

Federal Reserve Bank of New York
Staff Reports

Subprime Mortgage Pricing: The Impact of Race, Ethnicity,
and Gender on the Cost of Borrowing

Andrew Haughwout
Christopher Mayer
Joseph Tracy

Staff Report no. 368
April 2009

This paper presents preliminary findings and is being distributed to economists and other interested readers solely to stimulate discussion and elicit comments. The views expressed in the paper are those of the authors and are not necessarily reflective of views at the Federal Reserve Bank of New York or the Federal Reserve System. Any errors or omissions are the responsibility of the authors.

Subprime Mortgage Pricing: The Impact of Race, Ethnicity, and Gender on the Cost of Borrowing

Andrew Haughwout, Christopher Mayer, and Joseph Tracy

Federal Reserve Bank of New York Staff Reports, no. 368

April 2009

JEL classification: G21, D40

Abstract

Some observers have argued that minority borrowers and neighborhoods were targeted for expensive credit in 2004-06, the peak period for subprime lending. To investigate this claim, we take advantage of a new data set that merges demographic information on subprime borrowers with information on the mortgages they took out. In a sample of more than 75,000 adjustable-rate mortgages, we find no evidence of adverse pricing by race, ethnicity, or gender in either the initial rate or the reset margin. Indeed, if any pricing differential exists, minority borrowers appear to pay slightly lower rates, as do those borrowers in Zip codes with a larger percentage of black or Hispanic residents or a higher unemployment rate. Mortgage rates are also lower in locations that previously had higher rates of house price appreciation. These results suggest some economies of scale in subprime lending. Yet there are important caveats: we are unable to measure points and fees at loan origination, and the data do not indicate whether borrowers might have qualified for less expensive conforming mortgages.

Key words: subprime, mortgages, HMDA

Haughwout: Federal Reserve Bank of New York (e-mail: andrew.haughwout@ny.frb.org). Mayer: Columbia University (e-mail: cm310@columbia.edu). Tracy: Federal Reserve Bank of New York (e-mail: joseph.tracy@ny.frb.org). The authors wish to acknowledge especially valuable comments and discussions with Robert Avery, Glenn Canner, and Karen Pence. Ebiere Okah and Rembrandt Koning provided extremely careful and excellent research assistance as well as many suggestions. The Paul Milstein Center for Real Estate at Columbia Business School provided funding in support of this analysis. The views expressed in this paper are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System.

The subprime lending boom increased the ability of many Americans to get credit to purchase a house. Yet concerns persist that not all borrowers have been treated equally. Previous research suggests that subprime loans were particularly concentrated in neighborhoods with a high concentration of black and Hispanic residents (Mayer and Pence, 2007). Some commentators have been concerned that minority borrowers were steered into subprime loans in some cases when they might have qualified for cheaper conforming loans, or that minority borrowers were given subprime loans that had fees or rates that were too high.

Previous research on housing markets suggests that such concerns might be warranted. Beginning in the early 1990s, data collected from lenders through the Home Mortgage Disclosure Act (HMDA) indicate that black or Hispanic applicants were more likely to be rejected for a mortgage relative to a white applicant, even when controlling for credit scores or other observable individual risk factors (Munnell *et al* 1996). Subsequent research showed that minority borrowers might also have been more likely to default on loans, but these findings were less clear in that they did not control for basic ex-ante risk factors (Ladd, 1998). Even controlling for the likelihood of default, Canner *et al* (1991) argue that minorities still face reduced access to conventional lending markets.

Recent studies of consumer loans have amplified concerns that minorities still face disparate treatment when applying for credit. For example, Charles *et al* (2008) show that blacks pay appreciably higher rates than other borrowers when financing a new car. Some portion of the higher payments comes from a higher proportion of blacks who use more expensive finance companies, but even among borrowers with comparable risk profiles using finance companies, blacks still pay higher rates. Similarly, Ravina (2008) finds that black

borrowers on Prosper.com, a successful online lending market, pay rates that are over 1% higher than comparably risky white borrowers. Ravina attributes the higher rates for blacks to the fact that black lenders, who do not charge higher rates to black borrowers, are relatively underrepresented on Prosper.com relative to black borrowers.

Despite the size of the mortgage market, as well as previous evidence on racial and ethnic differences in access to lending for housing, there are no recent studies that we have found on mortgage rates for minority borrowers. Below, we examine mortgage rates charged to a group of subprime mortgage borrowers using an innovative new dataset created by merging information on the race, ethnicity and gender of mortgage borrowers (as reported under HMDA) with mortgage pricing and risk variables reported by LoanPerformance (LP). Through extensive work, we have been able to match approximately 70 percent of loans in LP to a unique mortgage in HMDA. The merged dataset allows us to examine racial, ethnic, and gender differences in mortgage lending, controlling for both the risk profile of the mortgage and the characteristics of the neighborhood where the property is located.

As the subprime market took off between 2000 and 2006, a variety of new products became available for financing housing. The available contracts were differentiated along many dimensions, including term, amortization schedule, and the allocation of future interest rate risk between borrower and lender. Because each of these features has effects on the value, timing and probability of repayments, the precise way that they are combined into products will affect their value to borrowers and lenders. Thus, if we are to understand the pricing of loans, it is important that we examine a specific part of the market at a particular time, so that loan features and credit conditions are common for all the contracts we observe.

We focus on so-called “2-28” mortgages originated in August 2005. The 2-28 is a hybrid adjustable rate mortgage (ARM) in which borrowers are charged an initial mortgage rate for two years, followed by biannual rate resets based on a margin over a short-term rate. The 2-28 was a very popular form of subprime borrowing, and accounted for approximately 80% of subprime variable-rate loans and over 53% of all subprime loans outstanding at the end of 2007 (Sherlund 2008). We choose August 2005 for our analysis because LP coverage of the subprime market was very high in that year, and August was the peak month for originations of subprime loans. As we will see below, our initial dataset includes over 80,000 subprime 2-28 loans originated in that month. In spite of its wide appeal and importance in the overall subprime universe, however, the 2-28 contract is only one part of a very large market – about 4.4 million first-lien mortgages were originated in 2005 – and our results may be specific to the part of the market we analyze.

There are at least three dimensions along which mortgage lenders may treat similar groups of borrowers differently. First, as discussed in much of the literature reviewed above, they may simply refuse to offer credit at all. Second, they may “steer” accepted applicants into less attractive or more costly products, like subprime loans. Finally, even at a particular time, they may price a given product differently for different borrowers.

Our approach precludes analysis of important questions related to selection into the 2-28 product. Our data do not allow us to determine whether some borrowers were “steered” into subprime mortgages, or into 2-28s in particular, based on their demographic characteristics. Instead, within the 2-28 product category, we examine determinants of both the initial interest rate, as well as the margin used to adjust the rate after two years. As well, we

add one more important caveat; we are unable to directly observe the points and fees paid when the borrower initially took out the mortgage, so it is possible that we are missing data that might show disparate treatment in loan origination costs.

