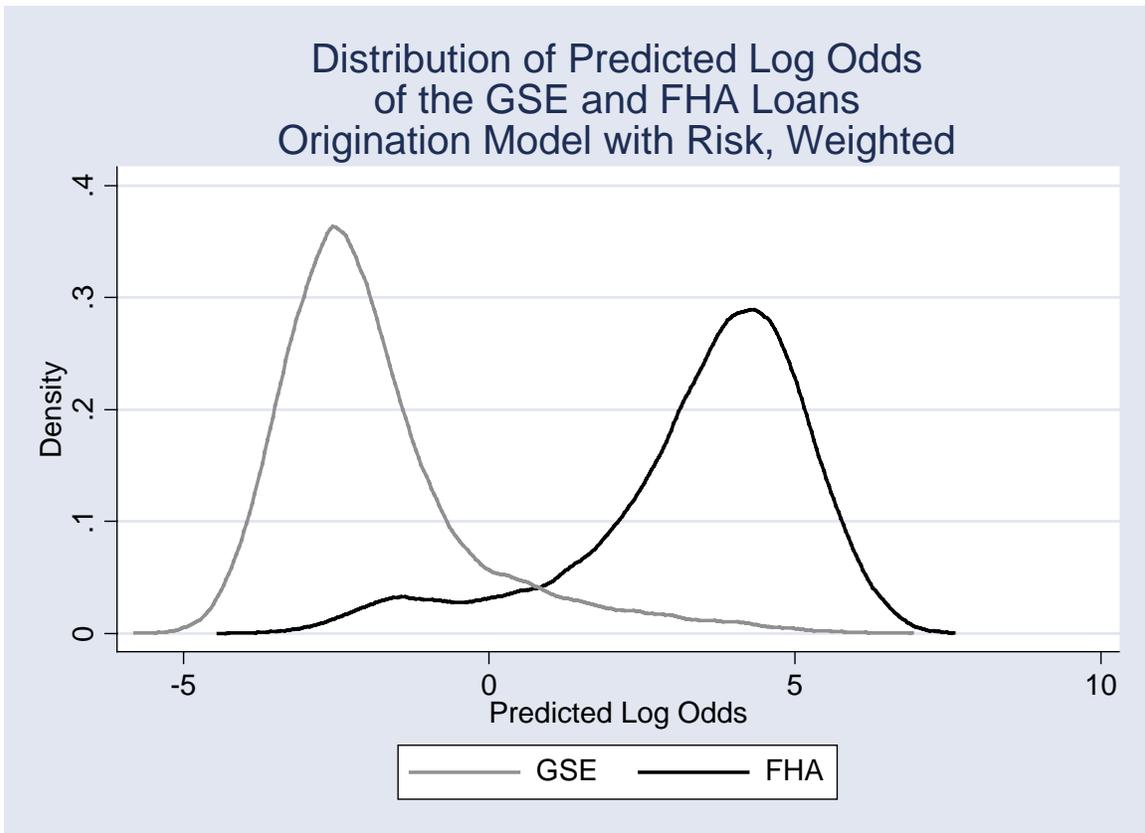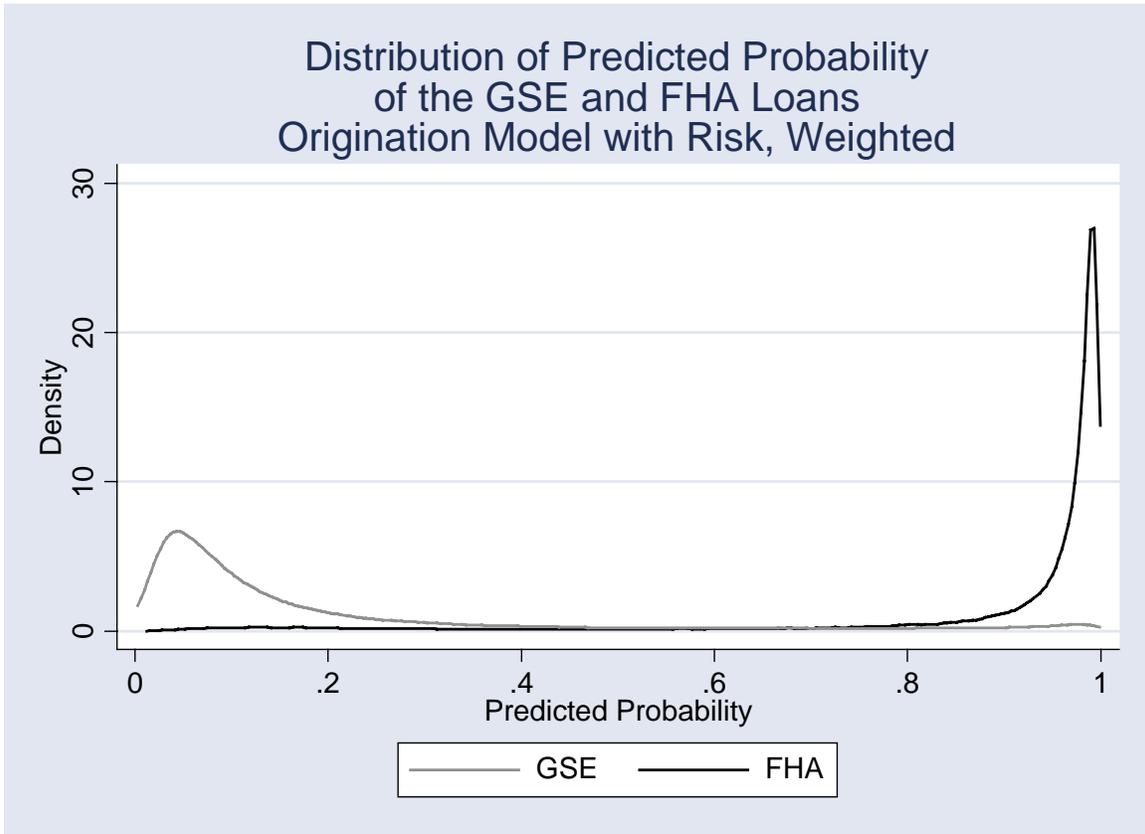
Exhibit 50: Analysis of Loans In Market Sectors Of Origination Model with Predicted Risk of Default Replacing FICO (Non-Parametric Tolerance Intervals)

FHA=1, GSE=0

On Matched, Conforming, Fixed-Rate FHA-Eligible FHA & GSE Loans with LTV 80-100%, Weighted
(n=114,780; weighted sample=674,238)

	GSE (non-overlap)	GSE in Overlap	Combined Overlap	FHA in Overlap	FHA (non-overlap)
Share of Loans (10% in Tails)	27%	5%	11%	6.58%	52%
Borrower Characteristics					
Unweighted Number of Borrowers	29,857	5,606	12,463	6,857	61,146
Weighted Number of Borrowers	180,978	32,007	76,341	44,334	349,158
Average Annual Income	$60,495	$48,608	$52,396	$55,132	$47,983
Median Annual Income	$56,000	$45,000	$46,000	$47,000	$44,000
Average Annual Income (Trimmed Top 1%)	$58,840	$47,564	$49,444	$50,891	$46,634
% Estimated Income Information	1%	1%	1%	1%	1%
Average FICO	726	662	671	677	642
% With FICO <620	6%	30%	27%	25%	37%
% With FICO 620-680	14%	22%	21%	20%	26%
% With FICO =>680	79%	48%	52%	54%	37%
% Missing FICO Information	0%	0%	0%	0%	0%
% White	75%	59%	61%	62%	49%
% Black	4%	15%	13%	13%	21%
% Hispanic	10%	19%	18%	18%	24%
% Other	8%	5%	5%	5%	3%
% Missing Race Information	3%	2%	3%	3%	2%
% Female	26%	35%	36%	36%	31%
% Age 19-34	34%	37%	38%	38%	40%
% Age 35-49	49%	51%	49%	48%	49%
% Age 50-64	14%	11%	11%	12%	10%
% Age >65	3%	1%	2%	2%	2%
% Missing Age Information	0%	0%	0%	0%	0%
Loan Characteristics					
Average Loan Amount	$127,226	$111,446	$112,394	$113,078	$111,730
Average LTV %	88	95	96	96	99
% With LTV<=80	20%	2%	1%	0%	0%
% With LTV 80-90	35%	8%	7%	6%	0%
% With LTV 90-96	45%	37%	26%	19%	0%
% With LTV 96-98	0%	54%	64%	71%	11%
% With LTV>98	0%	1%	2%	3%	89%
% Missing LTV Information	0%	0%	0%	0%	0%
Average Ratio of Loan Amount to FHA Loan Limit	70%	61%	61%	61%	62%
% With LoantoFHA Ratio <=.5	17%	32%	32%	32%	28%
% With LoantoFHA Ratio of .5 - 1	83%	68%	68%	68%	72%
% With LoantoFHA Ratio of 1 - 1.2	0%	0%	0%	0%	0%
% With LoantoFHA Ratio of >1.2	0%	0%	0%	0%	0%

Exhibit 50 (cont.): Analysis of Loans In Market Sectors Of Origination Model with Predicted Risk of Default Replacing FICO (Non-Parametric Tolerance Intervals)
FHA=1, GSE=0
On Matched, Conforming, Fixed-Rate FHA-Eligible FHA & GSE Loans with LTV 80-100%, Weighted
(n=114,780; weighted sample=674,238)

	GSE (non-overlap)	GSE in Overlap	Combined Overlap	FHA in Overlap	FHA (non-overlap)
Average PTI	0.20	0.21	0.20	0.20	0.21
% Originated in 1998	34%	32%	34%	35%	29%
% Originated in 1999	33%	34%	32%	30%	36%
% Originated in 2000	34%	34%	34%	34%	34%
Mortgaged Property Characteristics					
% New Construction	8%	9%	9%	9%	7%
% Unit Size 1	94%	93%	93%	93%	96%
% Unit Size 2	2%	2%	2%	2%	2%
% Unit Size 3	3%	3%	3%	3%	2%
% Unit Size 4	2%	2%	3%	3%	1%
Borrower Neighborhood Characteristics (1990 Census)					
% In Underserved Tracts	28%	41%	39%	37%	50%
% In High Cost MSA	11%	11%	12%	12%	10%
% In Medium Cost MSA	86%	86%	85%	85%	86%
% In Low Cost MSA	3%	3%	3%	3%	3%
% In Center City	28%	29%	28%	27%	30%
Average 5-yr House Price Appreciation, Lagged 1 year	113%	113%	112%	111%	111%
% In Area with Depreciation	11%	12%	13%	14%	11%
% In Area with Appreciation up to 20%	74%	73%	75%	76%	74%
% In Area with Appreciation over 20%	15%	15%	12%	9%	14%
% In Tracts with Income <90% of MSA Income	20%	29%	27%	25%	35%
% In Tracts with Income 90 - 120% of MSA Income	37%	38%	38%	38%	40%
% In Tracts with Income =>120% of MSA Income	43%	33%	35%	36%	25%
% In <10% Minority Tracts	44%	33%	30%	29%	31%
% In 10-30% Minority Tracts	34%	36%	38%	40%	32%
% In =>30% Minority Tracts	22%	31%	31%	31%	37%

Section 5: Tract Level Modeling

This section is an exploratory analysis motivated by the importance of identifying and understanding factors that determine the size of the FHA's market share. Previously, there was a simple comparison of various characteristics of the 11 MSAs that are in the study. However, the relationships between the FHA's market share and MSA characteristics were not consistent across the MSAs, perhaps due to the intentional heterogeneity in sampled MSAs. In addition, a sample of 11 MSAs did not provide adequate sample size for a meaningful regression analysis at the MSA level. In this section, neighborhoods denoted by census tracts are chosen as the unit of analysis to overcome small sample size. Multiple regression analysis is used to control for many factors at the same time. The results reported here support the findings of recent research done by Ambrose, Pennington-Cross and Yezer (2002) at the MSA level. For example, higher income and house prices have a negative impact of FHA market share, whereas percent racial minority and average default rate have a positive impact of FHA market share. However, the current model is simplistic and preliminary, and it could be extended into hierarchical models that capture both the MSA and neighborhood effects considerably better.

In their FHA wedge model and empirical results briefly discussed in the literature review section of this report, Ambrose, Pennington-Cross and Yezer show that FHA underwriting follows a national standard and the FHA share increases in weak markets. The authors conclude that conventional underwriting does not adjust to local factors in order to maintain market share. Rather, non-price credit rationing by conventional lenders leaves the FHA with the role of maintaining the mortgage credit supply in declining housing markets. The authors find that FHA's market share increases with increases in unemployment rate, increases in delinquency rate (lagged one year), higher average delinquency rate, higher volatility of house prices, higher shares of incomes below $20,000 and higher percentage of minority (lagged one year), and it decreases with increases in house prices (current and lagged one year), higher percentage of loan to income ratio greater than three, and higher black Gini coefficient.

Prior to conducting the regression analysis, the number of loans in the census tracts in the Experian/HMDA data set was checked to see if there were sufficient loans for a reasonable analysis. The universe was defined to be all FHA and GSE loans that were originated from 1998 through 2000 as listed in the Experian/HMDA data set. There were 209,486 such loans. Then the FHA/GSE data were reduced to the FHA eligible loans (171,288), which include 93,606 FHA loans. In all three data sets, there were enough tracts with adequate loans per tract for a decent tract level analysis. It should be noted that all census tracts that are in the 11 MSAs are not included in the final data set, but rather only a subset sampled in the Experian data. As shown in Exhibit 51, 80 percent of 4,245 census tracts in the total sample of all FHA and GSE loans, 74 percent of 4,198 census tracts in the FHA eligible sample, and 57 percent of 3,799 tracts in the FHA loan sample have 11 or more loans in each tract. Only 12 percent of tracts in the total sample, 15 percent of tracts in the FHA eligible sample and 27 percent of tracts in the FHA loans sample have five or fewer loans in each tract. Moreover, the distributions of loan counts for each of the 11 MSAs are similar to the overall distribution.

Exhibit 51: Tract Counts Corresponding to Loan Counts Among FHA and GSE Loans

Number of Loans in Tract	Number of Tracts		
	All Loans (n=209,486)	FHA Eligible Loans (n=171,288)	FHA Loans (n=93,606)
1-5	497	650	1,043
6-10	370	437	583
11-20	684	720	706
21-30	518	549	474
31-50	793	751	497
51-100	890	718	357
101+	493	373	139
Total Tracts	4,245	4,198	3,799

For each census tract, FHA's market share was calculated by taking the ratio of FHA loans to the FHA-eligible GSE and FHA loans. Then a few relevant census tract level variables from the 1990 Census were appended to the data set of tract counts and FHA share. The 1990 Census data were chosen because it is the most recent Census data that would have been known to lenders in 1998-2000. One obvious disadvantage of using the 1990 Census data is that the values of variables could have changed between 1990 and 1998 and the results might not show the exact extent to which the census tract characteristics are important in determining the FHA's market share.

The tract-level default rate of FHA loans from the Section 335 data is added to the data set with FHA's share, tract counts and Census characteristics. Section 335 data refer to the data HUD is required to collect on the status of the most recent 20 quarters of FHA loans by originating lender, census tract and year of origination as specified in Section 335 of the Affordable Housing Act of 1990. In order to gain sufficient sample size per census tract, the default rate was calculated by pooling years 1999 to 2001 and by taking the ratio of total default (delinquencies and foreclosures) to total loans in the entire time period of those three years.

Some variables such as the median household income and the median value of owner-occupied housing were only available in the top-coded form from the Census data and all variables in the model were used as obtained from the Census data. Some variables such as the poverty rate, which at first glance seemed to be a critical variable influencing FHA share, were highly correlated with other variables that could be included in the model. For example, a combination of minority rate, median household income and median value of owner-occupied housing, all of which were highly correlated with poverty rate, were better at explaining FHA share than poverty rate alone. Thus poverty rate was excluded from the model and other variables were included. Based on a study of correlations among potential explanatory variables, only seven independent variables were included in the model in addition to the MSA dummies. They are: minority rate, owner occupancy rate, median value of owner-occupied housing, default rate for FHA loans, percentage of households with heads aged 15 to 24, percentage of households with heads aged 65 or above and household median income. Exhibit 52 provides the descriptive statistics for these explanatory variables as well as the counts of census tracts in each of the 11 MSAs.

