Taking the Lie Out of Liar Loans

Michael LaCour-Little
Finance Department
5133 Mihaylo Hall
California State University
Fullerton, CA 92834-6848
Tel 657-278-4014
Fax 657-278-2161
Email: mlacour-little@fullerton.edu

and

Jing Yang
Finance Department
5149 Mihaylo Hall
California State University
Fullerton, CA 92834-6848
Tel 657-278-2924
Fax 657-278-2161
Email: jyang@fullerton.edu

August 30, 2009

Taking the Lie Out of Liar Loans

Abstract

We examine stated income loans originated by Bear Stearns affiliates during the recent housing market run-up and market collapse. After showing the extent to which these loans have higher default rates than do fully documented loans after controlling for other risk factors, we develop a measure for the extent of likely income over-statement. We then simulate a loan origination process that rejects stated income loan applications with high degrees of likely over-statement and calculate the reduction in default rates that might have been achieved had such an algorithm been in place.

1 Introduction

The recent surge in default and foreclosure rates in the recent U.S. residential mortgage has prompted considerable research with declining housing prices and negative equity areas of particular focus[1]. In addition, policymakers and academic researchers have focused on the proliferation of risky lending contracts including adjustable-rate instruments (ARMs), subprime and Alt-A quality credit (see GAO, 2006, Campbell and Cocco, 2003, Pavlov and Wachter, 2006, and LaCour-Little and Yang, 2009). Such loans may increase default risk by inducing greater borrowing when the principal repayment obligations can be deferred (e.g. interest-only or pay-option ARMs) or when future interest rate risk is involved (e.g. ARMs in general). While lax underwriting and misaligned incentives in the mortgage securitization process have been broadly blamed for the market meltdown, the explicit effect of reduced loan documentation has not been widely studied. In contrast, a recent web-search of the term "liar loan" produces 59,000 entries, most of which appear tied to popular press and blogger accounts of the mortgage crisis.

In this paper we examine 218,000 home purchase loans originated during 2000 - 2007 and securitized by Bear Stearns and affiliates, with loan information updated by May 2009.[2] We find that percentage of loans with full documentation dropped significantly from 42.4% to 21.6% over this time period, while those with reduced documentation (including stated-income, stated-assets, no-income, no-asset, or no- ratio, more on these categories later) increased dramatically from 11.5% to 69.3 %. For loans with self-reported information on key variables (such as income or assets), which we call "stated-doc" loans, almost all are ARMs and 74% are classified as Alt-A. For loans omitting information on traditional underwriting variables, including income, assets, and front-end

[1] As of the first quarter of 2009, a record 9.1% of home mortgages were in either default or the foreclosure process (Inside Mortgage Finance, 2009).

[2] Bear Stearns collapsed and was sold to JP Morgan in March 2008.

and back-end ratios, which we will call "no-doc" loans, almost all are ARMs and 97% are Alt-A. Stated-doc loans were also an important component of the subprime loan segment. According to Inside Mortgage Finance MBS Database, "about 32 percent of subprime mortgages securitized in the first four months of 2007 were originated with stated income, no documentation or so called no ratio underwriting, in which borrower income is not considered."[3] We believe our study here is the first to explore in detail the effects of loan documentation on default risk and analyze the underlying mechanisms that drive these results.

Residential mortgage loan applicants typically encounter varied documentation requirements with subtle differences across lenders and loan programs[4]. The following are some major categories: (1) "full doc", the strictest type under which the borrower must provide proof of income and perhaps income tax returns, W-2's, paycheck with YTD earnings information, verification of employment, and evidence of assets; (2) "lite doc", a common requirement for subprime loans in which the borrower provides bank statements to document income in lieu of paycheck stubs, W-2's, and 1040's, as well as a proof for employment; (3) SIVA (stated income/verified assets) in which borrower income is stated on the loan application, employment is verified, and assets are verified; (4) SISA (stated income/stated assets), where both income and assets are stated but not verified while employment is verified; (5) NORA (no ratio) in which employment and assets are verified, income appears on the application but debt ratios are not calculated; (6) NINA (no income/no asset) in which neither income nor assets are listed on the application but employment is still verified; and (7) "no doc", for which assets, income, and employment are all omitted from the loan application.