In contrast to previous findings, our results show that if anything, minority borrowers get slightly favorable terms, although the size of these effects are quite small. Black and Hispanic borrowers pay very slightly lower initial mortgage rates than other borrowers--about 2.5 basis points (0.025 percent) compared with a mean initial mortgage rate of 7.3 percent. Black and Hispanic borrowers also have slightly lower margins (about 1.7 to 5 basis points, or 0.0017 to 0.005 percent) compared to a mean margin of 5.9 percent. Asian borrowers pay slightly higher initial rates and reset margins (about 3 basis points). We find no appreciable differences in lending terms by the gender of the borrower. These results control for the mortgage risk characteristics and neighborhood composition. While many of these differences are statistically significant, they are economically insignificant.

A second important finding is that 2/28 mortgages were cheaper in Zip Codes with a higher percentage of Asian, black and Hispanic residents, as well as in counties with higher unemployment rates, once we control for the individual risk characteristics of the borrower. Taken in conjunction with the results in Mayer and Pence (2007) showing that high minority neighborhoods and locations with higher unemployment rates have a higher concentration of subprime loans, this is consistent with a small but positive credit supply shock in these neighborhoods. The fact that subprime loans were cheaper in high minority neighborhoods and counties with higher unemployment, possibly due to economies of scale, might help explain

why these neighborhoods had higher concentration of subprime loans.¹ Of course, these results cannot provide insight into whether some of the subprime borrowers would have qualified for a lower-cost conventional loan, as some have alleged.

Finally, our results also show that subprime loans were less expensive in metropolitan areas with greater past rates of house price appreciation. This finding is consistent with the idea that lenders may have expected higher rates of future house price appreciation in these neighborhoods and thus were willing to accept lower mortgage rates.² Such a finding can help explain why subprime mortgages were also more prevalent in high house price appreciation markets (Mayer and Pence, 2007), although the question of whether this was due to excessive expectations of borrowers and/or lenders can not be answered with our data.

The next section of the paper examines the data and describes the merging process for HMDA and LP data. Next we summarize the data and consider regressions of the cost of borrowing. The paper concludes with a brief policy discussion and a path for future research.

2. Data

Much of the innovative analysis in this paper is driven by our data. We merge together two datasets to examine the role of race, ethnicity, and gender in borrowing costs. The LoanPerformance (LP) database contains information on home location, mortgage amount,

¹ Previous research also found that FHA loans, the government insured predecessor to subprime lending, also were much more highly concentrated in minority neighborhoods.

² Piskorski and Tchisty (2008) develop a model showing that lenders optimally lend more in markets with faster rates of house price appreciation.

lending terms, mortgage risk factors, and monthly payments for the bulk of securitized loans. The Home Mortgage Disclosure Act requires lenders to report data on borrower demographics, income, geographic location, lender name, and in some cases, the mortgage yield for almost all loan applications in the United States. Therefore most LP mortgages should be contained in the HMDA database. We identify matching loans using the common data fields across the databases. Our analysis is limited to subprime mortgages to focus on the portion of the market with the riskiest loans. We also focus on mortgages originated in 2005, when LP appears to have the strongest coverage of the subprime market (Mayer and Pence, 2007).

LoanPerformance

First American LoanPerformance, a subsidiary of First American CoreLogic, Inc., provides information on securitized mortgages in subprime pools. The data do not include mortgages held in portfolio, securitized mortgages in prime, jumbo, or alt-A pools, or loans guaranteed by government agencies such as the Federal Housing Administration and the Veterans' Administration or by government-sponsored enterprises such as Fannie Mae, Freddie Mac or Ginnie Mae. The data also exclude loans handled by servicers that do not report to LoanPerformance.

We classify loans as subprime if they were packaged into a deal classified as subprime in LoanPerformance. The guidelines for what type of mortgage can be sold into a subprime pool vary across securitizers. In general, borrowers in subprime pools tend to have low credit scores and high loan-to-value ratios. On occasion, securitizers include a handful of near-prime or prime loans in these pools.

The LP data contain extensive information on the characteristics of the loan, such as the mortgage type, the interest rate, the loan purpose (purchase or refinance), and whether the loan has a prepayment penalty. However, data on upfront points and fees are not included. LP has less detailed information about the borrower, reporting the FICO credit score, the borrower's reported debt-to-income ratio, and the extent to which that income is documented. There is relatively little information about the property beyond the sale or appraised price, the type of property, and its state and Zip Code.

We use the following LP data fields: Zip Code, Origination Date, First Payment Date, Lien, Occupancy (owner-occupied or not), Purpose (purchase or refinance), Loan Amount, and Originator Name. As well we use the rating of the deal that the loan was packaged in. The 5-digit zip code and loan amount are retained as reported. The origination date is used as reported. However, a variable is created reporting whether the origination date is imputed or not. LP often imputes the origination date by assuming origination two months before the first payment date. Therefore we classify any loan with an origination date exactly two months before first payment date as having an imputed origination date. Lien is classified as first lien or subordinate. We restrict our analysis to first-lien mortgages.

HMDA

Under the Home Mortgage Disclosure Act, most originators must report basic attributes of the mortgage applications that they receive in metropolitan statistical areas to the Federal Financial Institutions Examination Council. These data are considered the most comprehensive source of mortgage data, and cover an estimated 80 percent of all home loans nationwide (Avery *et al*, 2007) and a higher share of loans originated in metropolitan statistical areas.

Depository institutions that are in the home lending business, have a home or branch office in an MSA, and have assets over a certain threshold (\$35 million in 2006) are required to report to HMDA. Mortgage and consumer finance companies that extend 100 or more home purchase or refinancing loans a year are also required to report for any MSA in which they receive 5 or more applications. We use the following HMDA data fields: Census Tract, Action Date, Loan Amount, Occupancy, Loan Purpose, Originator, High Cost APR, and Lien.

Merging and the Combined Sample

We use HMDA and LP loans originated in 2005. Only loan applications marked as originated, used for home purchase and refinance, and that are marked as one to four family properties are considered. Any HMDA loans marked as sold to Fannie Mae, Ginnie Mae, Freddie Mac, and Farmer Mac are not included since LP only reports loans that were privately securitized. Only loans originated on working days, that is, not weekends and not on holidays (as defined by the Office of Personal Management) are included. Finally, loans with missing purchase/refinance, occupancy, lien, or 5-digit zip code information are dropped.

We clean the originator name in LP to match originators in HMDA, which are likely accurate given that the HMDA data are reported by the originators themselves. We match the most common originator names in LP, comprising approximately 95% of loans with non-missing originator names, to the corresponding HMDA originator name by hand. However, the originator name is missing in LP about 60% of the time. In addition, we classify the originator name as missing in LP if the originator name cannot be hand matched into an originator name that is in the HMDA data.

Finally, we combine Census Tracts in HMDA into Zip Codes to match the geographic data in LP. We cross walk each Census Tract, as defined in the 2000 US Census, to 2006 Zip Codes using the proportion of a Census Tract that is fully contained in the Zip Code.