Exhibit 52: Tract Level Descriptive Statistics of FHA Eligible FHA and GSE Loans

Variables	# of Tracks	Mean	Std Err	Min	Max
FHA Share Among FHA Eligible Loans as Percent	4245	55.0	32.7	0	100.0
Percent Minority	4240	30.8	32.9	0	100.0
Percent Owner Occupied Housing Units	4240	62.5	23.4	0	100.0
Median Value of Owner Occupied Housing Unit ($1,000)	4245	$114	$85	0	$500
Percent of FHA Loans Defaulted	4077	6.9	9.7	0	100.0
Percent of Households with Head between 15 and 24	4240	4.4	4.2	0	92.7
Percent of Households with Head 65 or over	4240	21.8	10.9	0	91.0
Median Household Income ($1,000)	4245	$35	$16	0	$150

MSA	Freq
Baltimore	170
Chicago	530
Cleveland	588
Denver	336
Los Angeles	382
Oakland	203
Philadelphia	563
Portland	334
St Louis	391
Tampa	394
Washington DC	349

Exhibit 53 provides the regression results for the FHA's market share among the FHA eligible GSE and FHA loans. The initial estimation was a simple OLS model that included all seven independent variables but no MSA dummies. Analytic weights were also used to account for varying sample sizes of the FHA eligible loans in different census tracts. The regression results showed that the FHA's share in a census tract increases with higher minority rate, higher owner occupancy, lower median value of owner occupied housing, higher FHA default rate, higher percentage of households with heads between 15 and 24, lower percentage of households with elderly heads and lower median household income. All variables were significant at the five percent level but the explanatory power as shown by the adjusted R^2 of 0.34 was relatively low.

Next the MSA dummies were included to control for common characteristics of the census tracts in each of the 11 different MSAs in the model. Washington, DC was designated as the reference MSA. The explanatory power of the OLS model with the MSA dummies was better than the simple OLS model with an adjusted R^2 of 0.56. Signs of all the coefficients were as expected and consistent with the simple OLS model and they were all significant at the five percent level. In both regressions, multicollinearity was not a major concern since the correlation among independent variables was carefully studied and only variables with acceptable correlation with others were included.

Given the bounded values of the dependent variable—FHA share—between 0 and 100 percent, a logit regression model is more appropriate than a simple OLS model. However, the explanatory power of the logit regression model with the MSA dummies as shown in Exhibits 53 is poor despite the use of analytic weights derived from sample sizes and the inclusion of the MSA dummies. The pseudo-R^2s of the logit models are not directly comparable with the adjusted R^2s of the OLS models but the low pseudo-R^2s of the logit models are worrisome. In addition, signs of the coefficients for owner occupancy rate, percentage of households with heads between 15 and 24, percentage of household with elderly heads and median household income are reversed from the OLS models and only median value of owner occupied housing, percentage of households with elderly head and a few MSA dummies are significant at the five percent level.

Despite the questionable logit results, the general findings of the OLS models, especially the OLS model with the MSA dummies support the findings by Ambrose, Pennington-Cross and Yezer (2002). Variation across MSAs is important in explaining differences in FHA's share of the market. The FHA market share was also found to be higher for low income, minority, and young borrowers, and for lower house values.

Given the findings of the OLS models and the limitations of the logit model, the following recommendations are made for future research.

First, the matched sample of loans was drawn from an intentionally divergent, but small number of MSAs. Better results would be obtained by including more MSAs and a nationally representative set of MSAs. That way the included tracts would be more representative of a typical metropolitan area.

Exhibit 53: Regression of FHA Share as a Percent Among FHA Eligible FHA and GSE Loans

Independent Variables	OLS (n=4,054)			OLS with MSA Dummies (n=4,054)			Logit with MSA Dummies (n=4,054)		
	Coeff	Std Err		Coeff	Std Err		Coeff	Std Err	
Percent Minority	0.4468	0.0172	*	0.1579	0.0167	*	0.0102	0.0108	
Percent Owner Occupied Housing Units	0.1235	0.0368	*	0.0886	0.0319	*	-0.0174	0.0135	
Median Value of Owner Occupied Housing Unit ($1,000)	-0.1062	0.0115	*	-0.2537	0.0129	*	-0.0200	0.0035	*
Percent of FHA Loans Defaulted	0.5721	0.0679	*	0.5593	0.0584	*	0.0440	0.0323	
Percent of Households with Head between 15 and 24	1.5002	0.1870	*	1.0331	0.1555	*	-0.0709	0.0576	
Percent of Households with Head 65 or over	-0.1538	0.0445	*	-0.1536	0.0390	*	0.0666	0.0169	*
Median Household Income ($1,000)	-0.2039	0.0704	*	-0.3624	0.0672	*	0.0155	0.0211	
Baltimore				-9.7170	1.7465	*	-0.8948	1.5398	
Chicago				-37.4125	1.3112	*	-2.3497	0.7813	*
Cleveland				-48.6121	1.7112	*	-3.8838	0.8888	*
Denver				-27.4722	1.3665	*	-0.8219	1.0633	
Los Angeles				-4.7147	1.7952	*	-0.4271	0.9815	
Oakland				-25.7684	2.0367	*	-1.7505	0.8385	*
Philadelphia				-30.4187	1.4775	*	-2.6655	0.8221	*
Portland				-54.1434	1.5556	*	-1.6454	1.2031	
St Louis				-51.6383	1.5242	*	-5.7108	0.7891	*
Tampa				-39.8972	1.5496	*	-1.3264	1.3387	
Constant	52.0651	3.7722	*	116.0388	3.7216	*	8.6005	1.6160	*
R-Squared **	0.3383			0.5561			0.3466		
Log Likelihood							-313.0541		
Log Likelihood with Only Constant							-479.0995		

Notes:

* denotes significance at the 5 percent level.

** The OLS models have adjusted R-Squared's and the logit model has psuedo-R-Squared.

Second, Ambrose, Pennington-Cross and Yezer (2002) show that changes of neighborhood characteristics may be more important than static levels at any given time. Therefore, it is recommended that a time component be added to the neighborhood characteristics. For example, a panel of loans from 2001-2003 could use Census characteristics from both the 1990 and the 2000 Census data to provide the time element without running into the endogeneity problem faced by the 1998-2000 sample. It is expected that the GSEs will dominate strong and growing markets. But higher goals may push the GSEs towards more purchases in low-income and minority area with a preference for loans in such areas with increasing house values to offset low income or low FICO scores.

Third, many other variables can be added to the model to capture trends favorable or unfavorable to FHA originations. For instance, as house prices rise, low-income borrowers get priced out of the conventional market, but they stretch their buying power relative to their income by switching to FHA loans. LTV is an endogenous variable if borrowers jointly choose between an FHA or conventional loan and the loan amount. Following Ambrose and Pennington-Cross (2000), the LTV ratio could be treated as an instrument variable in the model. There could be experimentation with the average and the distribution of FICO scores. There may be a higher FHA share among those with missing FICO scores. Also, measurement of neighborhood income as well as ethnic integration such as the black Gini coefficient used by Ambrose, Pennington-Cross and Yezer can be used. Finally, measures of concentration among local lenders can be used because steering would be easier when there is less competition and fewer opportunities to shop.

In conclusion, the simplistic models support the finding of recent research but they could be extended to perform considerably better with hierarchical models that capture both the MSA and neighborhood effects.

References

Ambrose, Brent W. and Anthony Pennington-Cross (2000) "Local Economic Risk Factors and the Primary and Secondary Mortgage Markets," *Regional Science and Urban Economics*, 30: 683-701.

Ambrose, Brent W., Anthony Pennington-Cross and Anthony M. Yezer (2002) "Credit Rationing in the U.S. Mortgage Market: Evidence from Variation in FHA Market Shares," *J. of Urban Economics*, 51: 272-294.

Amemiya, Takeshi (1985) *Advanced Econometrics*, Harvard University Press.

Becker, Gary S. (1971) *The Economics of Discrimination, 2nd edition*. (Chicago: University of Chicago Press).

Berkovec, James A., Glenn B. Canner, Stuart A. Gabriel and Timothy H. Hannan (1998) "Discrimination, Competition, and Loan Performance in FHA Mortgage Lending," *Review of Economics and Statistics*, Vol. 80(May): 241-250.

Bunce, Harold, Charles A. Capone, Sue G. Neal, William J. Reeder, Randall M. Scheessele and Edward J. Szymanoski, (1995) *An Analysis of FHA's Single-Family Insurance Program*, U.S. Dept. of Housing and Urban Development, Office of Policy Development and Research, October.

Bunce, Harold (2000) "An Analysis of GSE Purchases of Mortgages for African-American Borrowers and Their Neighborhoods," HUD, PD&R, Housing Finance Working Paper Series (HF-011).

Cotterman, Robert F. (2001) *Neighborhood Effects in Mortgage Default Risk*, U.S. Dept. of Housing and Urban Development, Office of Policy Development and Research.

Cotterman, Robert F. (2002) *New Evidence on the Relationship Between Race and Mortgage Default: The Importance of Credit History Data*, U.S. Dept. of Housing and Urban Development, Office of Policy Development and Research.

Freeman, Lance, George Galster and Ron Malega (2003) "The Impact of Secondary Mortgage Market and GSE Purchases on Underserved Neighborhood Housing Markets: A Cleveland Case Study," paper presented at the ASSA-AREUEA Annual Conference in Washington, DC.

Gyourko, Joseph and Dapeng Hu (2002) "Spatial Distribution of Affordable Home Loan Purchases in Major Metropolitan Areas: Documentation and Analysis," *Regional Science and Urban Economics*, 32(5): 549-589.

Muolo, Paul, (2004) "A New Focus at Freddie," *National Mortgage News*, October 25,2004, pp. 1,86.

Onder, Zeynep (1998) "Individual and Neighborhood Effects on FHA Mortgage Activity: Evidence from HMDA Data," *Journal of Housing Economics*, 7:343-376.

Pennington-Cross, Anthony and Joseph Nichols, (2000) "Credit History and FHA-Conventional Choice," *Real Estate Economics,* 28: 307-336.

Appendix: Notes on Variables and Calculations

Exhibit A.1

	Notes on Calculations
Weighting Variable	OBS_POST_WGHT
Share of Loans	Weighted share of loans in given category divided by weighted number of borrowers.

MSA Characteristics

Designated High Cost/Low Cost (at least 1 year)	Based on FHA loan limits. MSAs with highest or lowest FHA single-unit loan limit by ACTN_DT designated high cost or low cost.
% Minority (2000)	Sum of individuals not categorized as white divided all individuals by MSA.
% Individuals in Poverty (2000)	Percent of individuals living in poverty by MSA.
Median Household Income (2000)	Median household income by MSA.
% Households Owner-Occupied (2000)	Owner-occupied households divided total number of households by MSA.
Median Owner-Occupied HH Value (2000)	Median owner-occupied household value.
# Tracts in Our Data	Count of unique tracts.
% Underserved Tracts in Our Data	Count of unique tracts in SERVED_N.

Borrower Characteristics

Unweighted Number of Borrowers	Count of loans.
Weighted Number of Borrowers	Weighted count of loans.
Average Annual Income	Weighted average ANN_INC_USE. Includes estimated incomes for those with missing ANN_INC based on middle of given income range information.
Median Annual Income	Weighted median average ANN_INC_USE. Includes estimated incomes for those with missing ANN_INC based on middle of given income range information.
Average Annual Income (Trimmed Top 1%)	Borrowers with highest 1% of ANN_INC_USE excluded. Weighted mean average of ANN_INC_USE on remainin 99% of borrowers.
% Estimated Income Information	Weighted count of loans with missing ANN_INC.

Average FICO	Weighted average EX_FICO.
% With FICO <620	Weighted count of loans where EX_FICO<620 divided by weighted number of borrowers.
% With FICO 620-680	Weighted count of loans where 620<=EX_FICO<680 divided by weighted number of borrowers.
% With FICO =>680	Weighted count of loans where 680<=EX_FICO divided by weighted number of borrowers.
% Missing FICO Information	1 - Sum of Percentage in above FICO ranges.

% White	Weighted count of loans in RACE_WHITE, or where BO_RACE=5 divided by weighted number of borrowers.
% Black	Weighted count of loans in RACE_BLACK, or where BO_RACE=3 divided by weighted number of borrowers.
% Hispanic	Weighted count of loans in RACE_HISP, or where BO_RACE=4 divided by weighted number of borrowers.
% Other	Weighted count of loans in RACE_OTHER, or where BO_RACE=1, 2, 6 divided by weighted number of borrowers.
% Missing Race Information	Weighted count of loans with BO_RACE=1, 7, 8 divided by weighted number of borrowers.

% Female	Weighted count loans where BO_SEX=2, divided by weighted number of borrowers.

Exhibit A.1

Notes on Calculations

% Age 19-34	Weighted count loans in AGE_1934, or where 19<=EX_AGE<35 divided by weighted number of borrowers. Where EX_AGE missing, included borrowers where EX_AGE_CODE=1.
% Age 35-49	Weighted count loans in AGE_3549, or where 36<=EX_AGE<50 divided by weighted number of borrowers. Where EX_AGE missing, included borrowers where EX_AGE_CODE=2.
% Age 50-64	Weighted count loans in AGE_5065, or where 50<=EX_AGE<65 divided by weighted number of borrowers. Where EX_AGE missing, included borrowers where EX_AGE_CODE=3.
% Age >65	Weighted count loans in AGE_65, or where 65<=EX_AGE divided by weighted number of borrowers. Where EX_AGE missing, included borrowers where EX_AGE_CODE=4.
% Missing Age Information	Weighted count of loans with no age or age range information divided by weighted number of borrowers.

Loan Characteristics

Average Loan Amount	Weighted average LN_AMNT.

Average LTV %	Weighted average EX_LTV. EX_LTV calculated as EX_LOAN_AMT_K divided by EX_SALE_AMT_K.
% With LTV<=80	Weighted count of loans where EX_LTV<=80 divided by weighted number of borrowers.
% With LTV 80-90	Weighted count of loans where 80<=EX_LTV<90 divided by weighted number of borrowers.
% With LTV 90-96	Weighted count of loans where 90<=EX_LTV<96 divided by weighted number of borrowers.
% With LTV 96-98	Weighted count of loans where 96<=EX_LTV<98 divided by weighted number of borrowers.
% With LTV=>98	Weighted count of loans where 98<=EX_LTV divided by weighted number of borrowers.
% Missing LTV Information	1 - Sum of Percentage in above LTV ranges.

Average Ratio of Loan Amount to FHA Loan Limit	Weighted average ROUNDEDLOANTOFHA. Calculated as LN_AMNT rounded up to the nearest $1,000, divided by the corresponding FHA loan limit based on EX_UNIT_SIZE and ACTN_DT. See notes below on EX_UNIT_SIZE assignments.
% With LoantoFHA Ratio <=.5	Weighted count of loans where ROUNDEDLOANTOFHA<=.5 divided by weighted number of borrowers.
% With LoantoFHA Ratio .5 - 1	Weighted count of loans where .5<ROUNDEDLOANTOFHA<=1 divided by weighted number of borrowers.
% With LoantoFHA Ratio 1 - 1.2	Weighted count of loans where 1<ROUNDEDLOANTOFHA<=1.2 divided by weighted number of borrowers.
% With LoantoFHA Ratio >1.2	Weighted count of loans where ROUNDEDLOANTOFHA>1.2 divided by weighted number of borrowers.

Average PTI	Weighted average PTI_CALC. Estimated annual payment on a 30-year fixed rate mortgage given EX_LOAN_AMT_K*1000 and the national average contract interest rates on the EX_SALE_DATE, divided by the ANN_INC (see above for notes on income estimation).

% Originated in 1998	Weighted count of loans with ACTN_DT in 1998, divided by weighted number of borrowers.
% Originated in 1999	Weighted count of loans with ACTN_DT in 1999, divided by weighted number of borrowers.
% Originated in 2000	Weighted count of loans with ACTN_DT in 2000, divided by weighted number of borrowers.

Exhibit A.1

Notes on Calculations

Mortgaged Property Characteristics

% New Construction	Weighted count of loans where NEW_CONSTR=1 divided by weighted number of borrowers.