Our focus here is the stated-income category (amusingly referred to as "self certified" in the U.K.).

[3] See article "Regulators Keep Pressing for Tougher Standards In Subprime Market; Stated Income Under Fire" Inside Mortgage Finance 24 (21), May 25, 2007.

[4] A "loan program" may reflect a particular investor's underwriting standards so that, for example, an application for a "low doc" loan slated for sale to Freddie or Fannie may be quite different from one slated for sale to a private-label conduit.

Traditionally, these loans were intended for self employed borrowers or those with seasonable income for which income is hard to document or verify. In the U.K. context, lenders reportedly sometimes also find it efficient to "fast track" borrowers with high credit scores and professional occupations with self certification (Cohen, 2009). In addition, lenders may have incentives to encourage brokers to solicit stated-income loans because such loans may produce "excessive rates and penalties" (Harney, 2009). The gist of this argument is that reduced documentation allowed unqualified borrowers to enter the market and/ or allowed those with marginal credit to qualify for unaffordable amounts, contributing to the elevated default rates we now observe.

The academic research on the topic of loan documentation has been limited. LaCour-Little (2007) confirms the traditional relationship posited between self-employment and use of reduced documentation loan programs using single-lender data from 2002; however, subsequent loan performance is not evaluated. Courchane (2007) uses a very large multi-lender dataset of 2004-2005 originations to estimate endogenous switching regressions to examine the effect of demographic and risk factors on loan pricing (as measured by the annual percentage rate or APR). She reports a 16 basis point premium for loans without full documentation in the subprime market segment but a 7 basis point price reduction in the prime loan category. Again, subsequent loan performance is not addressed. Pennington-Cross and Ho (2010, forthcoming) examines performance of both fixed and adjustable rate subprime mortgages using multi-lender data on securitized loans and reports that reduced documentation level is associated with both greater default and greater prepayment risk. The magnitude of the low doc effect is roughly a 40% increase in the marginal probability of early mortgage termination, whether by default or prepayment.

One study that explicitly addresses the issue is Rajan, Seru and Vig (2008), who focus on what they term "hard versus soft information" in the context of asset securitization. If a piece of information

can be readily documented or easily verified, it is hard information. Credit score is a good example (an objective measure that may be obtained at low cost in a matter of seconds). In contrast, a piece of information that is hard to document or verify is soft information, e.g. the risk of future borrower job loss. The authors argue that securitization makes it more difficult for the lenders to collect soft information due to their greater distance from the loan origination process. As a result, their increased reliance on hard information will produce moral hazard in differentiating the qualities of borrower who have the same hard information data but heterogeneous soft information, increasing default risk. Our study is consistent with their finding, showing that under mortgage securitization, lenders may have a tendency to reduce their reliance on even hard information, further weakening screening efficiency, aggravating the default risk.

In this study, we use recently updated (by May, 2009) data on residential mortgage loans originated during 2000-2007 and securitized by Bear Stearns and affiliates[5] to investigate the impact of loan documentation type on default risk, focusing on the stated income category. The data includes over 218,000 observations for which there is complete information on major loan characteristics such as loan balance, monthly payment, LTV, and credit score, which we call the "full sample", and a subsample with over 134,000 loans that have sufficient information from which to infer borrower income, which we call the "restricted sample". With the full sample, we find that stated-income is more widely used with alternative mortgage products and ARMs than other documentation types. Without full-documentation, especially when stated-income or stated-assets programs are used, we find the default rate is significantly higher, even after controlling for other risk factors such as credit score, local housing market affordability, housing cost growth rate, and capital market conditions. Using the restricted subsample we find that, as compared to full-doc loans, stated-doc loans are associated with a significantly higher ratio between inferred borrower income and local

[5] More information on the data is available at www.emcmortgage.com

average income measured at the MSA level. We find the higher the ratio of borrower income to the local average income, the higher the default rate. This pattern is even more pronounced in the stated-doc categories compared to full-doc loans. Results suggest that lenders may be able to reduce default risk by setting limits on the ratio between (verified or stated) borrower incomes and local average income levels. In other words, too large a discrepancy between borrower income and local average income suggests a suspicious transaction in which the borrower (and/or broker) may have exaggerated income to qualify for a loan that is greater than they can really afford.