Appendix 1 describes our merging process in more detail. The data used for our analysis, below, represents the set of unique, one-for-one merges between LP and HMDA.

Interest rates and spreads in LP and HMDA

The presence of interest rate information in LoanPerformance and an APR spread over Treasury rates data in HMDA provides an additional check on our merge and on the key variables in our analysis. HMDA reports the difference between the APR on each loan and the rate on comparable maturity Treasuries for all loans in which this spread is at least 300 basis points (3 percent).³ This APR, which is calculated by the institution reporting the HMDA data, is based on the full cost of the loan, including both interest costs and such up-front charges as points and fees, amortized over the full loan term.⁴ Future rate adjustments for all of the loans we analyze are tied to the 6-month LIBOR, plus the margin, and in every case this expected future rate will exceed the initial rate assuming no change in the 6-month LIBOR rate since origination. Thus, if we have correctly merged the same loan across the two data files, the initial interest rate provided in LP should *never* exceed the APR that is reported in the HMDA data.

³ See Avery *et al* (2006) page A126. The HMDA measure of spread is censored for loans with APRs that are less than 3 percent above the Treasury rate.

⁴ For details on calculating an APR see http://www.efunda.com/formulae/finance/apr_calculator.cfm

For each loan in our data, we calculated the spread between the initial interest rate as reported in LP and the comparable maturity Treasury (the same used to calculate the APR spread). Comparing this LP spread to the HMDA spread provides us with a data consistency check on our merge between the two data files. Table 1 reports the various combinations of spreads calculated from LP versus those obtained from the HMDA data. Start with mortgages in the southwest corner of the table. For these mortgages the HMDA APR spread is censored indicating that it was less than three percent. However, the LP spread – defined as the difference between the initial rate and the comparable Treasury - exceeded three percent. These mortgages fail this consistency check. Move now to the southeast corner. In this cell, both HMDA and LP indicate spreads of at least three percent. We divide this cell into two groups: those where the HMDA spread exceeds the LP spread (RHS), and those where the HMDA spread is less than the LP spread (LHS). Those mortgages in the LHS also fail this consistency check. For the cell in the northeast corner, all of the mortgages have a HMDA spread that equals or exceeds the LP spread so they pass the consistency check. Finally, for the mortgages in the northwest corner cell, the HMDA spread is censored and the LP spread is also less than three percent. We are not able to apply our consistency check on these mortgages (that is to verify if the HMDA spread is at least as large as the LP spread) since the HMDA spread is censored. For our estimation sample, we delete the mortgages that fail this consistency check.

3. Data Description

Our merged HMDA-LP dataset provides new information on subprime loan characteristics and demographic indicators. Table 2 reports unconditional mean characteristics of owner-occupant borrowers who took first-lien subprime hybrid 2/28 loans during August 2005. Several features of the data merit discussion. First, comparison of the “Female” and “Overall” columns indicates that loans with a female primary applicant are generally quite similar to those with a male primary applicant. We do, however, observe some differences by race and ethnicity.

The characteristics of the neighborhoods in which the collateral properties are located vary by type of borrower. Black borrowers generally secured their loans with properties located in areas that are more heavily black, experienced lower rates of house price appreciation during 2004, and had a lower rate of new housing permits. Collateral for loans made to Asian and Hispanic borrowers tended to be in neighborhoods that, on average, had experienced much higher recent house price appreciation and, for Hispanics, higher permit rates, presumably a reflection of the concentration of these borrowers in high-growth areas like Florida and California where house prices also experienced some of the fastest appreciation rates. Unemployment rates in the counties where all borrowers lived were all around 5%, although they were slightly higher for black and Hispanic borrowers.

The typical first-lien mortgage in our data was for a principal amount of about \$217,000, but the average loan made to black borrowers was about 12% smaller, while those made to Hispanic, and especially Asian, borrowers tended to be larger. Origination LTVs for these borrowers were broadly similar, although Asian borrowers were more likely to have LTVs of 95 or more, implying that the value of the house was the main source of this variation.

A substantial majority of the loans made to both Asian and Hispanic borrowers were for the purchase of a new property, as opposed to a refinance, and these borrowers provided full documentation less than half the time. Black borrowers provided full documentation nearly 70% of the time.

While debt-to-income ratios were broadly similar across these demographic groups, black borrowers had lower credit scores while Asian and Hispanic borrowers' scores tended to exceed the overall average. Black borrowers were more likely to have very low FICO scores (below 560) and less likely to have very high scores (above 719). In light of this, it is perhaps unsurprising that initial interest rates received by black borrowers averaged approximately 20 basis points higher than the overall average, while Asian and Hispanic borrowers' rates were 25-36 basis points lower than average. Margins for 2/28 loans, which is the amount added to the 6-month LIBOR rate to determine the adjustable rate in the future, follow a similar pattern, although the differences are smaller than they are for origination rates.

While these unconditional differences are of interest, they are very difficult to interpret on their own. The interest rate and margin charged on a given loan should be functions of the loan terms, the borrower's characteristics and the collateral property and its location. We next turn to a multivariate analysis of the determinants of the initial interest rate and reset margin.

4. Empirical specification and results

We investigate the pricing of subprime fixed-rate and adjustable-rate mortgages using the following regression specification.

$$r_{ijk} = X_i\beta_1 + Y_i\beta_2 + Z_j\beta_3 + \alpha_k + \varepsilon_{ij}$$

The dependent variables are the initial interest rate and the reset margin. The interest rate and margin always refer to the 1st-lien mortgage. While we know if a 2nd-lien exists, we do not know the rate on the 2nd-lien mortgage. In addition, neither LP nor HMDA report any up-front points that may be paid by the borrower. As a result, we have an incomplete picture of the full price of the mortgage(s).⁵ Our sample includes rates only for mortgages that were approved, and we do not control for any variation in denial rates across different locations. The vector X_i contains a set of indicators for the characteristics of the i^{th} borrower. We focus on two racial indicators (Asian and black), one ethnic indicator (Hispanic) and an indicator for the gender of the primary applicant.

The vector Y_i captures the risk profile of the i^{th} borrower. The basic risk characteristics we control for are the borrower's credit score (FICO), the initial combined loan-to-value (LTV) and debt-to-income (DTI) ratios, level of documentation used in the underwriting, whether the mortgage is for a purchase or a refinance, the loan amount, the presence and duration of a prepayment penalty, the type of property used as collateral and the loan type. We follow Haughwout *et al* (2008) in allowing the FICO, LTV and DTI variables to have nonlinear pricing effects. We include indicators for different intervals for each variable. In cases where a 2nd-lien

⁵ In future work we will try to back out the points and fees by comparing the reported APR from HMDA with a zero points/fees APR calculated from LP.

is present, the pricing may differ between the 1st-lien and 2nd-lien mortgages. Ideally, we would like an average interest rate weighted by the relative loan amounts. However, lacking information on the interest rate for the 2nd-lien, we interact the LTV and DTI variables for the presence of a 2nd-lien. The coefficients on the LTV and DTI variables, then, refer to cases where only a 1st-lien mortgage is present, and the coefficients on the 2nd-lien interactions show the degree to which differential pricing exists between mortgages with and without a 2nd-lien.