% Unit Size 1	Weighted count of loans where EX_UNIT_SIZE=C, G, H, I, J, K, L, M, N, O, or Missing, divided by weighted number of borrowers.
% Unit Size 2	Weighted count of loans where EX_UNIT_SIZE=D divided by weighted number of borrowers.
% Unit Size 3	Weighted count of loans where EX_UNIT_SIZE=E divided by weighted number of borrowers.
% Unit Size 4	Weighted count of loans where EX_UNIT_SIZE=F divided by weighted number of borrowers.

Borrower Neighborhood Characteristics (1990 Census)

% In Underserved Tracts	Weighted count of loans in SERVED_N=1 divided by weighted number of borrowers.

% In High Cost MSA	Weighted count of loans originating in MSAs with highest FHA single-unit loan limits at time of ACTN_DT.
% In Medium Cost MSA	Weighted count of loan originating in MSAs not designated as high cost or low cost at time of ACTN_DT.
% In Low Cost MSA	Weighted count of loans originating in MSAs with lowest FHA single-unit loan limits at time of ACTN_DT.

% In Center City	Unicon designation of center city by tract based on 1990 Census.

Average 5-yr House Price Appreciation, Lagged 1 year	Average HPIL5, or 5-year house price appreciation, lagged 1 year, of MSA where loan originated in year of origination. Appreciation from OFHEO House Price Index, lagged 1 year)
% in Area with Depreciation	Weighted count of Loans where HPIL5<1, divided by weighted number of borrowers.
% In Area with Appreciation up to 20%	Weighted count of Loans where 1<=HPIL5<1.2, divided by weighted number of borrowers.
% In Area with Appreciation over 20%	Weighted count of Loans where HPIL5>1.2, divided by weighted number of borrowers.

% In <90% Relative Income Tracts	Weighted count of loans where 1990 Census tract median household income divided by 1990 Census MSA median household income RELINC_C90T < .9, divided by weighted number of borrowers.
% In 90-120% Relative Income Tracts	Weighted count of loans where 1990 Census tract median household income divided by 1990 Census MSA median household income .9 =< RELINC_C90T <1.2, divided by weighted number of borrowers.
% In =>120% Relative Income Tracts	Weighted count of loans where 1990 Census tract median household income divided by 1990 Census MSA median household income RELINC_C90T => 1.2 divided by weighted number of borrowers.

% In <10% Minority Tracts	Weighted count of loans where tract level percent minority in 1990 C90T_PMINORTY < 10 divided by weighted number of borrowers.
% In 10-30% Minority Tracts	Weighted count of loans where tract level percent minority in 10 =< C90T_PMINORTY < 30 divided by weighted number of borrowers.
% In =>30% Minority Tracts	Weighted count of loans where tract level percent minority in C90T_PMINORTY => 30 divided by weighted number of borrowers.

Exhibit A.1

Notes on Calculations

Lending Market Characteristics

HHI (by HMDA lender ID, based on $ loaned)	Calculated based on weighted LN_AMNT by unique AGENCY+RESP_ID.
HHI (by HMDA lender ID, based on # loans)	Calculated based on weighted count of loans by unique AGENCY+RESP_ID.

% FHA ($ loaned)	Weighted share of LN_AMNT for loans in FHA_LOAN=1 over total weighted LN_AMNT.
% GSE ($ loaned)	Weighted share of LN_AMNT for loans in PUR_TYPE=1 or PUR_TYPE=3 and FHA_LOAN<>1 over total weighted LN_AMNT.

% FHA (# loans)	Weighted count of loans in FHA_LOAN=1 over total weighted number of borrowers.
% GSE (# loans)	Weighted count of loans in PUR_TYPE=1 or PUR_TYPE=3 and FHA_LOAN<>1 over total weighted number of borrowers.

Model Results

Overlap Range in Origination Model (MSA-level non-parametric tolerance intervals):

FHA Lower Limit (predicted log odds)	Low order statistic such that 90% of FHA loans remain with 95% confidence.
GSE Upper Limit (predicted log odds)	Low order statistic such that 90% of GSE loans remain with 95% confidence.

Overlap Range in Default Model (MSA-level non-parametric tolerance intervals):

FHA Lower Limit (predicted log odds)	Low order statistic such that 90% of FHA loans remain with 95% confidence.
GSE Upper Limit (predicted log odds)	Low order statistic such that 90% of GSE loans remain with 95% confidence.
Average Predicted Likelihood of Default (prediction)	Weighted average predicted likelihood of default based on default model.
Average Predicted Likelihood of Default (log odds)	Weighted log odds of predicted likelihood of default based on default model.

Origination and Default Model Market Sectors

Tolerance Interval Method

GSE	Loans in PUR_TYPE=1 or PUR_TYPE=3 and FHA_LOAN<>1 that are between GSE lower bound and FHA lower bound.
GSE in Overlap Range	Loans in PUR_TYPE=1 or PUR_TYPE=3 and FHA_LOAN<>1 that are between FHA lower bound and GSE upper bound.
Combined Overlap	Loans between FHA lower bound and GSE upper bound.
FHA in Overlap Range	Loans in FHA_LOAN=1 that are between FHA lower bound and GSE upper bound.
FHA	Loans in FHA_LOAN=1 that are between GSE upper bound and FHA upper bound.
Tail	Loans in PUR_TYPE=1 or PUR_TYPE=3 and FHA_LOAN<>1 that are below GSE lower bound or above GSE upper bound, and loans in FHA_LOAN=1 that are below FHA lower bound and above FHA upper bound.

Exhibit A.1

Notes on Calculations

Confidence Interval Method

FHA (non-overlap)	Loans in which the confidence interval (+/- 1.96*standard error) around the predicted probability of being FHA includes the value of 1.
GSE (non-overlap)	Loans in which the confidence interval (+/- 1.96*standard error) around the predicted probability of being FHA includes the value of 0.
Overlap	Loans in which the confidence interval (+/- 1.96*standard error) around the predicted probability of being FHA includes neither 1 or 0.

Market Sectors

All GSE	Loans in PUR_TYPE=1 or PUR_TYPE=3.
GSE With PMI	Loans in PUR_TYPE=1 or PUR_TYPE=3 and PMI_FLAG="Y".
GSE No PMI	Loans in PUR_TYPE=1 or PUR_TYPE=3 and PMI_FLAG="N".
All Depository	Loans in PUR_TYPE=0, 5, 6 or 8.
Depository With PMI	Loans in PUR_TYPE=0, 5, 6 or 8 and PMI_FLAG="Y".
Depository No PMI	Loans in PUR_TYPE=0, 5, 6 or 8 and PMI_FLAG="N".
All Other Investors	Loans in PUR_TYPE=7 or PUR_TYPE=9.
Other Investors No PMI	Loans in PUR_TYPE=7 or PUR_TYPE=9 and PMI_FLAG="Y".
Other Investors With PMI	Loans in PUR_TYPE=7 or PUR_TYPE=9 and PMI_FLAG="N".
All with PMI	PMI data downloaded from ftp website hud_pdr. Merged on HMDA_SEQ. Loans with PMI_FLAG="Y".
All FHA	Loans in FHA_LOAN=1.
All Subprime	Subprime data downloaded from http://www.huduser.org/datasets/manu.html. Merged on compressed AGENCY and RESP_ID. SUBPRIME=1 if lender classified as a primarily subprime lender in year of loan origination. Loans in SUBPRIME=1.

Exhibit A.2: Analysis of Matched and Unmatched Experian and HMDA Loans in Baltimore (1998-2000)

Charateristics	Experian Unmatched	Experian/HMDA Matched	HMDA Unmatched	All HMDA
Borrower Characteristics				
Unweighted Number of Borrowers	9,150	11,612	102,891	114,503
Weighted Number of Borrowers		121,587		
Average Annual Income	$63,569	$59,803	$66,765	$65,974
Median Annual Income	$62,500	$51,000	$57,000	$56,000
Average Annual Income (Trimmed Top 1%)	$63,139	$56,955	$63,958	$63,252
% Estimated Income Information	0%	2%	2%	2%
Average FICO	677	688		
% With FICO < 620	21%	24%		
% With FICO 620 - 679	13%	16%		
% With FICO => 680	41%	60%		
% Missing FICO Information	25%	0%		
% White		72%	70%	70%
% Black		22%	17%	17%
% Hispanic		1%	1%	1%
% Other		3%	4%	4%
% Missing Race Information		3%	8%	8%
% Female	24%	35%	29%	30%
% Male	69%	65%	65%	65%
% Missing Gender Information	6%	0%	6%	5%
% Age 19-34	22%	33%		
% Age 35-49	39%	49%		
% Age 50-64	11%	15%		
% Age >65	2%	3%		
% Missing Age Information	25%	0%		
Loan Characteristics				
Average Loan Amount	$117,149	$117,870	$112,185	$112,616
Average LTV %	75%	88%		
% With LTV <= 90	29%	39%		
% With LTV 90 - 96	9%	13%		
% With LTV 97 - 98	6%	9%		
% With LTV > 98	12%	39%		
% Missing LTV Information	45%	0%		
Average Ratio of Loan Amount to FHA Loan Limit	67%	67%		
% With LoantoFHA Ratio <=.5	33%	31%		
% With LoantoFHA Ratio of .6 - 1	52%	55%		
% With LoantoFHA Ratio of 1.1 - 1.2	10%	8%		
% With LoantoFHA Ratio > 1.2	5%	5%		
Average PTI	18%	19%	17%	17%
% Originated in 1998	19%	32%	33%	32%
% Originated in 1999	39%	34%	33%	34%
% Originated in 2000	42%	34%	34%	34%

Exhibit A.2: Analysis of Matched and Unmatched Experian and HMDA Loans in Baltimore (1998-2000)

Charateristics	Experian Unmatched	Experian/HMDA Matched	HMDA Unmatched	All HMDA
Mortgaged Property Characteristics				
% Old Construction	73%	86%		
% New Construction	20%	14%		
% Missing Construction Information	8%	0%		
% Unit Size 1	97%	96%		
% Unit Size 2	1%	2%		
% Unit Size 3	2%	1%		
% Unit Size 4	0%	0%		
Borrower Neighborhood Characteristics (1990 Census)				
% In Underserved Tracts		25%	24%	25%
% Not in Underserved Tracts		75%	73%	73%
% Missing Underserved Tracts		0%	2%	2%
% in High Cost Cities	0%	0%	0%	0%
% in Average Cost Cities	100%	100%	100%	100%
% in Low Cost Cities	0%	0%	0%	0%
% In Center City	24%	22%	19%	19%
% Not in Center City	76%	78%	79%	79%
% Missing Center City Information	0%	0%	2%	2%
Average 5-year Depreciation Lagged 1 year	108%	108%	108%	108%
% In Area with Depreciation	0%	0%	0%	0%
% In Area with Appreciation up to 20%	100%	100%	100%	100%
% In Area with Appreciation over 20%	0%	0%	0%	0%
% In <90% Relative Income Tracts	25%	19%	21%	21%
% In 90-120% Relative Income Tracts	43%	39%	27%	29%
% In >120% Relative Income Tracts	32%	42%	51%	50%
% In <10% Minority Tracts	49%	56%	50%	50%
% In 10-30% Minority Tracts	27%	26%	33%	32%
% In >30% Minority Tracts	24%	18%	15%	16%
% Missing Minority Tract Information	0%	0%	2%	2%

Note: Jumbo loans are excluded from this analysis. The column for the Experian/HMDA Matched loans presents weighted numbers but the columns for the Experian Unmatched, HMDA Unmatched and All HMDA loans present unweighted numbers.

Exhibit A.3: Analysis of Matched and Unmatched Experian and HMDA Loans in Chicago (1998-2000)

Charateristics	Experian Unmatched	Experian/HMDA Matched	HMDA Unmatched	All HMDA
Borrower Characteristics				
Unweighted Number of Borrowers	31,926	47,408	330,211	377,619
Weighted Number of Borrowers		389,963		
Average Annual Income	$89,377	$65,836	$68,894	$68,533
Median Annual Income	$87,500	$57,000	$59,000	$59,000
Average Annual Income (Trimmed Top 1%)	$89,377	$63,232	$65,630	$65,324
% Estimated Income Information	0%	2%	2%	2%
Average FICO	698	701		
% With FICO < 620	16%	18%		
% With FICO 620 - 679	14%	17%		
% With FICO => 680	53%	66%		
% Missing FICO Information	18%	0%		
% White		68%	65%	65%
% Black		10%	9%	9%
% Hispanic		14%	12%	12%
% Other		5%	7%	7%
% Missing Race Information		2%	7%	7%
% Female	19%	25%	24%	24%
% Male	74%	75%	71%	71%
% Missing Gender Information	7%	0%	5%	5%
% Age 19-34	27%	34%		
% Age 35-49	41%	49%		
% Age 50-64	11%	13%		
% Age >65	2%	3%		
% Missing Age Information	19%	0%		
Loan Characteristics				
Average Loan Amount	$136,438	$128,925	$125,600	$126,084
Average LTV %	63%	81%		
% With LTV <= 90	47%	62%		
% With LTV 90 - 96	6%	18%		
% With LTV 97 - 98	2%	6%		
% With LTV > 98	3%	14%		
% Missing LTV Information	42%	0%		
Average Ratio of Loan Amount to FHA Loan Limit	69%	67%		
% With LoantoFHA Ratio <=.5	27%	29%		
% With LoantoFHA Ratio of .6 - 1	59%	59%		
% With LoantoFHA Ratio of 1.1 - 1.2	11%	9%		
% With LoantoFHA Ratio > 1.2	4%	3%		
Average PTI	16%	19%	18%	18%
% Originated in 1998	25%	31%	31%	31%
% Originated in 1999	35%	34%	34%	34%
% Originated in 2000	40%	35%	35%	35%

Exhibit A.3: Analysis of Matched and Unmatched Experian and HMDA Loans in Chicago (1998-2000)

Charateristics	Experian Unmatched	Experian/HMDA Matched	HMDA Unmatched	All HMDA
Mortgaged Property Characteristics				
% Old Construction	85%	91%		
% New Construction	15%	9%		
% Missing Construction Information	0%	0%		
% Unit Size 1	92%	92%		
% Unit Size 2	2%	2%		
% Unit Size 3	4%	4%		
% Unit Size 4	2%	2%		
Borrower Neighborhood Characteristics (1990 Census)				
% In Underserved Tracts		26%	26%	26%
% Not in Underserved Tracts		74%	74%	74%
% Missing Underserved Tracts		0%	0%	0%
% in High Cost Cities	3%	2%	3%	2%
% in Average Cost Cities	97%	98%	97%	98%
% in Low Cost Cities	0%	0%	0%	0%
% In Center City	27%	30%	35%	34%
% Not in Center City	73%	70%	65%	66%
% Missing Center City Information	0%	0%	0%	0%
Average 5-year Depreciation Lagged 1 year	118%	118%	118%	118%
% In Area with Depreciation	0%	0%	0%	0%
% In Area with Appreciation up to 20%	100%	100%	100%	100%
% In Area with Appreciation over 20%	0%	0%	0%	0%
% In <90% Relative Income Tracts	17%	18%	20%	20%
% In 90-120% Relative Income Tracts	28%	31%	31%	31%
% In >120% Relative Income Tracts	55%	51%	49%	49%
% In <10% Minority Tracts	55%	49%	48%	48%
% In 10-30% Minority Tracts	27%	32%	32%	32%
% In >30% Minority Tracts	18%	19%	19%	19%
% Missing Minority Tract Information	0%	0%	0%	0%

Note: Jumbo loans are excluded from this analysis. The column for the Experian/HMDA Matched loans presents weighted numbers but the columns for the Experian Unmatched, HMDA Unmatched and All HMDA loans present unweighted numbers.