In summary, our study here contributes to the broader literature on residential mortgage loan performance by providing the first comprehensive study on the effects of loan documentation levels. Moreover, given our analysis we can simulate alternative loan origination policies to quantity the potential extent to which current defaults might have been avoided.

The structure of the paper is as follows. The next section explains the data and the major empirical methodologies we employ. Section 3 presents results of the empirical analysis. Section 4 presents conclusions and planned extensions.

2 Data and Methodology

2.1 Data

Our study relies on data from three sources: (1) loan-level data to identify factors such as loan type, documentation type, borrower credit score, and LTV; (2) MSA-level, for data on local housing market conditions such as housing price levels and the MSA median household income; (3) national-level, for data describing capital market conditions, such as the yield curve.

The loan-level data consists of loans securitized by Bear Stearns and its affiliates during 2000-2007,

restricting the sample to home purchase purpose, single-family dwelling units, with loan terms of thirty years. After deleting observations with missing data, we create a dataset of 218,589 loans, henceforth the "full sample". As income per se is not an available data element, we infer it from loan payment (PITI) and front-end ratio. We hypothesize that likely "income exaggeration" can be measured by the ratio of the inferred borrower income to MSA median household income. Of course, for the "no ratio" and "no doc" categories there is no front-end ratio from which to infer borrower income, so we discard these observations. There are also a small number of loans for which the ratio of borrower to MSA median household income is extremely high and view these as outliers that may be suspect. We initially delete those with a ratio greater than 10, but will come back to these later in the paper. The final subsample consists of 134,174 loans or about 61% of the full sample. We call this the "restricted sample". We calculate loan age (in months) as of April 2008 for non-defaulted loans and based on the date the loan was referred to a foreclosure attorney for defaulting loans. Thus we observe loan age at default or the point of data censoring.

Documentation level takes various forms in our data. We begin by classifying loans into three broad categories: (1) "full-doc", for loans with "full" marks in documentation type descriptions; (2) "stated-doc", for loans with "stated" marks in documentation type descriptions, including loans with stated-income, stated-income/stated-assets, and stated-income but verified assets; and (3) "no-doc", for loans with "no" marks in documentation type description, including loans with no-income, no-assets, no-ratio, and no-documentation at all. In the full sample, these three categories comprise 30%, 44% and 18% of the loans, respectively. In the restricted sample, the full-doc and stated-doc comprise 37.5% and 54.2% of the loans, respectively, and given the limited count of no-doc loans, our major comparison within the restricted sample data is between the stated-doc loans and the full-doc loans. Documentation type was missing in 8% of the loans in the full sample,

and about 8% of the loans in the restricted sample, so we discard these observations.

We also include MSA-level variables to control for market-specific factors. We include local housing price levels, which we measure with the publicly available OFHEO HPI; the 5-year average annual growth rate in MSA HPI; the MSA-median household income; local housing affordability, measured by the ratio between the MSA-median household income and the concurrent MSA HPI; and so forth. Finally, we include several capital market condition indicators as additional control variables, including the slope of the yield curve, the return on equity markets, and the level of mortgage rates, all measured as of the date of loan origination. The slope of the yield curve is calculated as the ratio of the 10-year Treasury bond rate and the 2-year Treasury note rate. The return on equity markets is measured by the 1-year return of S&P500 index. The level of mortgage rate is measured by the contract rate on 30-year, fixed-rate conventional home mortgage commitments, based on the Freddie Mac Primary Mortgage Market Survey data.

2.2 Methodology

The major questions we address are the following: (1) Does the lack of full documentation create additional default risk? (2) Is there evidence of "income exaggeration" among stated-documentation borrowers? (3) Does "income exaggeration" among stated-doc loan borrowers further increase default risk?