Finally, vector Z_j contains controls for the characteristics of the geographic area where the house is located. We follow Mayer and Pence (2007) and control at the Zip Code level for the average credit score (Vantage), percent Asian, percent black, percent Hispanic, and the homeownership rate. At the MSA level we control for the extent of house price appreciation over the prior year, house price risk, the unemployment rate and the degree of new building activity as proxied by the issuance of building permits relative to existing housing units. The α_k represent any location error components that remain after controlling for the observed borrower, mortgage and neighborhood characteristics. We report specifications that include MSA and Zip Code fixed effects to sweep out the α_k .

Summary statistics are provided in Appendix Tables A1. The results for pricing the initial interest rate are provided in Table 3. Specification (1) includes only borrower characteristics (X_i). Specification (2) adds controls for the risk profile of the mortgage (Y_i). Specification (3) adds controls for the neighborhood characteristics (Z_j). Finally, specification (4) checks for robustness by adding MSA fixed-effects, while specification (5) replaces the MSA fixed-effects

with Zip Code fixed-effects.⁶ We follow the same format when reporting the results for pricing the reset margin in Table 4.

In our sample of subprime 2/28 mortgages, we find modest differences in the average initial interest rates paid by different groups of borrowers. Interest rates for Asian and Hispanic borrowers on average were 41 and 31 basis points *lower* than for our left-out group of borrowers (specification (1) of Table 3).⁷ In contrast, interest rates for black borrowers were on average 16 basis points higher than our left-out group. In the case of women who are the primary applicant, the data indicate a 5 basis point higher average initial interest rate relative to the left-out group. In all cases, these differences are smaller for the reset margins (specification (1) of Table 4).

These unconditional differences in average interest rates could reflect systematic differences in the risk profiles of the mortgages underwritten for these different groups of borrowers, and/or differences in the characteristics of the geographic locations of these loans which might affect pricing. The extent to which these factors can explain the rate differences can be seen from expanding the estimation to include controls for these factors. We see in specification (2) of Table 3 that controlling for differences in the observed risk profiles of the mortgages significantly reduces the unexplained differences in average initial interest rates across our demographic groups. The 41 basis point lower rate for Asians is eliminated, while the 31 basis point lower rate for Hispanics is reduced to 11 basis points. Similarly, the 16 basis point

⁶ Specifications (2) through (5) contain three property type fixed effects and six loan product type fixed effects. Details are given in the table footnotes.

⁷ The left-out group of borrowers consists of primary applicants who are male, non-Asian, non-black, and non-Hispanic living in a single-family home and who took out a standard 2/28 mortgage.

higher average rate for blacks and the 5 basis point higher average rate for women are both reduced to zero. Controlling for the geographic characteristics in specification (3) of Table 3 further reduces the average rate difference for Hispanics from –11 basis points to –3 basis points. Adding MSA or Zip Code fixed effects has minimal further impact on these results.

The reset margin is a less transparent feature of the mortgage's price than the initial rate. Lenders who wanted to charge specific groups of borrowers different prices that did not correspond to verifiable risk factors might choose to do so with the margin.⁸ The data, however, do not provide any evidence that differential pricing by demographic characteristics of the borrower emerge in the determination of the reset margin. Specifications (2) through (5) of Table 4 show the same pattern that we saw for the initial interest rate. As we control for the characteristics of the mortgage as well as the geographic area the average residual differences in margins for our different types of borrowers become quite small in magnitude.

The results in Tables 3 and 4 indicate similar pricing of 2/28 subprime mortgages in terms of initial rates and reset margins for Asians, blacks, Hispanics and females as for our left-out group of borrowers. We carried out several robustness checks on these results. The results reported in Tables 3 and 4 are based on means of the pricing distributions. It is possible that disparate pricing practices, if they exist, may not be evident at the means, but may only manifest themselves when we look further out in the tails of the rate distributions.⁹ To check for this, we estimated quantile regressions for the initial rate and the reset margin for the 75th

⁸ In auto financing, disparate pricing practices have tended to manifest themselves in the dealer “markup” over the risk-adjusted rates quoted to the dealers by the lending companies. See Cohen (2006).

⁹ For example, Kofi Charles *et al* (2008) find evidence of racial disparities in pricing of auto loans by finance companies at the 75th percentile but not at the median or 25th percentile.

and 25th percentiles. The results from the quantile regressions are broadly similar to those from the mean regressions. The data provide no evidence that disparate pricing by demographic groups occurs for mortgages with high or low residual rates.

A concern might be that any differential pricing faced by women when securing a mortgage may be mitigated if she has a male co-applicant. To check for this possibility, our second robustness check was to restrict the female indicator to those female borrowers with *no* co-applicant on the mortgage. This does not significantly change our earlier findings of no positive residual price differences for women borrowers.

Another possibility is that first-time homebuyers are less skilled at negotiating mortgage rates (see Avery *et al* 2006, Bucks and Pence 2008). If in the subprime mortgage market blacks and/or Hispanics are overrepresented as first-time buyers, then the estimated coefficients on the indicators for these two groups would suffer from a positive left-out-variable bias. Neither the LP nor the HMDA data contain information on whether the applicant is a first-time buyer. However, we can identify a subset of applicants that definitely are *not* first-time buyers – refinances. We interacted the Asian, black, Hispanic and female indicators with an indicator for a refinance. In each case and for both outcome rate measures, the interaction is negative and significant. For Hispanics, the data indicate that the initial rate (reset margin) is on average 14 (9) basis points lower for a refinance as compared to a purchase. However, even purchase mortgages, Asians and Hispanics pay only 3 to 4 basis points higher initial rates and reset margins as compared to white male borrowers.¹⁰

¹⁰ There is no significant difference in the initial rate or in the reset margin on purchase mortgages for black borrowers.

Mayer and Pence (2007) find that subprime mortgage originations were more prevalent in locations with high concentrations of black and Hispanic residents. One possible explanation is that these same neighborhoods were the most credit constrained by the conforming mortgage market, so the development of the subprime market had a differential impact in these areas.¹¹ An alternative possible explanation is that deceptive practices were used to entice borrowers to take out subprime mortgages, and that these practices were relatively more effective in heavily minority neighborhoods. The first explanation is essentially a shift out in the supply of credit, while the second explanation is an induced shift out in the demand for credit.

Mayer and Pence (2007) could not investigate the merits of either of these explanations for their finding since their data did not permit controlling for the race/ethnicity of the individual borrower. While we can not provide a definitive explanation for the Mayer and Pence finding, our data shed some light on the relative merits of the different possible explanations. Specifications (3) and (4) of Tables 3 and 4 report the pricing effects from neighborhoods with higher concentrations of black and Hispanic residents holding constant the race/ethnicity of the primary applicant and the risk profile of the subprime mortgages. The data indicate that increases in the percent black and the percent Hispanic are generally associated, *ceteris paribus*, with lower interest rates and reset margins. Given that Mayer and Pence find positive quantity effects, the negative price effects are consistent with a shift in the supply of mortgage credit. That is, the development of subprime lending may have resulted in a differential expansion of mortgage credit in neighborhoods with high concentration of minorities.