ExhibitA.4: Analysis of Matched and Unmatched Experian and HMDA Loans in Cleveland (1998-2000)

Charateristics	Experian Unmatched	Experian/HMDA Matched	HMDA Unmatched	All HMDA
Borrower Characteristics				
Unweighted Number of Borrowers	17,230	24,795	67,932	92,727
Weighted Number of Borrowers		96,935		
Average Annual Income	$57,642	$57,318	$60,381	$59,599
Median Annual Income	$42,500	$49,000	$51,000	$50,000
Average Annual Income (Trimmed Top 1%)	$55,987	$55,230	$57,315	$56,756
% Estimated Income Information	0%	2%	2%	2%
Average FICO	679	700		
% With FICO < 620	15%	20%		
% With FICO 620 - 679	9%	14%		
% With FICO => 680	32%	66%		
% Missing FICO Information	44%	0%		
% White		86%	74%	77%
% Black		7%	10%	10%
% Hispanic		2%	2%	2%
% Other		2%	2%	2%
% Missing Race Information		3%	11%	9%
% Female	18%	22%	24%	24%
% Male	77%	78%	67%	69%
% Missing Gender Information	5%	0%	9%	7%
% Age 19-34	17%	33%		
% Age 35-49	29%	49%		
% Age 50-64	8%	15%		
% Age >65	2%	4%		
% Missing Age Information	44%	0%		
Loan Characteristics				
Average Loan Amount	$102,082	$106,126	$102,435	$103,510
Average LTV %	77%	83%		
% With LTV <= 90	61%	60%		
% With LTV 90 - 96	12%	17%		
% With LTV 97 - 98	4%	4%		
% With LTV > 98	9%	19%		
% Missing LTV Information	14%	0%		
Average Ratio of Loan Amount to FHA Loan Limit	58%	67%		
% With LoantoFHA Ratio <=.5	48%	34%		
% With LoantoFHA Ratio of .6 - 1	42%	52%		
% With LoantoFHA Ratio of 1.1 - 1.2	6%	7%		
% With LoantoFHA Ratio > 1.2	5%	7%		
Average PTI	19%	18%	16%	17%
% Originated in 1998	13%	34%	40%	33%
% Originated in 1999	33%	34%	34%	34%
% Originated in 2000	54%	32%	26%	32%

ExhibitA.4: Analysis of Matched and Unmatched Experian and HMDA Loans in Cleveland (1998-2000)

Charateristics	Experian Unmatched	Experian/HMDA Matched	HMDA Unmatched	All HMDA
Mortgaged Property Characteristics				
% Old Construction	41%	96%		
% New Construction	7%	4%		
% Missing Construction Information	52%	0%		
% Unit Size 1	96%	98%		
% Unit Size 2	1%	0%		
% Unit Size 3	4%	1%		
% Unit Size 4	0%	0%		
Borrower Neighborhood Characteristics (1990 Census)				
% In Underserved Tracts		25%	25%	25%
% Not in Underserved Tracts		75%	75%	75%
% Missing Underserved Tracts		0%	0%	0%
% in High Cost Cities	58%	35%	28%	35%
% in Average Cost Cities	42%	65%	72%	65%
% in Low Cost Cities	0%	0%	0%	0%
% In Center City	27%	19%	20%	20%
% Not in Center City	73%	81%	80%	80%
% Missing Center City Information	0%	0%	0%	0%
Average 5-year Depreciation Lagged 1 year	124%	124%	124%	124%
% In Area with Depreciation	0%	0%	0%	0%
% In Area with Appreciation up to 20%	0%	0%	0%	0%
% In Area with Appreciation over 20%	100%	100%	100%	100%
% In <90% Relative Income Tracts	32%	21%	22%	22%
% In 90-120% Relative Income Tracts	31%	34%	34%	34%
% In >120% Relative Income Tracts	37%	44%	44%	44%
% In <10% Minority Tracts	72%	81%	79%	79%
% In 10-30% Minority Tracts	13%	12%	12%	12%
% In >30% Minority Tracts	15%	7%	10%	9%
% Missing Minority Tract Information	0%	0%	0%	0%

Note: Jumbo loans are excluded from this analysis. The column for the Experian/HMDA Matched loans presents weighted numbers but the columns for the Experian Unmatched, HMDA Unmatched and All HMDA loans present unweighted numbers.

Exhibit A.5: Analysis of Matched and Unmatched Experian and HMDA Loans in Denver (1998-2000)

Charateristics	Experian Unmatched	Experian/HMDA Matched	HMDA Unmatched	All HMDA
Borrower Characteristics				
Unweighted Number of Borrowers	28,644	45,969	116,476	162,445
Weighted Number of Borrowers		169,772		
Average Annual Income	$80,714	$61,883	$66,192	$64,594
Median Annual Income	$62,500	$54,000	$57,000	$56,000
Average Annual Income (Trimmed Top 1%)	$79,030	$58,395	$63,482	$61,981
% Estimated Income Information	0%	2%	3%	3%
Average FICO	698	702		
% With FICO < 620	15%	16%		
% With FICO 620 - 679	17%	19%		
% With FICO => 680	57%	65%		
% Missing FICO Information	11%	0%		
% White		76%	71%	72%
% Black		3%	3%	3%
% Hispanic		14%	10%	12%
% Other		4%	5%	5%
% Missing Race Information		3%	11%	8%
% Female	20%	28%	27%	27%
% Male	74%	72%	66%	67%
% Missing Gender Information	6%	0%	8%	6%
% Age 19-34	30%	36%		
% Age 35-49	44%	47%		
% Age 50-64	12%	14%		
% Age >65	2%	3%		
% Missing Age Information	12%	1%		
Loan Characteristics				
Average Loan Amount	$147,139	$138,614	$129,547	$131,848
Average LTV %	60%	85%		
% With LTV <= 90	30%	49%		
% With LTV 90 - 96	2%	16%		
% With LTV 97 - 98	1%	10%		
% With LTV > 98	3%	25%		
% Missing LTV Information	64%	0%		
Average Ratio of Loan Amount to FHA Loan Limit	78%	75%		
% With LoantoFHA Ratio <=.5	13%	16%		
% With LoantoFHA Ratio of .6 - 1	68%	68%		
% With LoantoFHA Ratio of 1.1 - 1.2	14%	12%		
% With LoantoFHA Ratio > 1.2	6%	5%		
Average PTI	20%	22%	20%	20%
% Originated in 1998	24%	32%	34%	31%
% Originated in 1999	35%	34%	34%	34%
% Originated in 2000	41%	34%	33%	35%

Exhibit A.5: Analysis of Matched and Unmatched Experian and HMDA Loans in Denver (1998-2000)

Charateristics	Experian Unmatched	Experian/HMDA Matched	HMDA Unmatched	All HMDA
Mortgaged Property Characteristics				
% Old Construction	70%	81%		
% New Construction	30%	19%		
% Missing Construction Information	0%	0%		
% Unit Size 1	97%	96%		
% Unit Size 2	1%	2%		
% Unit Size 3	1%	1%		
% Unit Size 4	1%	1%		
Borrower Neighborhood Characteristics (1990 Census)				
% In Underserved Tracts		31%	27%	31%
% Not in Underserved Tracts		69%	72%	69%
% Missing Underserved Tracts		0%	0%	0%
% in High Cost Cities	5%	7%	7%	7%
% in Average Cost Cities	95%	93%	93%	93%
% in Low Cost Cities	0%	0%	0%	0%
% In Center City	20%	23%	21%	22%
% Not in Center City	80%	77%	79%	78%
% Missing Center City Information	0%	0%	0%	0%
Average 5-year Depreciation Lagged 1 year	119%	119%	119%	119%
% In Area with Depreciation	0%	0%	0%	0%
% In Area with Appreciation up to 20%	100%	100%	100%	100%
% In Area with Appreciation over 20%	0%	0%	0%	0%
% In <90% Relative Income Tracts	28%	25%	24%	26%
% In 90-120% Relative Income Tracts	20%	23%	22%	23%
% In >120% Relative Income Tracts	52%	52%	54%	51%
% In <10% Minority Tracts	44%	41%	41%	40%
% In 10-30% Minority Tracts	35%	42%	44%	44%
% In >30% Minority Tracts	20%	17%	14%	16%
% Missing Minority Tract Information	1%	0%	0%	0%

Note: Jumbo loans are excluded from this analysis. The column for the Experian/HMDA Matched loans presents weighted numbers but the columns for the Experian Unmatched, HMDA Unmatched and All HMDA loans present unweighted numbers.

Exhibit A.6: Analysis of Matched and Unmatched Experian and HMDA Loans in Los Angeles (1998-2000)

Charateristics	Experian Unmatched	Experian/HMDA Matched	HMDA Unmatched	All HMDA
Borrower Characteristics				
Unweighted Number of Borrowers	12,734	20,841	246,246	267,087
Weighted Number of Borrowers		268,173		
Average Annual Income	$78,768	$65,925	$72,366	$71,763
Median Annual Income	$62,500	$58,000	$60,000	$60,000
Average Annual Income (Trimmed Top 1%)	$78,768	$61,965	$68,470	$67,910
% Estimated Income Information	0%	3%	3%	3%
Average FICO	674	684		
% With FICO < 620	20%	22%		
% With FICO 620 - 679	18%	21%		
% With FICO => 680	42%	57%		
% Missing FICO Information	20%	0%		
% White		40%	41%	41%
% Black		6%	7%	7%
% Hispanic		34%	29%	30%
% Other		16%	14%	14%
% Missing Race Information		4%	9%	8%
% Female	24%	26%	25%	25%
% Male	67%	74%	70%	70%
% Missing Gender Information	9%	0%	5%	4%
% Age 19-34	21%	28%		
% Age 35-49	46%	56%		
% Age 50-64	11%	14%		
% Age >65	3%	3%		
% Missing Age Information	20%	0%		
Loan Characteristics				
Average Loan Amount	$153,444	$150,579	$131,779	$133,164
Average LTV %	64%	85%		
% With LTV <= 90	38%	52%		
% With LTV 90 - 96	6%	15%		
% With LTV 97 - 98	2%	11%		
% With LTV > 98	5%	22%		
% Missing LTV Information	49%	0%		
Average Ratio of Loan Amount to FHA Loan Limit	74%	74%		
% With LoantoFHA Ratio <=.5	17%	17%		
% With LoantoFHA Ratio of .6 - 1	67%	66%		
% With LoantoFHA Ratio of 1.1 - 1.2	14%	14%		
% With LoantoFHA Ratio > 1.2	2%	2%		
Average PTI	23%	23%	19%	20%
% Originated in 1998	25%	31%	31%	31%
% Originated in 1999	34%	34%	34%	34%
% Originated in 2000	40%	34%	34%	35%

Exhibit A.6: Analysis of Matched and Unmatched Experian and HMDA Loans in Los Angeles (1998-2000)

Charateristics	Experian Unmatched	Experian/HMDA Matched	HMDA Unmatched	All HMDA
Mortgaged Property Characteristics				
% Old Construction	95%	97%		
% New Construction	5%	3%		
% Missing Construction Information	0%	0%		
% Unit Size 1	94%	95%		
% Unit Size 2	2%	1%		
% Unit Size 3	2%	2%		
% Unit Size 4	2%	2%		
Borrower Neighborhood Characteristics (1990 Census)				
% In Underserved Tracts		50%	49%	49%
% Not in Underserved Tracts		50%	51%	51%
% Missing Underserved Tracts		0%	0%	0%
% in High Cost Cities	40%	41%	41%	41%
% in Average Cost Cities	60%	59%	59%	59%
% in Low Cost Cities	0%	0%	0%	0%
% In Center City	42%	42%	46%	46%
% Not in Center City	58%	58%	54%	54%
% Missing Center City Information	0%	0%	0%	0%
Average 5-year Depreciation Lagged 1 year	101%	99%	99%	99%
% In Area with Depreciation	60%	66%	66%	65%
% In Area with Appreciation up to 20%	40%	34%	34%	35%
% In Area with Appreciation over 20%	0%	0%	0%	0%
% In <90% Relative Income Tracts	34%	25%	24%	25%
% In 90-120% Relative Income Tracts	31%	35%	32%	32%
% In >120% Relative Income Tracts	34%	40%	44%	43%
% In <10% Minority Tracts	1%	1%	1%	1%
% In 10-30% Minority Tracts	25%	31%	35%	35%
% In >30% Minority Tracts	74%	68%	64%	64%
% Missing Minority Tract Information	0%	0%	0%	0%

Note: Jumbo loans are excluded from this analysis. The column for the Experian/HMDA Matched loans presents weighted numbers but the columns for the Experian Unmatched, HMDA Unmatched and All HMDA loans present unweighted numbers.

Exhibit A.7: Analysis of Matched and Unmatched Experian and HMDA Loans in Oakland (1998-2000)

Charateristics	Experian Unmatched	Experian/HMDA Matched	HMDA Unmatched	All HMDA
Borrower Characteristics				
Unweighted Number of Borrowers	7,882	13,683	85,946	99,629
Weighted Number of Borrowers		96,068		
Average Annual Income	$89,663	$70,587	$81,032	$79,339
Median Annual Income	$87,500	$63,000	$71,000	$70,000
Average Annual Income (Trimmed Top 1%)	$89,663	$66,811	$77,927	$76,349
% Estimated Income Information	0%	2%	3%	3%
Average FICO	698	708		
% With FICO < 620	13%	14%		
% With FICO 620 - 679	15%	18%		
% With FICO => 680	50%	69%		
% Missing FICO Information	22%	0%		
% White		52%	51%	51%
% Black		7%	6%	7%
% Hispanic		18%	12%	13%
% Other		19%	19%	19%
% Missing Race Information		4%	11%	10%
% Female	22%	27%	25%	25%
% Male	68%	73%	69%	70%
% Missing Gender Information	10%	0%	6%	5%
% Age 19-34	20%	27%		
% Age 35-49	44%	54%		
% Age 50-64	12%	16%		
% Age >65	3%	3%		
% Missing Age Information	22%	0%		
Loan Characteristics				
Average Loan Amount	$167,078	$167,811	$131,659	$136,255
Average LTV %	62%	81%		
% With LTV <= 90	55%	68%		
% With LTV 90 - 96	2%	14%		
% With LTV 97 - 98	2%	7%		
% With LTV > 98	3%	11%		
% Missing LTV Information	38%	0%		
Average Ratio of Loan Amount to FHA Loan Limit	80%	83%		
% With LoantoFHA Ratio <=.5	13%	11%		
% With LoantoFHA Ratio of .6 - 1	62%	61%		
% With LoantoFHA Ratio of 1.1 - 1.2	22%	23%		
% With LoantoFHA Ratio > 1.2	3%	4%		
Average PTI	20%	23%	17%	18%
% Originated in 1998	26%	33%	32%	31%
% Originated in 1999	37%	36%	35%	36%
% Originated in 2000	37%	31%	33%	33%

Exhibit A.7: Analysis of Matched and Unmatched Experian and HMDA Loans in Oakland (1998-2000)

Charateristics	Experian Unmatched	Experian/HMDA Matched	HMDA Unmatched	All HMDA
Mortgaged Property Characteristics				
% Old Construction	91%	95%		
% New Construction	9%	5%		
% Missing Construction Information	0%	0%		
% Unit Size 1	95%	96%		
% Unit Size 2	1%	1%		
% Unit Size 3	2%	1%		
% Unit Size 4	2%	1%		
Borrower Neighborhood Characteristics (1990 Census)				
% In Underserved Tracts		52%	48%	50%
% Not in Underserved Tracts		48%	52%	50%
% Missing Underserved Tracts		0%	0%	0%
% in High Cost Cities	80%	74%	75%	76%
% in Average Cost Cities	20%	26%	25%	24%
% in Low Cost Cities	0%	0%	0%	0%
% In Center City	18%	15%	17%	17%
% Not in Center City	82%	85%	83%	83%
% Missing Center City Information	0%	0%	0%	0%
Average 5-year Depreciation Lagged 1 year	112%	110%	111%	111%
% In Area with Depreciation	26%	33%	32%	31%
% In Area with Appreciation up to 20%	37%	36%	35%	36%
% In Area with Appreciation over 20%	37%	31%	33%	33%
% In <90% Relative Income Tracts	42%	31%	28%	30%
% In 90-120% Relative Income Tracts	27%	29%	34%	33%
% In >120% Relative Income Tracts	30%	40%	38%	37%
% In <10% Minority Tracts	2%	2%	3%	2%
% In 10-30% Minority Tracts	33%	43%	49%	48%
% In >30% Minority Tracts	65%	55%	48%	50%
% Missing Minority Tract Information	0%	0%	0%	0%

Note: Jumbo loans are excluded from this analysis. The column for the Experian/HMDA Matched loans presents weighted numbers but the columns for the Experian Unmatched, HMDA Unmatched and All HMDA loans present unweighted numbers.