2.2.1 Issue one: does the lack of full documentation create additional default risk?

We develop the following three specifications of logit regressions of loan default[6], to explore this issue using the full sample data:

$$D_{default} = \alpha_f + \beta_f \ D_{fulldoc} + \sum_{j=1}^{v} \gamma_{fj} V_j + \vartheta_f, \quad (1)$$

$$D_{default} = \alpha_s + \beta_s \ D_{stateddoc} + \sum_{j=1}^{v} \gamma_{sj} V_j + \vartheta_s, \quad (2)$$

$$D_{default} = \alpha_n + \beta_n \ D_{nodoc} + \sum_{j=1}^{v} \gamma_{nj} V_j + \vartheta_n. \quad (3)$$

$D_{default}$ is the default dummy, which takes on the value of 1 if the data shows that the loan has defaulted; $D_{fulldoc}$, $D_{stateddoc}$ and D_{nodoc} are the dummies for full-doc, stated-doc and no-doc types, respectively; set V contains control variables that are expected to also affect default probability, including loan-level factors such as credit score, MSA-level factors such as local housing market affordability, and capital market factors such as 1-year return in S&P 500 index; α_f, α_s and α_n are the intercepts; β_f, β_s and β_n are the coefficients for the documentation type dummies; γ_{fj}, γ_{sj} and γ_{nj} $(j = 1, ..., v)$ are coefficients for control variables; and finally, ϑ_f, ϑ_s and ϑ_n are the error terms. We will test the following hypothesis:

- [Hypothesis 1] Stated-doc and no-doc loans are more likely to default than full-doc loans. In regressions 1, 2 and 3, this means that the stated-doc dummy and the no-doc dummy positively affect the default dummy, while the full-doc dummy negatively affects the default dummy, that is, $\beta_f < 0$, $\beta_s > 0$ and $\beta_n > 0$.

[6] We measure default by the indicator variable "Referred to foreclosure attorney" contained in the data. There is also a field indicating the date the loan was referred to the foreclosure attorney, so we can determine loan age at time of default. Other authors have used the first instance of a 90-day delinquency, the occurrence of the filing of a notice of default, or similar measures intended to capture serious loan delinquency and pending foreclosure. None of these definitions implies that the loan actually proceeds to a foreclosure sale, of course, as the borrower may always reinstate the loan, pay off the loan, and/or sell the property prior to auction date. Capturing those outcomes in detail is important for measurement of loss severity, as opposed to default rates, which is our focus here.

2.2.2 Issue two: is there evidence of "income exaggeration" by stated-documentation borrowers?

We develop the following simple regressions, to explore this issue using the restricted sample data:

$$Incratio = \kappa_f + \rho_f \; D_{fulldoc} + \varepsilon_f, \tag{4}$$

$$Incratio = \kappa_s + \rho_s \; D_{stateddoc} + \varepsilon_s, \tag{5}$$

where $Incratio$ is the inferred-MSA median income ratio, that is, the ratio between inferred borrower income and the MSA median household income; $D_{fulldoc}$ and $D_{stateddoc}$ are the dummies for full-doc and stated-doc, respectively; κ_f and κ_s are the intercepts; ρ_f and ρ_s are the coefficients for the documentation type dummies; and finally, ε_f and ε_s are the error terms. We will test the following hypothesis:

- [Hypothesis 2] Stated-doc loans have higher inferred-MSA median income ratio than full-doc loans. In regressions 4 and 5, this means that the income ratio is positively affected by the stated-doc dummy while negatively affected by the full-doc dummy, that is, $\rho_f < 0$ and $\rho_s > 0$.

2.2.3 Issue three: does "income exaggeration" among stated-doc borrowers further increase default risk?

To explore this issue, we first develop the following regression for the full-doc subsample and the stated-doc subsample in the restricted sample:

$$D_{default} = \pi + \mu \; Incratio + \sum_{j=1}^{k} \varpi_j S_j + \xi, \tag{6}$$

where $D_{default}$ is the default dummy; $Incratio$ is the inferred-MSA median income ratio; π is the intercept; set S contains control variables that are expected to also affect default probability; μ is

the coefficient for $Incratio$; ϖ_j $(j = 1, ..., \upsilon)$ are the coefficient for control variables; and finally, ξ is the error terms. We will test the following hypothesis:

- [Hypothesis 3] The default rate is increasing in the inferred-MSA median income ratio for the stated-doc subsample, with a stronger effect than that for the full-doc subsample. In regression 6, this means that with the stated-doc subsample data, $\mu > 0$, and in addition, μ is positive and larger in magnitude than when estimated using the full-doc subsample data.