¹¹ See Ladd (1998), Charles and Hurst (2002) and Gabriel and Rosenthal (2005).

Our findings regarding the pricing of risk characteristics of the mortgages are also of interest and we briefly summarize these findings now. Specification (2) of Table 3 shows how the initial interest rate on a 2/28 varies with the characteristics of the mortgage. The results line up well with the findings in Haughwout *et al* (2008) on how these same characteristics affect early default rates. Haughwout *et al* (2008) report that early defaults rise in a nonlinear fashion as the FICO score deteriorates and as the LTV increases, but are relatively insensitive to DTI.¹² These patterns in early default risks are reflected in the upfront pricing based on the mortgage's FICO and LTV. The pricing effects of variation in DTI are inconsistent with intuition, but are small in magnitude compared to the FICO and LTV effects. The interaction between the LTV and an indicator for the presence of a 2nd-lien suggest that the 1st-lien mortgage is priced at a discount to what would be indicated by the combined LTV across both mortgages.¹³ This is true for both the initial interest rate as well as the reset margin.

Turning to the remaining loan characteristics, Haughwout *et al* (2008) report that early defaults are higher for mortgages with limited documentation, and lower for refinances as compared to mortgages for new purchases. This again matches the pattern in pricing of the initial interest rates on 2/28 mortgages. Mortgages with limited documentation are assessed around 50 basis points in higher interest rates, and a higher reset margin of around 40 basis points. Similarly, controlling for observed risk characteristics, interest rates are lower for refinances – both cash-out and no cash-out – as compared to mortgages for new purchases.

¹² Haughwout *et al* (2008) control for the updated LTV which is a function of the initial LTV and the house price appreciation since the mortgage was underwritten.

¹³ A typical example would be that the 1st-lien mortgage has an LTV of 80, and the 2nd-lien mortgage can bring the combined LTV to well in excess of 95. The data indicate that the 1st-lien mortgage in this case is typically priced comparable to an 80 LTV mortgage that does not have a 2nd-lien present.

The magnitudes range from 12 to 31 basis points for the interest rate and reset margin depending on whether geographic controls are included.¹⁴

The final two attributes of the mortgage are the loan balance and the presence of prepayment penalties. For 2/28 mortgages, the interest rate as well as reset margin decline with the size of the loan. Each additional \$10,000 in principal balance reduces the interest rate from around 0.8 to 1.7 basis points. The data indicate that borrowers who are willing to accept a prepayment penalty can reduce the interest rate by 27 to 41 basis points.¹⁵ Mortgages with prepayment penalties will be more attractive to borrowers who expect to keep the mortgage for a longer period of time. Haughwout *et al* (2008), though, find that early default rates are higher for mortgages with prepayment penalties. The optimal pricing for a prepayment penalty, then, depends on the relative tradeoff between lower prepayment risk and higher default risk.

The pricing of 2/28 mortgages is influenced by the house price dynamics in the local housing market. The initial interest rate is lower in markets that experienced a greater degree of house price appreciation over the prior year. If lenders expect these areas to continue to outperform on price appreciation, then the rate of future equity buildup will be higher for these mortgages which could justify the lower initial interest rate. Holding constant the degree of

¹⁴ We do not know if the refinanced mortgage is with the same lender in which case the reduced rate may also reflect the value of an on-going business relationship. Alternatively, the refinance effect may reflect better negotiating skills as discussed earlier.

¹⁵ This assumes that the duration of the prepayment penalty is 24 months – that is, it covers the period up to the first rate reset. The LP data does not contain information on the points involved in the prepayment penalty.

past house price appreciation, increases in the within market dispersion of two year house price changes leads to higher reset margins.¹⁶

One final note is that for 2/28 mortgages the degree of risk-based price differentiation for the initial interest rate tends to be higher than for the reset margin. It is possible that lenders price the reset margin based on the expected risk profile of the mortgage given that it survives the first two years. If the borrower's FICO score improves and house price appreciation reduces the current LTV, then the lender may take this factor into account when setting the reset margin. This would result in smaller coefficients in the margin regression as compared to the initial interest rate regression. Further progress on this issue will require working with the lender identification information.

5. Conclusion

In a sample of over 75 thousand 2/28 subprime mortgages we were able to merge LP and HMDA data together to provide a more detailed picture of loan pricing. Our results provide no evidence of adverse pricing by race, ethnicity or sex of the borrower in either the initial rate or the reset margin. If any pricing differential exists, minority borrowers appear to pay slightly lower rates. We also find that borrowers in Zip Codes with a higher percentage of black or Hispanic residents or a higher unemployment rate actually pay slightly lower mortgage rates.

¹⁶ Our 2-year house price risk is derived from the variance estimate produced by the MSA specific repeat-sale analysis. We have standardized this variable to have a zero mean and unit standard deviation across MSAs.

Mortgage rates are also lower in locations that experienced higher past rates of house price appreciation.

These results suggest appreciable scope for additional research. First, and foremost, it is important to determine whether mortgages originated to minority borrowers had higher upfront costs. In future work, we plan to use the reported APR in HMDA and the initial interest rate, reset margin and interest rate caps reported in LP to infer the upfront points and fees charged on these mortgages. We also plan on using information about the names of the lenders in the HMDA data to consider the role of regulated lenders and also unregulated mortgage brokers in the origination process.

Finally, these results suggest the possibility that subprime lending did serve as a positive supply shock for credit in locations with higher unemployment rates and minority residents. These results are consistent with economies of scale in subprime lending. We believe that further research is needed to understand better how this additional credit impacted these locations. Policy responses today often consider how to limit subprime lending in the future, but it is important to understand any positives that may also have occurred along with the downsides of subprime lending.