Exhibit A.8: Analysis of Matched and Unmatched Experian and HMDA Loans in Philadelphia (1998-2000)

Charateristics	Experian Unmatched	Experian/HMDA Matched	HMDA Unmatched	All HMDA
Borrower Characteristics				
Unweighted Number of Borrowers	23,183	30,296	160,392	190,688
Weighted Number of Borrowers		196,784		
Average Annual Income	$57,515	$60,959	$67,379	$65,588
Median Annual Income	$42,500	$52,000	$58,000	$56,000
Average Annual Income (Trimmed Top 1%)	$56,621	$54,192	$64,308	$62,623
% Estimated Income Information	0%	2%	3%	3%
Average FICO	681	692		
% With FICO < 620	20%	23%		
% With FICO 620 - 679	11%	15%		
% With FICO => 680	42%	62%		
% Missing FICO Information	26%	0%		
% White		75%	72%	71%
% Black		14%	9%	11%
% Hispanic		4%	3%	3%
% Other		4%	4%	4%
% Missing Race Information		4%	12%	10%
% Female	24%	30%	26%	27%
% Male	69%	70%	66%	66%
% Missing Gender Information	7%	0%	8%	7%
% Age 19-34	22%	31%		
% Age 35-49	38%	51%		
% Age 50-64	11%	15%		
% Age >65	2%	3%		
% Missing Age Information	27%	0%		
Loan Characteristics				
Average Loan Amount	$106,718	$108,463	$107,692	$106,432
Average LTV %	75%	85%		
% With LTV <= 90	39%	52%		
% With LTV 90 - 96	10%	19%		
% With LTV 97 - 98	3%	6%		
% With LTV > 98	9%	23%		
% Missing LTV Information	37%	0%		
Average Ratio of Loan Amount to FHA Loan Limit	69%	71%		
% With LoantoFHA Ratio <=.5	37%	32%		
% With LoantoFHA Ratio of .6 - 1	43%	49%		
% With LoantoFHA Ratio of 1.1 - 1.2	9%	9%		
% With LoantoFHA Ratio > 1.2	12%	10%		
Average PTI	19%	17%	16%	16%
% Originated in 1998	25%	32%	32%	32%
% Originated in 1999	36%	34%	34%	34%
% Originated in 2000	40%	34%	34%	35%

Exhibit A.8: Analysis of Matched and Unmatched Experian and HMDA Loans in Philadelphia (1998-2000)

Charateristics	Experian Unmatched	Experian/HMDA Matched	HMDA Unmatched	All HMDA
Mortgaged Property Characteristics				
% Old Construction	86%	92%		
% New Construction	14%	8%		
% Missing Construction Information	0%	0%		
% Unit Size 1	95%	95%		
% Unit Size 2	2%	2%		
% Unit Size 3	3%	3%		
% Unit Size 4	0%	0%		
Borrower Neighborhood Characteristics (1990 Census)				
% In Underserved Tracts		24%	21%	24%
% Not in Underserved Tracts		76%	79%	76%
% Missing Underserved Tracts		0%	0%	0%
% in High Cost Cities	0%	0%	0%	0%
% in Average Cost Cities	100%	100%	100%	100%
% in Low Cost Cities	0%	0%	0%	0%
% In Center City	30%	26%	18%	21%
% Not in Center City	70%	74%	82%	79%
% Missing Center City Information	0%	0%	0%	0%
Average 5-year Depreciation Lagged 1 year	105%	105%	105%	105%
% In Area with Depreciation	0%	0%	0%	0%
% In Area with Appreciation up to 20%	100%	100%	100%	100%
% In Area with Appreciation over 20%	0%	0%	0%	0%
% In <90% Relative Income Tracts	33%	23%	19%	21%
% In 90-120% Relative Income Tracts	32%	37%	33%	33%
% In >120% Relative Income Tracts	35%	40%	49%	46%
% In <10% Minority Tracts	62%	69%	69%	69%
% In 10-30% Minority Tracts	23%	21%	21%	21%
% In >30% Minority Tracts	15%	10%	10%	11%
% Missing Minority Tract Information	0%	0%	0%	0%

Note: Jumbo loans are excluded from this analysis. The column for the Experian/HMDA Matched loans presents weighted numbers but the columns for the Experian Unmatched, HMDA Unmatched and All HMDA loans present unweighted numbers.

Exhibit A.9: Analysis of Matched and Unmatched Experian and HMDA Loans in Portland (1998-2000)

Charateristics	Experian Unmatched	Experian/HMDA Matched	HMDA Unmatched	All HMDA
Borrower Characteristics				
Unweighted Number of Borrowers	19,254	37,023	62,545	99,568
Weighted Number of Borrowers		102,918		
Average Annual Income	$88,321	$61,797	$65,106	$63,865
Median Annual Income	$87,500	$54,000	$56,000	$55,000
Average Annual Income (Trimmed Top 1%)	$88,321	$59,246	$62,239	$61,100
% Estimated Income Information	0%	2%	2%	2%
Average FICO	705	712		
% With FICO < 620	13%	13%		
% With FICO 620 - 679	17%	16%		
% With FICO => 680	60%	70%		
% Missing FICO Information	10%	0%		
% White		86%	78%	81%
% Black		1%	1%	1%
% Hispanic		4%	3%	4%
% Other		6%	6%	6%
% Missing Race Information		3%	12%	8%
% Female	17%	24%	24%	24%
% Male	77%	76%	68%	71%
% Missing Gender Information	7%	0%	8%	5%
% Age 19-34	24%	35%		
% Age 35-49	38%	47%		
% Age 50-64	10%	14%		
% Age >65	2%	3%		
% Missing Age Information	26%	0%		
Loan Characteristics				
Average Loan Amount	$142,355	$137,479	$122,380	$128,001
Average LTV %	59%	84%		
% With LTV <= 90	33%	59%		
% With LTV 90 - 96	7%	16%		
% With LTV 97 - 98	3%	8%		
% With LTV > 98	2%	17%		
% Missing LTV Information	54%	0%		
Average Ratio of Loan Amount to FHA Loan Limit	84%	81%		
% With LoantoFHA Ratio <=.5	7%	9%		
% With LoantoFHA Ratio of .6 - 1	70%	71%		
% With LoantoFHA Ratio of 1.1 - 1.2	15%	12%		
% With LoantoFHA Ratio > 1.2	8%	8%		
Average PTI	18%	22%	18%	20%
% Originated in 1998	25%	35%	38%	34%
% Originated in 1999	36%	33%	32%	33%
% Originated in 2000	39%	32%	30%	33%

Exhibit A.9: Analysis of Matched and Unmatched Experian and HMDA Loans in Portland (1998-2000)

Charateristics	Experian Unmatched	Experian/HMDA Matched	HMDA Unmatched	All HMDA
Mortgaged Property Characteristics				
% Old Construction	73%	84%		
% New Construction	27%	16%		
% Missing Construction Information	0%	0%		
% Unit Size 1	98%	97%		
% Unit Size 2	1%	2%		
% Unit Size 3	1%	1%		
% Unit Size 4	0%	0%		
Borrower Neighborhood Characteristics (1990 Census)				
% In Underserved Tracts		27%	25%	26%
% Not in Underserved Tracts		73%	72%	72%
% Missing Underserved Tracts		0%	3%	2%
% in High Cost Cities	0%	0%	0%	0%
% in Average Cost Cities	100%	100%	100%	100%
% in Low Cost Cities	0%	0%	0%	0%
% In Center City	28%	36%	31%	34%
% Not in Center City	72%	64%	65%	64%
% Missing Center City Information	0%	0%	3%	2%
Average 5-year Depreciation Lagged 1 year	143%	144%	145%	144%
% In Area with Depreciation	0%	0%	0%	0%
% In Area with Appreciation up to 20%	0%	0%	0%	0%
% In Area with Appreciation over 20%	100%	100%	100%	100%
% In <90% Relative Income Tracts	19%	25%	25%	25%
% In 90-120% Relative Income Tracts	42%	41%	42%	41%
% In >120% Relative Income Tracts	38%	34%	33%	33%
% In <10% Minority Tracts	76%	71%	71%	71%
% In 10-30% Minority Tracts	21%	25%	23%	24%
% In >30% Minority Tracts	3%	4%	3%	3%
% Missing Minority Tract Information	0%	0%	3%	2%

Note: Jumbo loans are excluded from this analysis. The column for the Experian/HMDA Matched loans presents weighted numbers but the columns for the Experian Unmatched, HMDA Unmatched and All HMDA loans present unweighted numbers.

Exhibit A.10: Analysis of Matched and Unmatched Experian and HMDA Loans in St Louis (1998-2000)

Charateristics	Experian Unmatched	Experian/HMDA Matched	HMDA Unmatched	All HMDA
Borrower Characteristics				
Unweighted Number of Borrowers	29,556	38,409	86,160	124,569
Weighted Number of Borrowers		125,054		
Average Annual Income	$59,411	$56,011	$56,338	$56,463
Median Annual Income	$62,500	$48,000	$48,000	$48,000
Average Annual Income (Trimmed Top 1%)	$58,769	$54,472	$53,766	$54,019
% Estimated Income Information	0%	1%	2%	2%
Average FICO	689	691		
% With FICO < 620	15%	23%		
% With FICO 620 - 679	10%	14%		
% With FICO => 680	39%	63%		
% Missing FICO Information	36%	0%		
% White		80%	79%	79%
% Black		14%	8%	10%
% Hispanic		1%	1%	1%
% Other		3%	2%	2%
% Missing Race Information		2%	9%	7%
% Female	17%	29%	25%	26%
% Male	76%	71%	68%	69%
% Missing Gender Information	7%	0%	7%	5%
% Age 19-34	22%	34%		
% Age 35-49	31%	48%		
% Age 50-64	9%	15%		
% Age >65	2%	4%		
% Missing Age Information	36%	0%		
Loan Characteristics				
Average Loan Amount	$99,852	$96,681	$91,218	$93,386
Average LTV %	70%	81%		
% With LTV <= 90	68%	78%		
% With LTV 90 - 96	3%	4%		
% With LTV 97 - 98	5%	2%		
% With LTV > 98	4%	16%		
% Missing LTV Information	19%	0%		
Average Ratio of Loan Amount to FHA Loan Limit	72%	70%		
% With LoantoFHA Ratio <=.5	33%	35%		
% With LoantoFHA Ratio of .6 - 1	47%	46%		
% With LoantoFHA Ratio of 1.1 - 1.2	10%	8%		
% With LoantoFHA Ratio > 1.2	11%	10%		
Average PTI	17%	17%	16%	16%
% Originated in 1998	25%	33%	36%	33%
% Originated in 1999	35%	33%	33%	34%
% Originated in 2000	40%	33%	31%	33%

Exhibit A.10: Analysis of Matched and Unmatched Experian and HMDA Loans in St Louis (1998-2000)

Charateristics	Experian Unmatched	Experian/HMDA Matched	HMDA Unmatched	All HMDA
Mortgaged Property Characteristics				
% Old Construction	55%	95%		
% New Construction	6%	5%		
% Missing Construction Information	39%	0%		
% Unit Size 1	98%	98%		
% Unit Size 2	1%	1%		
% Unit Size 3	1%	1%		
% Unit Size 4	1%	0%		
Borrower Neighborhood Characteristics (1990 Census)				
% In Underserved Tracts		26%	25%	25%
% Not in Underserved Tracts		74%	69%	71%
% Missing Underserved Tracts		0%	6%	4%
% in High Cost Cities	0%	0%	0%	0%
% in Average Cost Cities	100%	100%	100%	100%
% in Low Cost Cities	0%	0%	0%	0%
% In Center City	27%	26%	20%	21%
% Not in Center City	73%	74%	74%	74%
% Missing Center City Information	0%	0%	6%	4%
Average 5-year Depreciation Lagged 1 year	121%	120%	120%	120%
% In Area with Depreciation	0%	0%	0%	0%
% In Area with Appreciation up to 20%	25%	33%	36%	33%
% In Area with Appreciation over 20%	75%	67%	64%	67%
% In <90% Relative Income Tracts	23%	23%	27%	25%
% In 90-120% Relative Income Tracts	33%	34%	36%	35%
% In >120% Relative Income Tracts	44%	43%	37%	39%
% In <10% Minority Tracts	72%	73%	75%	74%
% In 10-30% Minority Tracts	16%	16%	11%	13%
% In >30% Minority Tracts	12%	11%	8%	9%
% Missing Minority Tract Information	0%	0%	6%	4%

Note: Jumbo loans are excluded from this analysis. The column for the Experian/HMDA Matched loans presents weighted numbers but the columns for the Experian Unmatched, HMDA Unmatched and All HMDA loans present unweighted numbers.

Exhibit A.11: Analysis of Matched and Unmatched Experian and HMDA Loans in Tampa (1998-2000)

Charateristics	Experian Unmatched	Experian/HMDA Matched	HMDA Unmatched	All HMDA
Borrower Characteristics				
Unweighted Number of Borrowers	35,756	45,669	89,181	134,850
Weighted Number of Borrowers		144,608		
Average Annual Income	$55,605	$54,512	$57,694	$56,749
Median Annual Income	$42,500	$43,000	$46,000	$45,000
Average Annual Income (Trimmed Top 1%)	$54,418	$51,802	$54,236	$53,411
% Estimated Income Information	0%	3%	3%	3%
Average FICO	679	683		
% With FICO < 620	16%	25%		
% With FICO 620 - 679	12%	18%		
% With FICO => 680	34%	57%		
% Missing FICO Information	39%	0%		
% White		80%	72%	75%
% Black		5%	4%	4%
% Hispanic		9%	6%	7%
% Other		3%	3%	3%
% Missing Race Information		3%	14%	10%
% Female	14%	29%	29%	29%
% Male	80%	71%	61%	65%
% Missing Gender Information	6%	0%	10%	7%
% Age 19-34	15%	25%		
% Age 35-49	30%	47%		
% Age 50-64	12%	20%		
% Age >65	4%	8%		
% Missing Age Information	39%	0%		
Loan Characteristics				
Average Loan Amount	$95,804	$89,769	$87,514	$88,541
Average LTV %	80%	85%		
% With LTV <= 90	56%	55%		
% With LTV 90 - 96	16%	18%		
% With LTV 97 - 98	6%	6%		
% With LTV > 98	9%	22%		
% Missing LTV Information	13%	0%		
Average Ratio of Loan Amount to FHA Loan Limit	83%	80%		
% With LoantoFHA Ratio <=.5	22%	23%		
% With LoantoFHA Ratio of .6 - 1	49%	53%		
% With LoantoFHA Ratio of 1.1 - 1.2	12%	9%		
% With LoantoFHA Ratio > 1.2	17%	14%		
Average PTI	19%	17%	15%	16%
% Originated in 1998	21%	31%	34%	31%
% Originated in 1999	32%	34%	35%	34%
% Originated in 2000	47%	34%	32%	35%

Exhibit A.11: Analysis of Matched and Unmatched Experian and HMDA Loans in Tampa (1998-2000)

Charateristics	Experian Unmatched	Experian/HMDA Matched	HMDA Unmatched	All HMDA
Mortgaged Property Characteristics				
% Old Construction	82%	88%		
% New Construction	18%	12%		
% Missing Construction Information	0%	0%		
% Unit Size 1	98%	98%		
% Unit Size 2	0%	0%		
% Unit Size 3	1%	1%		
% Unit Size 4	0%	1%		
Borrower Neighborhood Characteristics (1990 Census)				
% In Underserved Tracts		30%	31%	30%
% Not in Underserved Tracts		70%	69%	70%
% Missing Underserved Tracts		0%	0%	0%
% in High Cost Cities	0%	0%	0%	0%
% in Average Cost Cities	18%	61%	63%	60%
% in Low Cost Cities	82%	39%	37%	40%
% In Center City	29%	34%	30%	31%
% Not in Center City	71%	66%	70%	69%
% Missing Center City Information	0%	0%	0%	0%
Average 5-year Depreciation Lagged 1 year	117%	116%	115%	116%
% In Area with Depreciation	0%	0%	0%	0%
% In Area with Appreciation up to 20%	53%	66%	68%	65%
% In Area with Appreciation over 20%	47%	34%	32%	35%
% In <90% Relative Income Tracts	25%	26%	27%	26%
% In 90-120% Relative Income Tracts	31%	32%	33%	32%
% In >120% Relative Income Tracts	44%	42%	40%	41%
% In <10% Minority Tracts	57%	59%	61%	61%
% In 10-30% Minority Tracts	34%	32%	31%	31%
% In >30% Minority Tracts	9%	8%	8%	8%
% Missing Minority Tract Information	0%	0%	0%	0%

Note: Jumbo loans are excluded from this analysis. The column for the Experian/HMDA Matched loans presents weighted numbers but the columns for the Experian Unmatched, HMDA Unmatched and All HMDA loans present unweighted numbers.