We further explore the effect of income ratio on stated-doc loans' default risk by conducting a "income ratio cap" sensitivity analysis, using the restricted sample. We analyze the default rate of stated-doc loans versus that of full-doc loans, setting different upper boundaries on the income ratio. We will call these "income ratio caps". For a specific cap $\widehat{H_i}$ $(i = 1, ..., n)$, we extract a subsample with exclusively stated-doc loans, and a subsample with exclusively full-doc loans. With n different caps, we could get n cap-specific stated–doc-loan subsamples, and n cap-specific full–doc-loan subsamples. We calculate the cap-corresponding mean default rate for loans with each documentation type, and run the following regressions for stated-doc loans and also for full-doc loans:

$$\overline{D_{default}} = \phi + \theta \ \widehat{H} + \delta. \tag{7}$$

$\overline{D_{default}}$ is the cap-specific mean default rate; $\widehat{H}$ is the inferred-MSA median income ratio cap; ϕ is the intercept; θ is the coefficient; and δ is the error term. We will test the following hypothesis:

- [Hypothesis 4] A more restrictive policy on income exaggeration (a lower income ratio cap) will reduce stated-doc loan default risk, and the effect is stronger among stated-doc loans than for full-doc loans. In regression 7, this means that the average default rate will be positively affected by the income ratio cap for the stated-doc loans, and the effect is stronger than for

full-doc loans, that is, $\theta > 0$ for the stated-doc subsample, and θ is more positive and larger in magnitude for the stated-doc subsample than for the full-doc subsample.

Testing these four hypotheses will be the main focus of our empirical analysis. Along the way, we explore related issues such as the relationship between doc type and other loan characteristics, interactions between product type and doc type, and variation with housing market and capital market conditions. The following section presents results.

3 Empirical Results

3.1 Descriptive Statistic Results

3.1.1 Time trend

We first examine time trends in the data for loans originated during 2000-2007. Figure 1 illustrates the time trend for the loans in the full sample. Full-doc type remained the dominant category with a market share (in terms of loan number) consistently above 40% (peaking with a 60.2% in 2004), until 2005 (roughly the peak of the housing bubble). After 2005, its market share declined sharply. By 2007, its market share was only 21.6%, lower than the shares of other two types (46.0% for stated-doc and 23.3% for no-doc). In contrast, stated-doc loans were very rare before 2002, but grew at an accelerating pace becoming the dominant category in 2005 with over a 40% market share. No-doc type also experienced fast growth since 2005 and was the second largest category by 2007. These results suggest a regime shift somewhere around the year of 2005, concurrent with the rise of alternative mortgage products (AMPS). The interaction between these two patterns will be addressed later.

Figure 1 illustrates that 2005 is also a turning point for loan origination where we observe that

stated-doc loans had become the riskiest category with mean default rates exceeding those of the other two documentation types. Among loans originated in 2000, the full-doc, stated-doc and no-doc default rates (by May, 2009) are 37.3%, 0.0% and 14.7%, respectively, while among loans originated in 2007, these rates were 6.0%, 10.7% and 10.4%, respectively. This suggests that the rapid growth in the stated-doc type increased risk, consistent with Hypothesis 1.

The increase in the relative riskiness of stated-doc loans cannot be explained by credit score and original LTV alone, the two traditional risk factors associated with default. As illustrated in Figure 1, stated-doc loans originated in 2000 had mean FICO scores of 739, as compared to 593 for full-doc borrowers, and 631 for no-doc borrowers. But this pattern changed dramatically over time. For loans originated in 2003, mean FICO scores had converged to 638 versus 645 and 650. From 2003 to 2005, the credit scores of all-type loan borrowers increased consistent with lender reliance on hard information, as in Rajan, Seru and Vig (2008). Since then, the credit scores of stated-doc and no-doc borrowers have stayed rather stable while the scores of full-doc borrowers have declined slightly.