References

- Avery, Robert B., Kenneth P. Brevoort, and Glenn B. Canner. 2006. "Higher Priced Home Lending and the 2005 HMDA Data". *Federal Reserve Bulletin*, 92(September): a123-a166.
- Avery, Robert B., Kenneth P. Brevoort, and Glenn B. Canner. 2007. "The 2006 HMDA Data." *Federal Reserve Bulletin*, Vol. 93. A73-A109.
- Bucks, Brian and Karen Pence. 2008 "Do Borrowers Know Their Mortgage Terms?" *Journal of Urban Economics*, 64(2):218-233.
- Canner, Glenn B., Stuart A. Gabriel, and J. Michael Wooley. 1991. "Race, Default Risk and Mortgage Lending: A Study of FHA and Conventional Loan Markets." *Southern Economics Journal*, 58(1):249-262.
- Charles, Kerwin Kofi and Erik Hurst. 2002. "The Transition to Homeownership and the Black-White Wealth Gap." *The Review of Economics and Statistics*, 84(2):281-297.
- Charles, Kerwin Kofi, Erik Hurst, and Melvin Stephens Jr. 2008. "Rates for Vehicle Loans: Race and Loan Source." *The American Economic Review*, 98(2):315-320.
- Cohen, Marc A. 2006. "Imperfect Competition in Auto Lending: Subjective Markup, Racial Disparity, and Class Action Litigation." Vanderbilt Law and Economics Research Paper No. 07-01, December.
- Gabriel, Stuart and Stuart Rosenthal. 2005. "Homeownership in the 1980s and 1990s: Aggregate Trends and Racial Gaps." *Journal of Urban Economics* Vol. 57, no. 1, January.
- Haughwout, Andrew, Richard Peach and Joseph Tracy. 2008 "Juvenile Delinquent Mortgages: Bad Credit or Bad Economy?" *Journal of Urban Economics*, 64: 246-257.
- Ladd, Helen. 1998. "Evidence on Discrimination in Credit Markets." *Journal of Economic Perspectives*, 1(Spring):223-234.
- Mayer, Christopher and Karen Pence. 2007. "Subprime Mortgages: What, Where, and To Whom." NBER Working Paper No. 14083.

Munnell, Alicia H., Geoffrey M. B. Tootell, Lynne E. Browne, and James McEneaney. 1996. "Mortgage Lending in Boston: Interpreting HMDA Data." *The American Economic Review*, 86(1):25-53.

Piskorski, Tomasz and Alexei Tchisty. 2008. "Stochastic House Price Appreciation and Optimal Mortgage Lending," Columbia Business School mimeo.

Ravina, Enrichetta. 2008. "Love & Loans: The Effect of Beauty and Personal Characteristics in Credit Markets." New York University Department of Finance Working Paper Series, July.

Sherlund, Shane. "The Past, Present and Future of Subprime Mortgages." Finance and Economics Discussion Series, no. 2008-63, Federal Reserve Board.

Table 1. Spreads in LP and HMDA – 2/28 ARMs

		HMDA	
		Spread<3	Spread>3
LP	Spread<3	5,349	41,204
	Spread>3	2,571	479^a 29,191 ^b
Total		78,794	
Sample size		75,744	

Notes: Observations in **bold** fail the consistency test and are dropped from the estimation sample.

^a LP spread > HMDA spread.

^b LP spread ≤ HMDA spread.

Table 2. Mean Characteristics by Borrower Race, Ethnicity and Gender

	<i>Primary borrower identified as:</i>				
	Overall	Female	Black	Asian	Hispanic
<i>Observations</i>	75,744	28,489	12,892	2,655	15,647
<i>Borrower characteristics:</i>					
Asian	3.51	3.92	0	100	0.47
Hispanic	20.66	19.19	2.18	2.79	100
Female	37.61	100	51.66	42.11	34.95
<i>Loan characteristics:</i>					
Initial interest rate	7.32	7.37	7.53	6.96	7.06
Margin over 6-month LIBOR	5.88	5.90	5.96	5.76	5.81
FICO	618.95	617.04	606.41	645.01	618.97
Percent < 560	15.57	16.96	19.17	7.72	10.88
Percent > 719	5.16	5.10	2.99	10.88	8.21
LTV	87.43	86.70	88.47	89.40	88.39
Percent < 80	19.64	21.58	17.48	12.84	17.55
Percent ≥ 95	42.74	41.21	45.59	49.11	48.13
DTI ²	40.88	41.62	41.24	41.77	41.64
Percent < 40	54.47	52.08	53.24	45.72	46.92
Percent ≥ 50	7.98	8.47	9.06	7.61	6.65
Percent full documentation	60.81	57.87	68.52	43.46	43.01
Percent purchase	49.59	48.79	51.54	61.66	58.01
Loan amount (\$10,000)	21.69	20.87	19.01	32.88	25.30
Percent with prepayment penalty	74.03	73.04	67.28	75.59	80.98
Months penalty in effect ³	24.04	24.10	23.94	23.23	23.67
<i>Neighborhood characteristics:</i>					
Average credit score	736.46	733.27	706.58	750.62	724.68
Percent Asian	3.84	3.91	3.06	11.00	5.07
Percent black	14.66	17.32	37.79	10.09	11.42
Percent Hispanic	16.42	16.15	11.17	21.14	35.14
Homeownership rate	63.59	62.95	60.49	62.45	59.25
House price appreciation in prior year	10.19	10.12	8.74	14.06	14.58
House price risk ⁴	0	0.012	-0.046	-0.109	-0.002
Unemployment rate	5.12	5.13	5.27	5.07	5.23
Lagged permits in County / 100 units	3.14	2.99	2.45	2.87	3.60

Notes: Subprime 2/28 mortgages – owner-occupants only.

¹ For loans with a minimum spread of three.

² Back-end ratio. Missing for 25.97 percent of loans – generally low and no-doc loans.

³ For loans with prepayment penalties.

⁴ Standardized (two-year) variance of OFHEO index.

Table 3. Initial Interest Rate: Subprime 2/28 Mortgages, Owner-Occupied

	(1)	(2)	(3)	(4)	(5)
<i>Borrower characteristics:</i>					
Asian	-0.413** (0.021)	-0.016 (0.016)	0.027* (0.016)	0.030* (0.016)	0.019 (0.017)
Black	0.157** (0.010)	0.006 (0.008)	-0.026** (0.009)	-0.026** (0.009)	-0.017* (0.009)
Hispanic	-0.315** (0.010)	-0.106** (0.007)	-0.029** (0.008)	-0.026** (0.008)	-0.029** (0.009)
Female	0.055** (0.008)	-0.002 (0.006)	-0.002 (0.006)	-0.004 (0.006)	-0.008 (0.0060)
<i>Loan characteristics:</i>					
FICO: missing		2.376** (0.088)	2.344** (0.087)	2.279** (0.086)	2.213** (0.093)
< 560		1.854** (0.016)	1.838** (0.016)	1.820** (0.016)	1.788** (0.016)
560 – 589		1.113** (0.015)	1.096** (0.015)	1.082** (0.015)	1.058** (0.016)
590 – 619		0.717** (0.015)	0.699** (0.014)	0.689** (0.014)	0.675** (0.015)
620 – 649		0.419** (0.014)	0.406** (0.014)	0.402** (0.014)	0.395** (0.015)
650 – 679		0.232** (0.015)	0.225** (0.015)	0.221** (0.014)	0.215** (0.015)
680 – 719		0.082** (0.016)	0.078** (0.016)	0.078** (0.015)	0.078** (0.016)
LTV: 80 – 84		0.117** (0.010)	0.060** (0.010)	0.035** (0.010)	0.033** (0.010)
85 – 89		0.392** (0.012)	0.326** (0.012)	0.298** (0.012)	0.290** (0.013)
90 – 94		0.568** (0.011)	0.486** (0.011)	0.452** (0.011)	0.449** (0.012)
95+		1.092** (0.012)	0.988** (0.012)	0.932** (0.012)	0.916** (0.013)
95+ • 2 nd -lien		-0.993** (0.012)	-0.961** (0.012)	-0.911** (0.012)	-0.887** (0.012)
DTI: missing		-0.056** (0.008)	-0.056** (0.008)	-0.053** (0.008)	-0.040** (0.008)
40 – 44		-0.028** (0.010)	-0.023** (0.010)	-0.012 (0.010)	0.003 (0.011)
45 – 49		-0.034** (0.010)	-0.025** (0.010)	-0.013 (0.010)	0.002 (0.010)
50+		-0.091** (0.013)	-0.084** (0.013)	-0.064** (0.013)	-0.047** (0.014)