Exhibit A.12: Analysis of Matched and Unmatched Experian and HMDA Loans in Washington DC (1998-2000)

Charateristics	Experian Unmatched	Experian/HMDA Matched	HMDA Unmatched	All HMDA
Borrower Characteristics				
Unweighted Number of Borrowers	24,214	32,027	241,153	273,180
Weighted Number of Borrowers		268,219		
Average Annual Income	$73,987	$68,312	$77,243	$76,197
Median Annual Income	$62,500	$60,000	$67,000	$66,000
Average Annual Income (Trimmed Top 1%)	$72,535	$66,228	$74,158	$73,197
% Estimated Income Information	0%	2%	2%	2%
Average FICO	691	699		
% With FICO < 620	17%	19%		
% With FICO 620 - 679	14%	16%		
% With FICO => 680	48%	65%		
% Missing FICO Information	21%	0%		
% White		59%	60%	60%
% Black		21%	16%	17%
% Hispanic		8%	6%	6%
% Other		8%	8%	8%
% Missing Race Information		3%	10%	9%
% Female	22%	34%	29%	30%
% Male	68%	66%	64%	65%
% Missing Gender Information	10%	0%	6%	6%
% Age 19-34	18%	34%		
% Age 35-49	27%	50%		
% Age 50-64	7%	13%		
% Age >65	2%	3%		
% Missing Age Information	46%	0%		
Loan Characteristics				
Average Loan Amount	$150,348	$146,561	$124,448	$127,072
Average LTV %	71%	87%		
% With LTV <= 90	34%	45%		
% With LTV 90 - 96	9%	13%		
% With LTV 97 - 98	6%	13%		
% With LTV > 98	7%	29%		
% Missing LTV Information	44%	0%		
Average Ratio of Loan Amount to FHA Loan Limit	75%	74%		
% With LoantoFHA Ratio <=.5	18%	18%		
% With LoantoFHA Ratio of .6 - 1	64%	65%		
% With LoantoFHA Ratio of 1.1 - 1.2	15%	14%		
% With LoantoFHA Ratio > 1.2	3%	4%		
Average PTI	20%	21%	17%	17%
% Originated in 1998	20%	30%	30%	29%
% Originated in 1999	35%	34%	34%	34%
% Originated in 2000	45%	36%	37%	37%

Exhibit A.12: Analysis of Matched and Unmatched Experian and HMDA Loans in Washington DC (1998-2000)

Charateristics	Experian Unmatched	Experian/HMDA Matched	HMDA Unmatched	All HMDA
Mortgaged Property Characteristics				
% Old Construction	70%	81%		
% New Construction	28%	19%		
% Missing Construction Information	1%	0%		
% Unit Size 1	97%	95%		
% Unit Size 2	2%	2%		
% Unit Size 3	1%	2%		
% Unit Size 4	1%	1%		
Borrower Neighborhood Characteristics (1990 Census)				
% In Underserved Tracts		36%	32%	33%
% Not in Underserved Tracts		64%	61%	61%
% Missing Underserved Tracts		0%	7%	6%
% in High Cost Cities	0%	0%	0%	0%
% in Average Cost Cities	100%	100%	100%	100%
% in Low Cost Cities	0%	0%	0%	0%
% In Center City	10%	13%	14%	13%
% Not in Center City	90%	87%	79%	80%
% Missing Center City Information	0%	0%	7%	6%
Average 5-year Depreciation Lagged 1 year	105%	104%	104%	104%
% In Area with Depreciation	0%	0%	0%	0%
% In Area with Appreciation up to 20%	100%	100%	100%	100%
% In Area with Appreciation over 20%	0%	0%	0%	0%
% In <90% Relative Income Tracts	24%	24%	29%	29%
% In 90-120% Relative Income Tracts	43%	39%	35%	35%
% In >120% Relative Income Tracts	33%	37%	36%	36%
% In <10% Minority Tracts	18%	17%	16%	16%
% In 10-30% Minority Tracts	54%	51%	52%	52%
% In >30% Minority Tracts	28%	33%	25%	26%
% Missing Minority Tract Information	0%	0%	7%	6%

Note: Jumbo loans are excluded from this analysis. The column for the Experian/HMDA Matched loans presents weighted numbers but the columns for the Experian Unmatched, HMDA Unmatched and All HMDA loans present unweighted numbers.

Exhibit A.13: Analysis of Loans by MSA

On All Matched, Conforming Loans, Weighted
(n=347,732; weighted sample=1,980,080)

	Baltimore	Chicago	Cleveland	Denver	Los Angeles	Oakland	Philadelphia	Portland	St. Louis	Tampa	Washington DC
Share of Loans	18%	58%	14%	25%	40%	14%	29%	15%	19%	21%	40%
MSA Characteristics											
Designated High, Medium or Low Cost (at least 1 year)	Medium	High	High	High	High	High	Medium	Medium	Medium	Low	Medium
% Minority (2000)	31.5%	32.6%	21.7%	18.0%	48.8%	41.3%	26.5%	12.5%	20.6%	15.3%	38.0%
% Individuals in Poverty (2000)	10%	11%	11%	8%	18%	10%	11%	10%	10%	11%	7%
Median Household Income	$49,938	$51,680	$42,089	$51,191	$42,189	$59,365	$47,536	$47,077	$44,437	$37,406	$62,216
% Households Owner-Occupied	59.7%	50.9%	60.2%	57.1%	41.1%	53.2%	63.2%	52.9%	61.6%	53.6%	54.7%
Median Owner-Occupied Household Value	$134,900	$166,200	$119,400	$176,600	$209,300	$289,100	$121,200	$170,000	$99,400	$93,800	$178,900
Number of Tracts in Our Data	176	549	616	337	386	207	575	335	405	399	351
% of Tracts Underserved in Our Data	53%	53%	46%	54%	66%	68%	50%	39%	49%	45%	55%
Borrower Characteristics											
Unweighted Number of Borrowers	11,612	47,408	24,795	45,969	20,841	13,683	30,296	37,023	38,409	45,669	32,027
Weighted Number of Borrowers	121,587	389,963	96,935	169,772	268,173	96,068	196,784	102,918	125,054	144,608	268,219
Average Annual Income	$59,803	$65,836	$57,318	$61,883	$65,925	$70,587	$60,959	$61,797	$56,011	$54,512	$68,312
Median Annual Income	$51,000	$57,000	$49,000	$54,000	$58,000	$63,000	$52,000	$54,000	$48,000	$43,000	$60,000
Average Annual Income (Trimmed Top 1%)	$57,630	$63,016	$55,137	$59,528	$62,878	$68,222	$58,406	$59,266	$53,813	$51,452	$66,101
% Estimated Income Information	2%	2%	2%	2%	3%	2%	2%	2%	1%	3%	2%
Average FICO	688	701	700	702	684	708	692	712	691	683	699
% With FICO <620	24%	18%	20%	16%	22%	14%	23%	13%	23%	25%	19%
% With FICO 620-680	16%	17%	14%	19%	21%	18%	15%	16%	14%	18%	16%
% With FICO =>680	60%	66%	66%	65%	57%	69%	62%	70%	63%	57%	65%
% Missing FICO Information	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
% White	72%	68%	86%	76%	40%	52%	75%	86%	80%	80%	59%
% Black	22%	10%	7%	3%	6%	7%	14%	1%	14%	5%	21%
% Hispanic	1%	14%	2%	14%	34%	18%	4%	4%	1%	9%	8%
% Other	3%	5%	2%	4%	16%	19%	4%	6%	3%	3%	8%
% Missing Race Information	3%	2%	3%	3%	4%	4%	4%	3%	2%	3%	3%
% Female	35%	25%	22%	28%	26%	27%	30%	24%	29%	29%	34%
% Age 19-34	33%	34%	33%	36%	28%	27%	31%	35%	34%	25%	34%
% Age 35-49	49%	49%	49%	47%	56%	54%	51%	47%	48%	47%	50%
% Age 50-64	15%	13%	15%	14%	14%	16%	15%	14%	15%	20%	13%
% Age >65	3%	3%	4%	3%	3%	3%	3%	3%	4%	8%	3%
% Missing Age Information	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%
Loan Characteristics											
Average Loan Amount	$117,870	$128,925	$106,126	$138,614	$150,579	$167,811	$108,463	$137,479	$96,681	$89,769	$146,561
Average LTV %	88	81	83	85	85	81	85	84	81	85	87
% With LTV<=80	30%	47%	39%	36%	39%	54%	39%	43%	75%	41%	34%
% With LTV 80-90	9%	15%	21%	14%	13%	13%	14%	16%	3%	14%	11%
% With LTV 90-96	13%	18%	17%	16%	15%	14%	19%	16%	4%	18%	13%
% With LTV 96-98	9%	6%	4%	10%	11%	7%	6%	8%	2%	6%	13%
% With LTV>98	39%	14%	19%	25%	22%	11%	23%	17%	16%	22%	29%
% Missing LTV Information	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Average Ratio of Loan Amount to FHA Loan Limit	0.67	0.67	0.67	0.75	0.74	0.83	0.71	0.81	0.70	0.80	0.74
% With LoantoFHA Ratio <=.5	31%	29%	34%	16%	17%	11%	32%	9%	35%	23%	18%
% With LoantoFHA Ratio of .5 - 1	55%	59%	52%	68%	66%	61%	49%	71%	46%	53%	65%
% With LoantoFHA Ratio of 1 - 1.2	8%	9%	7%	12%	14%	23%	9%	12%	8%	9%	14%
% With LoantoFHA Ratio of >1.2	5%	3%	7%	5%	2%	4%	10%	8%	10%	14%	4%

Exhibit A.13: Analysis of Loans by MSA

On All Matched, Conforming Loans, Weighted
(n=347,732; weighted sample=1,980,080)

	Baltimore	Chicago	Cleveland	Denver	Los Angeles	Oakland	Philadelphia	Portland	St. Louis	Tampa	Washington DC
Average PTI	0.19	0.19	0.18	0.22	0.23	0.23	0.17	0.22	0.17	0.17	0.21
% Originated in 1998	32%	31%	34%	32%	31%	33%	32%	35%	33%	31%	30%
% Originated in 1999	34%	34%	34%	34%	34%	36%	34%	33%	33%	34%	34%
% Originated in 2000	34%	35%	32%	34%	34%	31%	34%	32%	33%	34%	36%
Mortgaged Property Characteristics											
% New Construction	14%	9%	4%	19%	3%	5%	8%	16%	5%	12%	19%
% Unit Size 1	96%	92%	98%	96%	95%	96%	95%	97%	98%	98%	95%
% Unit Size 2	2%	2%	0%	2%	1%	1%	2%	2%	1%	0%	2%
% Unit Size 3	1%	4%	1%	1%	2%	1%	3%	1%	1%	1%	2%
% Unit Size 4	0%	2%	0%	1%	2%	1%	0%	0%	0%	1%	1%
Borrower Neighborhood Characteristics (1990 Census)											
Count of Tracts	176	549	616	337	386	207	575	335	405	399	351
% of Tracts that Are Underserved	53%	53%	46%	54%	66%	68%	50%	39%	49%	45%	55%
% In Underserved Tracts	25%	26%	25%	31%	50%	52%	24%	27%	26%	30%	36%
% In High Cost MSA	0%	2%	35%	7%	41%	74%	0%	0%	0%	0%	0%
% In Medium Cost MSA	100%	98%	65%	93%	59%	26%	100%	100%	100%	61%	100%
% In Low Cost MSA	0%	0%	0%	0%	0%	0%	0%	0%	0%	39%	0%
% In Center City	22%	30%	19%	23%	42%	15%	26%	36%	26%	34%	13%
Average 5-yr House Price Appreciation, Lagged 1 year	108%	118%	124%	119%	99%	110%	105%	144%	120%	116%	104%
% In Area with Depreciation	0%	0%	0%	0%	66%	33%	0%	0%	0%	0%	0%
% In Area with Appreciation up to 20%	100%	100%	0%	100%	34%	36%	100%	0%	33%	66%	100%
% In Area with Appreciation over 20%	0%	0%	100%	0%	0%	31%	0%	100%	67%	34%	0%
Lagged Depreciation in 1998	105%	118%	124%	119%	84%	98%	101%	150%	118%	111%	101%
Lagged Depreciation in 1999	107%	118%	124%	119%	98%	109%	104%	147%	120%	115%	103%
Lagged Depreciation in 2000	111%	118%	123%	119%	114%	125%	109%	137%	123%	121%	109%
% In Tracts with Income <90% of MSA Income	19%	18%	21%	25%	25%	31%	23%	25%	23%	26%	24%
% In Tracts with Income 90 - 120% of MSA Income	39%	31%	34%	23%	35%	29%	37%	41%	34%	32%	39%
% In Tracts with Income =>120% of MSA Income	42%	51%	44%	52%	40%	40%	40%	34%	43%	42%	37%
% In <10% Minority Tracts	56%	49%	81%	41%	1%	2%	69%	71%	73%	59%	17%
% In 10-30% Minority Tracts	26%	32%	12%	42%	31%	43%	21%	25%	16%	32%	51%
% In =>30% Minority Tracts	18%	19%	7%	17%	68%	55%	10%	4%	11%	8%	33%
Lending Market Characteristics											
FHA Market Share (by Count)	66%	36%	43%	50%	52%	29%	43%	32%	34%	42%	58%
GSE Market Share (by Count)	34%	64%	57%	50%	48%	71%	57%	68%	66%	58%	42%
FHA Market Share (by $)	58%	33%	39%	46%	48%	25%	33%	30%	26%	35%	52%
GSE Market Share (by $)	42%	67%	61%	54%	52%	75%	67%	70%	74%	65%	48%
HHI by Count	243	158	348	211	245	301	229	314	182	239	255
HHI by $	238	165	373	207	244	305	234	319	190	243	256

Exhibit A.14: Analysis of Loans by MSA

On Matched, Conforming, Fixed-Rate FHA-Eligible Loans, Weighted
(n=238,158; weighted sample=1,369,923)