Table 3. Initial Interest Rate: Subprime 2/28 Mortgages, Owner-Occupied – continued

	(1)	(2)	(3)	(4)	(5)
40 – 44 • 2 nd -lien		0.096** (0.016)	0.098** (0.016)	0.097** (0.016)	0.088** (0.017)
45 – 49 • 2 nd -lien		0.061** (0.015)	0.062** (0.015)	0.063** (0.015)	0.056** (0.016)
50+ • 2 nd -lien		0.019 (0.023)	0.018 (0.023)	0.012 (0.023)	-0.003 (0.024)
Limited documentation		0.516** (0.006)	0.532** (0.006)	0.539** (0.006)	0.547** (0.007)
No documentation		0.562** (0.063)	0.585** (0.063)	0.596** (0.062)	0.609** (0.065)
Refinance – cash out		-0.169** (0.007)	-0.148** (0.007)	-0.124** (0.007)	-0.118** (0.008)
Refinance – no cash out		-0.204** (0.013)	-0.211** (0.013)	-0.196** (0.013)	-0.178** (0.014)
Loan amount (\$10,000)		-0.017** (0.000)	-0.011** (0.000)	-0.008** (0.000)	-0.008** (0.000)
Prepayment penalty		-0.220** (0.018)	-0.177** (0.018)	-0.172** (0.018)	-0.198** (0.020)
Months penalty in effect		-0.003** (0.001)	-0.004** (0.001)	-0.008** (0.001)	-0.009** (0.001)
<i>Neighborhood characteristics:</i>					
Average credit score (x10)			-0.017** (0.001)	-0.012** (0.001)	
Percent Asian (x10)			-0.040** (0.006)	-0.006 (0.006)	
Percent black (x10)			-0.007** (0.002)	-0.004 (0.003)	
Percent Hispanic (x10)			-0.028** (0.002)	-0.025** (0.003)	
Homeownership rate (x10)			-0.002 (0.002)	-0.012** (0.003)	
House price appreciation in prior year			-0.014** (0.000)		
House price risk			0.001 (0.003)		
Unemployment rate			-0.010** (0.002)		
Lagged permits in County / 100 units			-0.002** (0.001)		
R-square	0.026	0.458	0.469	0.489	0.572
MSA fixed effects	No	No	No	Yes	No
Zip code fixed effects	No	No	No	No	Yes

Notes: Number of mortgages is 75,744. Specifications (2)-(5) contain three property type fixed effects – condo (8.1%), 2-4 unit (5.2%) and townhouse (0.4%); and six product type fixed effects – 2 yr IO (7.4%), 3 yr IO (0.02%), 5 yr IO (21.2%), 10 yr IO (0.2%), IO unknown period (0.07%) and ARM balloon (5.6%).

** significant at the 5 percent level

* significant at the 10 percent level

Table 4. Margin to 6-month LIBOR: Subprime 2/28 Mortgages, Owner-Occupied

	(1)	(2)	(3)	(4)	(5)
<i>Borrower characteristics:</i>					
Asian	-0.123** (0.020)	0.005 (0.018)	0.019 (0.018)	0.036** (0.018)	0.025 (0.019)
Black	0.071** (0.010)	-0.010 (0.009)	-0.017* (0.010)	-0.015 (0.010)	-0.001 (0.011)
Hispanic	-0.081** (0.009)	-0.057** (0.008)	-0.050** (0.009)	-0.030** (0.009)	-0.030** (0.010)
Female	0.028** (0.007)	-0.008 (0.007)	-0.008 (0.007)	-0.012* (0.006)	-0.010 (0.007)
<i>Loan characteristics:</i>					
FICO: missing		1.279** (0.098)	1.252** (0.097)	1.211** (0.096)	1.101** (0.105)
< 560		1.140** (0.018)	1.131** (0.018)	1.113** (0.017)	1.081** (0.018)
560 – 589		0.701** (0.017)	0.696** (0.017)	0.681** (0.017)	0.659** (0.018)
590 – 619		0.462** (0.016)	0.455** (0.016)	0.443** (0.016)	0.427** (0.017)
620 – 649		0.231** (0.016)	0.226** (0.016)	0.224** (0.016)	0.207** (0.016)
650 – 679		0.118** (0.016)	0.114** (0.016)	0.113** (0.016)	0.097** (0.017)
680 – 719		0.019 (0.017)	0.017 (0.017)	0.024 (0.017)	0.014 (0.018)
LTV: 80 – 84		0.146** (0.011)	0.137** (0.011)	0.126** (0.011)	0.120** (0.012)
85 – 89		0.204** (0.013)	0.192** (0.013)	0.185** (0.013)	0.171** (0.014)
90 – 94		0.290** (0.012)	0.275** (0.012)	0.265** (0.012)	0.255** (0.013)
95+		0.705** (0.013)	0.688** (0.014)	0.660** (0.014)	0.636** (0.015)
95+ • 2 nd -lien		-0.732** (0.013)	-0.725** (0.013)	-0.698** (0.013)	-0.672** (0.014)
DTI: missing		-0.363** (0.009)	-0.363** (0.009)	-0.362** (0.009)	-0.350** (0.009)
40 – 44		-0.029** (0.012)	-0.030** (0.012)	-0.021** (0.011)	-0.016 (0.012)
45 – 49		-0.011 (0.011)	-0.017 (0.011)	-0.009 (0.011)	-0.003 (0.012)
50+		-0.038** (0.015)	-0.039** (0.015)	-0.022 (0.015)	-0.019 (0.016)

Table 4. Margin to 6-month LIBOR: Subprime 2/28 Mortgages, Owner-Occupied – continued