	Baltimore	Chicago	Cleveland	Denver	Los Angeles	Oakland	Philadelphia	Portland	St. Louis	Tampa	Washington DC
Share of Loans	14%	41%	8%	17%	27%	8%	21%	10%	13%	14%	29%
MSA Characteristics											
Designated High, Medium or Low Cost (at least 1 year)	Medium	High	High	High	High	High	Medium	Medium	Medium	Low	Medium
% Minority (2000)	31.5%	32.6%	21.7%	18.0%	48.8%	41.3%	26.5%	12.5%	20.6%	15.3%	38.0%
% Individuals in Poverty (2000)	10%	11%	11%	8%	18%	10%	11%	10%	10%	11%	7%
Median Household Income	$49,938	$51,680	$42,089	$51,191	$42,189	$59,365	$47,536	$47,077	$44,437	$37,406	$62,216
% Households Owner-Occupied	59.7%	50.9%	60.2%	57.1%	41.1%	53.2%	63.2%	52.9%	61.6%	53.6%	54.7%
Median Owner-Occupied Household Value	$134,900	$166,200	$119,400	$176,600	$209,300	$289,100	$121,200	$170,000	$99,400	$93,800	$178,900
Number of Tracts in Our Data	176	544	578	335	381	201	571	335	400	399	351
% of Tracts Underserved in Our Data	53%	53%	43%	54%	66%	69%	50%	39%	49%	45%	55%
Borrower Characteristics											
Unweighted Number of Borrowers	9,007	33,467	12,988	31,808	14,460	7,770	23,141	24,144	27,052	30,654	23,667
Weighted Number of Borrowers	94,142	274,627	54,722	115,453	183,360	52,449	144,380	67,386	89,242	97,150	197,011
Average Annual Income	$53,292	$59,890	$51,511	$55,868	$58,705	$61,722	$51,116	$54,907	$47,487	$43,576	$61,343
Median Annual Income	$46,000	$53,000	$46,000	$49,000	$52,000	$55,000	$45,000	$48,000	$42,000	$37,000	$54,000
Average Annual Income (Trimmed Top 1%)	$51,285	$57,686	$50,003	$53,713	$56,076	$59,801	$49,171	$52,583	$45,774	$41,258	$59,214
% Estimated Income Information	2%	2%	2%	3%	3%	2%	2%	2%	1%	3%	2%
Average FICO	683	702	705	703	683	710	685	716	686	677	693
% With FICO <620	26%	18%	18%	16%	23%	13%	25%	12%	25%	27%	21%
% With FICO 620-680	16%	17%	14%	19%	21%	17%	16%	16%	15%	19%	17%
% With FICO =>680	58%	66%	68%	66%	56%	69%	59%	72%	60%	54%	62%
% Missing FICO Information	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
% White	69%	67%	91%	74%	36%	49%	72%	85%	78%	78%	56%
% Black	24%	10%	4%	3%	7%	8%	17%	1%	16%	6%	23%
% Hispanic	1%	16%	2%	16%	40%	23%	5%	5%	1%	10%	10%
% Other	3%	5%	1%	4%	14%	16%	4%	6%	2%	3%	8%
% Missing Race Information	2%	2%	2%	2%	4%	4%	4%	3%	2%	3%	3%
% Female	38%	26%	22%	30%	26%	29%	35%	27%	33%	33%	37%
% Age 19-34	34%	34%	34%	36%	29%	27%	32%	36%	35%	26%	35%
% Age 35-49	48%	49%	47%	47%	55%	52%	50%	45%	46%	45%	49%
% Age 50-64	15%	14%	15%	14%	13%	17%	15%	15%	15%	20%	13%
% Age >65	3%	3%	4%	3%	3%	3%	4%	4%	4%	9%	3%
% Missing Age Information	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%
Loan Characteristics											
Average Loan Amount	$103,610	$117,651	$94,527	$125,825	$135,896	$146,557	$89,132	$122,121	$78,635	$70,035	$130,953
Average LTV %	89	81	83	86	87	82	86	84	82	86	88
% With LTV<=80	26%	45%	39%	33%	32%	47%	33%	40%	70%	36%	30%
% With LTV 80-90	8%	13%	18%	12%	11%	11%	12%	14%	3%	12%	9%
% With LTV 90-96	11%	18%	17%	15%	15%	14%	19%	16%	4%	15%	11%
% With LTV 96-98	10%	7%	4%	11%	13%	10%	7%	10%	2%	7%	16%
% With LTV>98	45%	17%	22%	29%	29%	18%	29%	20%	20%	30%	35%
% Missing LTV Information	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%

Exhibit A.14: Analysis of Loans by MSA
On Matched, Conforming, Fixed-Rate FHA-Eligible Loans, Weighted
(n=238,158; weighted sample=1,369,923)

	Baltimore	Chicago	Cleveland	Denver	Los Angeles	Oakland	Philadelphia	Portland	St. Louis	Tampa	Washington DC
Average Ratio of Loan Amount to FHA Loan Limit	0.59	0.60	0.59	0.67	0.67	0.72	0.58	0.72	0.57	0.62	0.66
% With LoantoFHA Ratio <=.5	37%	33%	35%	19%	21%	16%	39%	11%	43%	30%	22%
% With LoantoFHA Ratio of .5 - 1	63%	67%	65%	81%	79%	84%	61%	89%	57%	70%	78%
% With LoantoFHA Ratio of 1 - 1.2	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
% With LoantoFHA Ratio of >1.2	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Average PTI	0.19	0.19	0.17	0.22	0.23	0.23	0.17	0.22	0.16	0.17	0.21
% Originated in 1998	34%	29%	45%	31%	31%	34%	32%	38%	36%	31%	29%
% Originated in 1999	33%	36%	24%	36%	36%	38%	34%	32%	32%	35%	33%
% Originated in 2000	33%	35%	31%	33%	33%	28%	33%	30%	31%	35%	38%
Mortgaged Property Characteristics											
% New Construction	11%	8%	1%	18%	3%	5%	6%	15%	3%	9%	16%
% Unit Size 1	96%	91%	98%	95%	94%	95%	94%	97%	98%	98%	95%
% Unit Size 2	2%	2%	1%	2%	2%	1%	2%	2%	1%	1%	3%
% Unit Size 3	2%	5%	1%	1%	2%	2%	3%	1%	1%	1%	2%
% Unit Size 4	0%	2%	0%	2%	2%	2%	0%	0%	1%	1%	1%
Borrower Neighborhood Characteristics (1990 Census)											
Count of Tracts	176	544	578	335	381	201	571	335	400	399	351
% of Tracts that Are Underserved	53%	53%	43%	54%	66%	69%	50%	39%	49%	45%	55%
% In Underserved Tracts	29%	28%	27%	35%	57%	59%	30%	32%	29%	36%	40%
% In High Cost MSA	0%	2%	34%	8%	43%	75%	0%	0%	0%	0%	0%
% In Medium Cost MSA	100%	98%	66%	92%	57%	25%	100%	100%	100%	61%	100%
% In Low Cost MSA	0%	0%	0%	0%	0%	0%	0%	0%	0%	39%	0%
% In Center City	24%	31%	20%	23%	41%	17%	31%	39%	28%	34%	14%
Average 5-yr House Price Appreciation, Lagged 1 year	108%	118%	124%	119%	99%	110%	105%	145%	120%	116%	104%
% In Area with Depreciation	0%	0%	0%	0%	67%	34%	0%	0%	0%	0%	0%
% In Area with Appreciation up to 20%	100%	100%	0%	100%	33%	38%	100%	0%	36%	65%	100%
% In Area with Appreciation over 20%	0%	0%	100%	0%	0%	28%	0%	100%	64%	35%	0%
Lagged Depreciation in 1998	105%	118%	124%	119%	84%	98%	101%	150%	118%	111%	101%
Lagged Depreciation in 1999	107%	118%	124%	119%	98%	109%	104%	147%	120%	115%	103%
Lagged Depreciation in 2000	111%	118%	123%	119%	114%	125%	109%	137%	123%	121%	109%
% In Tracts with Income <90% of MSA Income	21%	20%	22%	28%	29%	38%	28%	29%	26%	31%	26%
% In Tracts with Income 90 - 120% of MSA Income	42%	33%	37%	24%	37%	28%	40%	42%	37%	35%	42%
% In Tracts with Income =>120% of MSA Income	37%	47%	41%	48%	34%	34%	32%	29%	37%	34%	32%
% In <10% Minority Tracts	53%	48%	82%	36%	1%	1%	67%	69%	71%	58%	15%
% In 10-30% Minority Tracts	26%	32%	12%	44%	27%	39%	22%	27%	16%	31%	50%
% In =>30% Minority Tracts	20%	20%	6%	20%	73%	60%	12%	4%	12%	10%	35%
Lending Market Characteristics											
FHA Market Share (by Count)	71%	38%	45%	54%	59%	38%	51%	35%	41%	55%	65%
GSE Market Share (by Count)	29%	62%	55%	46%	41%	62%	49%	65%	59%	45%	35%
FHA Market Share (by $)	68%	37%	44%	53%	57%	37%	46%	36%	36%	54%	63%
GSE Market Share (by $)	32%	63%	56%	47%	43%	63%	54%	64%	64%	46%	37%
HHI by Count	261	160	336	230	253	290	248	309	191	263	270
HHI by $	258	166	355	227	250	291	253	312	197	267	273

Exhibit A.15: Analysis of Loans by MSA

On Matched, Conforming, Fixed-Rate FHA-Eligible FHA & GSE Loans with LTV 80-100%, Weighted

(n=114,780; weighted sample=674,238)

	Baltimore	Chicago	Cleveland	Denver	Los Angeles	Oakland	Philadelphia	Portland	St. Louis	Tampa	Washington DC
Share of Loans	9%	18%	3%	10%	15%	3%	11%	5%	3%	7%	16%
MSA Characteristics											
Designated High, Medium or Low Cost (at least 1 year)	Medium	High	High	High	High	High	Medium	Medium	Medium	Low	Medium
% Minority (2000)	31.5%	32.6%	21.7%	18.0%	48.8%	41.3%	26.5%	12.5%	20.6%	15.3%	38.0%
% Individuals in Poverty (2000)	10%	11%	11%	8%	18%	10%	11%	10%	10%	11%	7%
Median Household Income	$49,938	$51,680	$42,089	$51,191	$42,189	$59,365	$47,536	$47,077	$44,437	$37,406	$62,216
% Households Owner-Occupied	59.7%	50.9%	60.2%	57.1%	41.1%	53.2%	63.2%	52.9%	61.6%	53.6%	54.7%
Median Owner-Occupied Household Value	$134,900	$166,200	$119,400	$176,600	$209,300	$289,100	$121,200	$170,000	$99,400	$93,800	$178,900
Number of Tracts in Our Data	173	509	487	331	354	180	553	332	323	394	335
% of Tracts Underserved in Our Data	53%	51%	39%	54%	68%	75%	50%	39%	52%	44%	56%
Borrower Characteristics											
Unweighted Number of Borrowers	5,678	14,376	4,898	19,147	8,191	3,437	12,103	11,465	7,165	15,313	13,007
Weighted Number of Borrowers	58,910	118,379	20,576	67,352	102,336	22,621	73,390	31,693	22,632	47,833	108,516
Average Annual Income	$50,185	$55,655	$48,536	$52,462	$55,193	$57,551	$48,065	$52,885	$41,344	$40,889	$57,446
Median Annual Income	$45,000	$51,000	$45,000	$47,000	$50,000	$53,000	$44,000	$48,000	$37,000	$36,000	$51,000
Average Annual Income (Trimmed Top 1%)	$48,517	$54,290	$47,399	$50,652	$52,782	$56,221	$46,729	$51,051	$40,180	$39,094	$55,512
% Estimated Income Information	1%	1%	0%	2%	2%	2%	1%	1%	0%	1%	1%
Average FICO	662	673	673	681	659	685	661	696	651	651	668
% With FICO <620	33%	26%	27%	22%	31%	20%	33%	17%	35%	35%	29%
% With FICO 620-680	19%	21%	19%	24%	24%	23%	18%	21%	21%	23%	20%
% With FICO =>680	49%	53%	54%	55%	45%	57%	49%	63%	44%	42%	50%
% Missing FICO Information	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
% White	63%	56%	90%	69%	29%	45%	67%	83%	72%	74%	47%
% Black	31%	15%	4%	4%	8%	11%	21%	2%	23%	7%	30%
% Hispanic	2%	23%	3%	22%	52%	30%	5%	7%	2%	13%	13%
% Other	2%	4%	1%	3%	8%	12%	3%	6%	2%	3%	7%
% Missing Race Information	2%	2%	2%	2%	3%	3%	4%	3%	2%	3%	3%
% Female	40%	25%	21%	29%	24%	28%	34%	24%	36%	34%	39%
% Age 19-34	38%	39%	44%	42%	32%	33%	37%	44%	41%	33%	38%
% Age 35-49	48%	49%	44%	46%	56%	52%	49%	44%	47%	47%	50%
% Age 50-64	12%	10%	10%	9%	10%	13%	12%	11%	10%	14%	11%
% Age >65	2%	2%	2%	1%	2%	2%	2%	2%	2%	5%	2%
% Missing Age Information	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%
Loan Characteristics											
Average Loan Amount	$104,637	$120,350	$97,725	$128,746	$136,960	$149,282	$90,581	$128,538	$72,803	$74,736	$131,448

Exhibit A.15: Analysis of Loans by MSA

On Matched, Conforming, Fixed-Rate FHA-Eligible FHA & GSE Loans with LTV 80-100%, Weighted
(n=114,780; weighted sample=674,238)

	Baltimore	Chicago	Cleveland	Denver	Los Angeles	Oakland	Philadelphia	Portland	St. Louis	Tampa	Washington DC
Average LTV %	97	94	95	95	95	93	95	94	97	96	96
% With LTV<=80	4%	7%	6%	6%	7%	13%	7%	9%	2%	4%	5%
% With LTV 80-90	6%	18%	16%	12%	10%	14%	12%	18%	5%	11%	7%
% With LTV 90-96	10%	27%	21%	16%	15%	19%	19%	20%	9%	18%	11%
% With LTV 96-98	12%	12%	6%	17%	18%	15%	8%	13%	8%	10%	19%
% With LTV>98	68%	36%	51%	48%	50%	39%	54%	40%	76%	57%	57%
% Missing LTV Information	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Average Ratio of Loan Amount to											
FHA Loan Limit	0.60	0.61	0.61	0.69	0.67	0.73	0.59	0.75	0.53	0.65	0.66
% With LoantoFHA Ratio <=.5	36%	29%	30%	16%	18%	11%	37%	5%	52%	22%	20%
% With LoantoFHA Ratio of .5 - 1	64%	71%	70%	84%	82%	89%	63%	95%	48%	78%	80%
% With LoantoFHA Ratio of 1 - 1.2	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
% With LoantoFHA Ratio of >1.2	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Average PTI	0.19	0.20	0.19	0.23	0.24	0.24	0.17	0.23	0.17	0.18	0.22
% Originated in 1998	35%	29%	45%	30%	32%	36%	34%	36%	16%	26%	29%
% Originated in 1999	33%	36%	23%	36%	36%	38%	34%	32%	42%	36%	33%
% Originated in 2000	31%	35%	32%	33%	32%	27%	33%	32%	42%	38%	38%
Mortgaged Property Characteristics											
% New Construction	10%	5%	1%	14%	3%	4%	4%	13%	4%	8%	16%
% Unit Size 1	96%	90%	99%	95%	95%	96%	94%	97%	98%	98%	95%
% Unit Size 2	2%	2%	0%	2%	2%	1%	2%	2%	0%	1%	2%
% Unit Size 3	2%	6%	1%	1%	1%	2%	3%	1%	1%	1%	2%
% Unit Size 4	0%	2%	0%	2%	2%	2%	0%	0%	0%	1%	1%
Borrower Neighborhood Characteristics (1990 Census)											
Count of Tracts	173	509	487	331	354	180	553	332	323	394	335
% of Tracts that Are Underserved	53%	51%	39%	54%	68%	75%	50%	39%	52%	44%	56%
% In Underserved Tracts	33%	34%	30%	42%	65%	65%	34%	37%	41%	37%	44%
% In High Cost MSA	0%	1%	33%	7%	43%	72%	0%	0%	0%	0%	0%
% In Medium Cost MSA	100%	99%	67%	93%	57%	28%	100%	100%	100%	59%	100%
% In Low Cost MSA	0%	0%	0%	0%	0%	0%	0%	0%	0%	41%	0%
% In Center City	26%	35%	25%	24%	40%	14%	33%	40%	34%	37%	11%
Average 5-yr House Price Appreciation, Lagged 1 year	108%	118%	124%	119%	99%	109%	105%	145%	121%	116%	104%
% In Area with Depreciation	0%	0%	0%	0%	68%	36%	0%	0%	0%	0%	0%
% In Area with Appreciation up to 20%	100%	100%	0%	100%	32%	38%	100%	0%	16%	62%	100%
% In Area with Appreciation over 20%	0%	0%	100%	0%	0%	27%	0%	100%	84%	38%	0%
Lagged Depreciation in 1998	105%	118%	124%	119%	84%	98%	101%	150%	118%	111%	101%
Lagged Depreciation in 1999	107%	118%	124%	119%	98%	109%	104%	147%	120%	115%	103%
Lagged Depreciation in 2000	111%	118%	123%	119%	114%	125%	109%	137%	123%	121%	109%
% In Tracts with Income <90% of MSA Income	23%	24%	24%	32%	32%	42%	32%	34%	37%	32%	27%
% In Tracts with Income 90 - 120% of MSA Income	46%	37%	44%	26%	39%	28%	43%	41%	45%	36%	45%
% In Tracts with Income =>120% of MSA Income	31%	39%	32%	42%	29%	30%	25%	25%	19%	32%	28%