	(1)	(2)	(3)	(4)	(5)
40 – 44 • 2 nd -lien		0.095** (0.018)	0.095** (0.018)	0.101** (0.018)	0.097** (0.019)
45 – 49 • 2 nd -lien		0.105** (0.017)	0.106** (0.017)	0.115** (0.017)	0.108** (0.017)
50+ • 2 nd -lien		-0.036 (0.026)	-0.035 (0.026)	-0.039 (0.025)	-0.042 (0.027)
Limited documentation		0.401** (0.007)	0.398** (0.007)	0.404** (0.007)	0.402** (0.008)
No documentation		0.463** (0.071)	0.459** (0.070)	0.454** (0.069)	0.432** (0.073)
Refinance – cash out		-0.166*** (0.008)	-0.164** (0.008)	-0.137** (0.008)	-0.134** (0.009)
Refinance – no cash out		-0.211** (0.015)	-0.204** (0.015)	-0.186** (0.015)	-0.179** (0.016)
Loan amount (\$10,000)		-0.008** (0.000)	-0.006** (0.000)	-0.005** (0.000)	-0.005** (0.000)
Prepayment penalty		-0.110** (0.020)	-0.113** (0.020)	-0.091** (0.020)	-0.103** (0.022)
Months penalty in effect		0.005** (0.001)	0.005** (0.001)	-0.000 (0.001)	-0.001 (0.001)
Neighborhood characteristics:					
Average credit score (x10)			-0.018** (0.001)	-0.011** (0.002)	
Percent Asian (x10)			-0.015** (0.007)	-0.005 (0.007)	
Percent black (x10)			-0.021** (0.003)	-0.011** (0.003)	
Percent Hispanic (x10)			-0.019** (0.002)	-0.011** (0.003)	
Homeownership rate (x10)			-0.002 (0.003)	-0.003 (0.003)	
House price appreciation in prior year			-0.000 (0.001)		
House price risk			0.041** (0.003)		
Unemployment rate			-0.022** (0.002)		
Lagged permits in County / 100 units			-0.003** (0.001)		
R-square	0.003	0.221	0.226	0.263	0.372
MSA fixed effects	No	No	No	Yes	No
Zip Code fixed effects	No	No	No	No	Yes

Notes: Number of mortgages is 75,744. Specifications (2)-(5) contain three property type fixed effects – condo (8.1%), 2-4 unit (5.2%) and townhouse (0.4%); and six product type fixed effects – 2 yr IO (7.4%), 3 yr IO (0.02%), 5 yr IO (21.2%), 10 yr IO (0.2%), IO unknown period (0.07%) and ARM balloon (5.6%).

** significant at the 5 percent level

* significant at the 10 percent level

Table A1. Summary Statistics: Subprime 2/28 Mortgages, Owner-Occupied

	Mean	Std Dev	Min.	Max.
<i>Borrower characteristics:</i>				
Asian	0.035	0.182	0	1
Black	0.170	0.376	0	1
Hispanic	0.207	0.405	0	1
Female	0.376	0.484	0	1
<i>Loan characteristics:</i>				
Initial interest rate	7.323	1.063	4.25	12.75
Interest rate margin over 6-month LIBOR	5.878	0.988	1.25	11.35
FICO: missing	0.001	0.033	0	1
< 560	0.156	0.362	0	1
560 – 589	0.151	0.358	0	1
590 – 619	0.205	0.404	0	1
620 – 649	0.205	0.404	0	1
650 – 679	0.140	0.347	0	1
680 – 719	0.090	0.287	0	1
LTV: 80 – 84	0.176	0.381	0	1
85 – 89	0.083	0.276	0	1
90 – 94	0.117	0.321	0	1
95+	0.427	0.495	0	1
95+ • 2 nd -lien	0.320	0.467	0	1
DTI: missing	0.260	0.438	0	1
40 – 44	0.172	0.377	0	1
45 – 49	0.204	0.403	0	1
50+	0.080	0.271	0	1
40 – 44 • 2 nd -lien	0.065	0.246	0	1
45 – 49 • 2 nd -lien	0.081	0.273	0	1
50+ • 2 nd -lien	0.024	0.155	0	1
Limited documentation	0.390	0.488	0	1
No documentation	0.002	0.045	0	1
Refinance – cash out	0.450	0.497	0	1
Refinance – no cash out	0.054	0.226	0	1
Loan amount (\$10,000)	21.694	13.436	1.25	154
Prepayment penalty	0.740	0.438	0	1
Months penalty in effect ²	24.036	4.882	5	60
<i>Neighborhood characteristics:</i>				
Average credit score	736.46	44.94	600.9	873.0
Percent Asian	3.84	5.63	0	65.09
Percent black	14.66	20.98	0	98.18
Percent Hispanic	16.42	19.66	0	97.87
Homeownership rate	63.59	15.48	0	99.27
House price appreciation in prior year	10.19	7.98	-1.69	29.11
House price risk (standardized, 2-year)	0	1	-2.54	7.12
Unemployment rate	5.12	1.37	2.3	16
Lagged permits in County / 100 units	3.14	3.07	0	30.07

Notes: Number of mortgages is 75,744.

¹ Conditional on the APR spread being greater than or equal to three.

² Conditional on an existing prepayment penalty.

Appendix 1: Matching LP to HMDA

We match LP into HMDA in multiple stages as described below:

Stage 1

Only those Loans in LP with non-missing originators are considered. LP loans are matched to HMDA loans with the same purpose, occupancy, and lien status. The HMDA loan must be within +/- \$1,000 of the LP loan in order to be considered. For LP loans with non-imputed dates only HMDA loans within +/- 5 working days are considered, for loans with imputed dates HMDA loans within the same month of origination are considered. LP loans are only matched to HMDA loans with the same first 4 digits of the LP loan's zip code. Lastly, if an LP loan matches to multiple HMDA loans a tie breaker is attempted using the subprime variable.

After finding all possible HMDA matches for each LP loan, the LP loans are then classified as non-matches, one-to-one matches, or multiple matches. Any LP loan that has no corresponding HMDA loans using the above criteria is a non-match. Any loan that matches to either multiple HMDA loans or to a HMDA loan that another LP loan also matches to is a multiple match. Lastly, any LP loan that matches to a HMDA, with no other LP loans matching to the given HMDA loan, is a one-to-one match.

After Stage 1 all LP loans classified as one-to-one matches, and their corresponding HMDA loans, are set aside and not considered in future matches. All other loans are then considered in future stages.

Stage 2

Stage 2 is exactly like Stage 1 except that the originator name field is ignored. This means LP loans both with and without originator name information are considered. As well LP loans can match to HMDA loans with any originator name. As with Stage 1, all one-to-one matches are set aside and not considered in future stages.

Stage 3

Stage 3 is exactly like Stage 1 except that the zip code is matched to 5-digits not just 4-digits and origination amount must be exactly the same for the LP and HMDA origination amount. This stage tries to break multiple matches that may have occurred in Stage 1.

Stage 4

Stage 4 is exactly like Stage 2 except that the zip code is matched to 5-digits not 4-digits and origination amount must be exactly the same for the LP and HMDA origination amount. This stage tries to break multiple matches that may have occurred in Stage 2.

Stage 5

Stage 5 is exactly like Stage 1 except that the origination amount is matched to within +/- 2.5% of the LP origination amount.

Stage 6

Stage 6 is exactly like Stage 2 except that the origination amount is matched to within +/- 2.5% of the LP origination amount.

The one-to-one matches from each of the 6 stages above are aggregated into a dataset for further analysis in our paper.