Exhibit A.15: Analysis of Loans by MSA

On Matched, Conforming, Fixed-Rate FHA-Eligible FHA & GSE Loans with LTV 80-100%, Weighted

(n=114,780; weighted sample=674,238)

	Baltimore	Chicago	Cleveland	Denver	Los Angeles	Oakland	Philadelphia	Portland	St. Louis	Tampa	Washington DC
% In <10% Minority Tracts	48%	40%	78%	28%	0%	1%	64%	66%	63%	53%	13%
% In 10-30% Minority Tracts	28%	35%	15%	47%	22%	35%	22%	28%	23%	35%	49%
% In =>30% Minority Tracts	24%	25%	7%	25%	78%	65%	13%	5%	14%	12%	38%
Lending Market Characteristics											
FHA Market Share (by Count)	81%	50%	58%	68%	69%	53%	63%	49%	83%	68%	74%
GSE Market Share (by Count)	19%	50%	42%	32%	31%	47%	37%	51%	17%	32%	26%
FHA Market Share (by $)	78%	48%	56%	65%	67%	51%	57%	48%	82%	67%	72%
GSE Market Share (by $)	22%	52%	44%	35%	33%	49%	43%	52%	18%	33%	28%
HHI by Count	371	229	412	312	335	498	355	419	309	443	317
HHI by $	359	232	401	303	332	490	359	416	300	427	318
Model Results											
Overlap Range in Origination Model (MSA-level non-parametric tolerance intervals):											
FHA Lower Limit (predicted log odds)	1.28	-0.63	-0.04	0.42	0.51	-0.36	0.61	-0.58	1.92	0.38	0.42
GSE Upper Limit (predicted log odds)	1.93	0.26	0.52	0.92	0.95	0.49	0.69	0.07	1.25	1.09	2.18
Overlap Range in Default Model (MSA-level non-parametric tolerance intervals):											
FHA Lower Limit (predicted log odds)	-4.95	-4.67	-5.05	-5.06	-4.05	-4.11	-4.63	-5.43	-4.90	-4.70	-4.52
GSE Upper Limit (predicted log odds)	-2.35	-2.37	-2.45	-2.89	-1.81	-1.87	-2.10	-3.27	-2.53	-2.28	-1.91
Average Predicted Likelihood of Default (sectors by MSA-level non-parametric tolerance intervals):											
Prediction	0.07	0.05	0.04	0.04	0.09	0.07	0.08	0.02	0.06	0.06	0.07
Log Odds of Prediction	-3.30	-3.67	-3.94	-3.87	-2.74	-3.07	-3.16	-4.49	-3.34	-3.38	-3.10

Exhibit A.16: Origination Model Results by MSA, FHA vs. GSE

FHA=1, GSE=0
On Matched, Conforming, Fixed-Rate FHA-Eligible FHA
& GSE Loans with LTV 80-100%, Weighted
(n=114,780; weighted sample=674,238)

Parameter	Label	Baltimore		Chicago		Cleveland		Denver	
		Estimate	Pr > ChiSq	Estimate	Pr > ChiSq	Estimate	Pr > ChiSq	Estimate	Pr > ChiSq
Intercept	Intercept	10.1406	<.0001	5.0731	0.1897	30.2718	0.1426	-2.8098	1
ex_fico	FICO Score	-0.0023	0.0018	-0.0040	<.0001	-0.0012	0.1498	-0.0030	<.0001
FICOs620	Marginal spline for FICO above 620	-0.0040	<.0001	-0.0070	<.0001	-0.0085	<.0001	-0.0068	<.0001
EX_LTV	Loan to Value Ratio	0.1073	<.0001	0.0420	<.0001	0.0422	<.0001	0.0689	<.0001
LTVs95	Marginal spline for LTV above 95%	0.1490	<.0001	1.3540	<.0001	1.0340	<.0001	1.3028	<.0001
LTVs98	Marginal spline for LTV above 98%	-0.0221	0.849	0.3964	<.0001	2.2482	<.0001	1.1378	<.0001
LTVs99	Marginal spline for LTV above 99%	-1.5116	<.0001	-3.0819	<.0001	-5.1921	<.0001	-4.5666	<.0001
roundedloantoFHA	Ratio of Borrower's Loan to FHA Loan Limit	-3.5119	<.0001	-0.8190	<.0001	-0.5280	0.0048	-3.2048	<.0001
PTI_calc	Payment to Income Ratio- Calculated	9.4115	<.0001	4.4399	<.0001	10.9841	<.0001	1.5792	0.0059
PTIs20	Marginal spline for PTI above .20	-1.9696	0.1353	-0.8469	0.258	-8.2696	<.0001	2.2519	0.0256
PTIs28	Marginal spline for PTI above .28	-12.8952	<.0001	-4.0954	<.0001	-6.5015	0.0187	-3.3959	<.0001
age_3549	Borrower Aged 35-49	-0.1681	<.0001	-0.0613	0.0077	0.0383	0.5019	-0.0334	0.3055
age_5064	Borrower Aged 50-64	0.0462	0.4629	-0.2735	<.0001	-0.1420	0.1418	0.0149	0.7751
age_65	Borrower Aged 65 and up	-0.1703	0.2786	-0.5330	<.0001	-0.5474	0.0077	-0.1623	0.188
female	Female Borrower	0.0714	0.0962	-0.0718	0.006	0.0478	0.4732	0.1852	<.0001
race_black	African American Borrower	0.7146	<.0001	0.2068	<.0001	0.4447	0.0079	0.3015	0.0012
race_hisp	Hispanic Borrower	-0.4129	0.0047	0.4911	<.0001	-0.5168	0.0027	0.4143	<.0001
race_other	Other Borrower (Native American, Asian, or Other)	-0.9418	<.0001	0.1784	0.0006	-0.0412	0.873	-0.1294	0.1053
new_constr	New Construction	0.5384	<.0001	0.8159	<.0001	1.1604	<.0001	0.1502	0.0003
relinc_c90t	Census Tract family income relative toMSA median family income in 1990	0.3546	<.0001	0.5204	<.0001	0.0713	0.5298	0.1042	0.0514
c90t_pminorty	Pct minority	-0.0023	0.0638	-0.00089	0.1522	-0.00124	0.6403	-0.00448	0.0005
ccity_n	Center City tract indicator	-0.0839	0.1661	-0.3281	<.0001	0.4612	<.0001	-0.3793	<.0001
high_cost	High Cost City based on FHA Loan Limits	0		0.9311	<.0001	-1.1064	<.0001	-0.9500	<.0001
low_cost	Low Cost City based on FHA Loan Limits	0		0		0		0	
hpiL5	OFHEO House Price 5-yr change lagged 1 yr	-17.807	<.0001	-7.3407	0.0248	-28.9899	0.0802	-0.7702	1
year98	Loan Originated in 1998	-0.3499	<.0001	0.3416	<.0001	-0.151	0.1032	0.7873	0.9994
year00	Loan Originated in 2000	0.0631		-0.2178	<.0001	0		0.0782	0.9999

	Baltimore	Chicago	Cleveland	Denver
Percent Concordant	96	95.6	96.2	96.3
Percent Discordant	3.7	4.3	3.7	3.6
Somers' D	0.923	0.914	0.925	0.927
Pseudo R2				
Log Likelihood- Intercept only	57982.285	164102.0	27991.954	84835.492
Log Likelihood- Intercept and Covariates	20302.349	61151.143	9988.68	30099.463

Exhibit A.16: Origination Model Results by MSA, FHA vs. GSE

FHA=1, GSE=0
On Matched, Conforming, Fixed-Rate FHA-Eligible FHA
& GSE Loans with LTV 80-100%, Weighted
(n=114,780; weighted sample=674,238)

Parameter	Label	Los Angeles		Oakland		Philadelphia		Portland	
		Estimate	Pr > ChiSq	Estimate	Pr > ChiSq	Estimate	Pr > ChiSq	Estimate	Pr > ChiSq
Intercept	Intercept	1.6413	0.0054	-8.8687	<.0001	2.4176	1	4.3408	<.0001
ex_fico	FICO Score	-0.0060	<.0001	-0.00055	0.555	-0.0029	<.0001	-0.0034	<.0001
FICOs620	Marginal spline for FICO above 620	-0.0013	0.0335	-0.00838	<.0001	-0.0045	<.0001	-0.0070	<.0001
EX_LTV	Loan to Value Ratio	0.0396	<.0001	0.1080	<.0001	0.0192	<.0001	-0.0205	<.0001
LTVs95	Marginal spline for LTV above 95%	1.4105	<.0001	1.0893	<.0001	1.2932	<.0001	1.0889	<.0001
LTVs98	Marginal spline for LTV above 98%	1.7610	<.0001	1.5571	<.0001	0.9944	<.0001	3.9935	<.0001
LTVs99	Marginal spline for LTV above 99%	-6.4561	<.0001	-5.4819	<.0001	-3.0859	<.0001	-8.9198	<.0001
roundedloantoFHA	Ratio of Borrower's Loan to FHA Loan Limit	-2.5480	<.0001	-2.3103	<.0001	-1.7924	<.0001	-2.3587	<.0001
PTI_calc	Payment to Income Ratio- Calculated	0.8707	0.0745	-0.2169	0.8453	7.4840	<.0001	-1.1456	0.1938
PTIs20	Marginal spline for PTI above .20	6.5349	<.0001	6.9786	<.0001	-7.0385	<.0001	12.3161	<.0001
PTIs28	Marginal spline for PTI above .28	-7.2178	<.0001	-8.0909	<.0001	0.7379	0.6533	-10.7563	<.0001
age_3549	Borrower Aged 35-49	-0.1188	<.0001	-0.1868	<.0001	0.0526	0.1135	-0.1351	0.0027
age_5064	Borrower Aged 50-64	-0.1140	0.0095	-0.5394	<.0001	-0.1130	0.0267	-0.2736	0.0002
age_65	Borrower Aged 65 and up	-0.3950	<.0001	-0.8553	0.0003	0.1108	0.2646	-0.5933	0.0007
female	Female Borrower	0.0727	0.0143	0.3674	<.0001	-0.0552	0.104	0.1660	0.0007
race_black	African American Borrower	0.5595	<.0001	0.4026	<.0001	0.5907	<.0001	0.9908	<.0001
race_hisp	Hispanic Borrower	0.6026	<.0001	0.2902	<.0001	0.4324	<.0001	0.6349	<.0001
race_other	Other Borrower (Native American, Asian, or Other)	-0.3900	<.0001	-0.4366	<.0001	0.2371	0.0054	-0.0706	0.4312
new_constr	New Construction	0.1458	0.0859	1.0636	<.0001	0.5059	<.0001	0.3385	<.0001
relinc_c90t	Census Tract family income relative toMSA median family income in 1990	0.4614	<.0001	-0.3330	0.0025	-0.2230	0.0006	-0.1242	0.2165
c90t_pminorty	Pct minority	0.0020	0.0021	-0.0035	0.0213	-0.00006	0.9513	-0.00656	0.0054
ccity_n	Center City tract indicator	-0.4069	<.0001	-1.0021	<.0001	-0.6176	<.0001	-0.3265	<.0001
high_cost	High Cost City based on FHA Loan Limits	0.00877	0.8705	-0.3077	0.0013	0		0	
low_cost	Low Cost City based on FHA Loan Limits	0		0		0		0	
hpiL5	OFHEO House Price 5-yr change lagged 1 yr	-2.2656	<.0001	0.2084	0.6082	-3.2938	1	-0.0624	0.9107
year98	Loan Originated in 1998	0.1008	0.3207	0.1884	0.1085	-0.3826	0.9999	0.0414	0.511
year00	Loan Originated in 2000	0		0		-0.0786	1	0.0726	

	Los Angeles	Oakland	Philadelphia	Portland
Percent Concordant	96.7	96.2	96.6	95.9
Percent Discordant	3.2	3.7	3.2	4
Somers' D	0.935	0.926	0.934	0.919
Pseudo R2				
Log Likelihood- Intercept only	127072.36	31295.441	96976.448	43906.871
Log Likelihood- Intercept and Covariates	42630.343	11307.473	31632.759	15457.333

Exhibit A.16: Origination Model Results by MSA, FHA vs. GSE

FHA=1, GSE=0
On Matched, Conforming, Fixed-Rate FHA-Eligible FHA
& GSE Loans with LTV 80-100%, Weighted
(n=114,780; weighted sample=674,238)

Parameter	Label	St. Louis		Tampa		Washington DC	
		Estimate	Pr > ChiSq	Estimate	Pr > ChiSq	Estimate	Pr > ChiSq
Intercept	Intercept	25.0190	<.0001	1.1421	0.5943	9.4565	<.0001
ex_fico	FICO Score	-0.0015	0.2372	-0.0035	<.0001	-0.0009	0.02
FICOs620	Marginal spline for FICO above 620	-0.0117	<.0001	-0.0063	<.0001	-0.0051	<.0001
EX_LTV	Loan to Value Ratio	-0.0069	0.5759	0.0148	0.0012	0.0742	<.0001
LTVs95	Marginal spline for LTV above 95%	1.6116	<.0001	1.0861	<.0001	0.9738	<.0001
LTVs98	Marginal spline for LTV above 98%	2.6220	<.0001	3.0699	<.0001	1.3762	<.0001
LTVs99	Marginal spline for LTV above 99%	-7.6022	<.0001	-6.9572	<.0001	-4.0123	<.0001
roundedloantoFHA	Ratio of Borrower's Loan to FHA Loan Limit	-3.2373	<.0001	-1.8254	<.0001	-2.5015	<.0001
PTI_calc	Payment to Income Ratio- Calculated	13.5799	<.0001	9.7138	<.0001	2.0073	<.0001
PTIs20	Marginal spline for PTI above .20	-12.6018	0.0008	-6.1031	<.0001	5.1734	<.0001
PTIs28	Marginal spline for PTI above .28	1.4148	0.8306	-4.2681	0.0006	-6.2898	<.0001
age_3549	Borrower Aged 35-49	-0.0192	0.8261	-0.1463	0.0003	-0.0422	0.0829
age_5064	Borrower Aged 50-64	-0.0905	0.5026	-0.4112	<.0001	0.1271	0.001
age_65	Borrower Aged 65 and up	-1.0525	<.0001	-0.7970	<.0001	0.3570	<.0001
female	Female Borrower	0.0853	0.3699	0.0824	0.0308	0.3802	<.0001
race_black	African American Borrower	0.7144	<.0001	1.2220	<.0001	0.5846	<.0001
race_hisp	Hispanic Borrower	1.2385	<.0001	0.5643	<.0001	-0.0749	0.0403
race_other	Other Borrower (Native American, Asian, or Other)	0.1104	0.719	0.2710	0.0058	-0.6132	<.0001
new_constr	New Construction	-0.4048	0.1334	0.6360	<.0001	0.1403	<.0001
relinc_c90t	Census Tract family income relative toMSA median family income in 1990	-0.1850	0.3873	-0.4078	<.0001	0.2156	<.0001
c90t_pminorty	Pct minority	0.0048	0.142	-0.00247	0.0639	0.0054	<.0001
ccity_n	Center City tract indicator	-0.1004	0.2332	-0.1604	<.0001	0.2267	<.0001
high_cost	High Cost City based on FHA Loan Limits	0		0		0	
low_cost	Low Cost City based on FHA Loan Limits	0		-0.2102	0.0539	0	
hpiL5	OFHEO House Price 5-yr change lagged 1 yr	-20.7532	<.0001	-1.0348	0.5709	-15.2966	<.0001
year98	Loan Originated in 1998	-0.0451	0.8113	0.3396	0.0006	-0.3336	<.0001
year00	Loan Originated in 2000	-0.0040		-0.0604		0	

	St. Louis	Tampa	Washington DC
Percent Concordant	98	95.8	94
Percent Discordant	1.8	4	5.9
Somers' D	0.962	0.918	0.881
Pseudo R2			
Log Likelihood- Intercept only	20594.433	59693.73	124235.05
Log Likelihood- Intercept and Covariates	4906.692	22214.132	55272.026