In terms of original LTV, the stated-doc loans used to be the "safest" loans, with mean original LTV of only 56.5%, as compared to 76.7% (full-doc) and 73.1% (no-doc), for loans originated in 2000. For loans originated in 2007, however, full-doc, stated-doc and no-doc had mean original LTVs of 75.3%, 73.1% and 65.2%, respectively, a substantial increase for the stated-doc loans.

Given these patterns, how can we explain the higher default rates of stated-doc loans? We address this question with the restricted sample. Figure 2 shows that on average, loans in the restricted sample are similar as those in the full sample: as compared to full-doc loans, stated-doc loans migrated from lower- to higher-default risk loans, which cannot be explained by credit scores and original LTV ratios. However, stated-doc loans do show higher ratios between borrower income and

MSA median household income than do full-doc loans since 2001, a pattern consistent with income exaggeration. Income exaggeration may shift lending to less qualified borrowers. This provides some support for the implication of Hypothesis 2 that stated-doc loans are associated with income exaggeration as well as Hypothesis 3 that income exaggeration may by the reason that stated-doc loans become the riskiest category over time.

3.1.2 Descriptive statistics

We next present descriptive statistics on loan characteristics by documentation level for loans originated during the entire origination period 2000-2007.

Table 1 shows results for the full sample. Stated doc is the most common category (44%), followed by full-doc (30%) and then no-doc (18%). The distribution of doc type is further illustrated in Figure 3. Most of the loans are ARMs, while FRMs comprise less than one percent. The majority (63%) are ALT-A, a category which overlaps with AMPS, which comprise 43% of the loan population. About one-third of all loans are subprime.

Cross-documentation type comparisons are highlighted in Table 2. In general, doc type seems to interact with LTV, credit score and loan type, affecting default rates. While no-doc loans have, on average, the highest credit scores, the lowest LTV, the lowest subprime percentage, the highest ALT-A and AMPS percentages, the most recent origination years, and relatively larger loan size. This pattern suggests higher underwriting standards for such loans, consistent with their lower mean default rates. Likewise, full-doc loans have the lowest borrower credit score, the highest LTV, the highest subprime percentage, the lowest ALT-A and AMPS percentages, tend to be older, and smaller in loan size. This suggests underwriting standards that require lower quality loans to provide more documentation. This category has the lowest mean default rate by May,

2009. Finally, stated-doc loans have features in between the other two categories, except that the loan size and the default rate are the largest across the three groups. This patterns suggests that lenders were able to successfully screen high and low-quality borrowers but may be relatively less effective in underwriting those in the middle-range, particularly since in that category borrowers could misrepresent their income.

Turning to the restricted sample, we report additional information including inferred-MSA median income ratio, local housing market conditions such as housing cost and affordability, as well as capital market conditions. Table 3 shows that, with the restricted sample as a whole, stated-doc is still a much more frequent category (54%) than full-doc (36%). In general, inferred income is a little lower than the local area median, so the median inferred-MSA median income ratio lower is 0.81, although the mean is a little bit higher at 1.10.

Table 4 compares stated-doc with full-doc loans. Across the two groups, stated-doc loans have much higher mean inferred-income to MSA median ratio than do full-doc loans, 1.20 versus 0.94, with the difference significant at 1% level. As was the case with the full sample, stated-doc loans also have a significantly higher default rate than do full-doc loans. The income ratio difference supports Hypothesis 2, and the coexistence of an income ratio difference and a default risk difference are consistent with Hypothesis 3. Furthermore, Table 4 also shows that stated-doc loans are more concentrated in areas with higher and more rapidly increasing housing costs and areas with lower affordability, and become more frequently used when stock market returns were higher or yield curve was flatter, that is, when the capital market is less constrained.

3.1.3 Statistics by income ratio range

Since findings thus far with respect to the inferred income-MSA median are interesting, we further explore loan characteristics and the doc type effects by examining loans in each range of this income ratio. To explore the issue more fully, we enlarge the restricted sample to include those loans with inferred-MSA median income ratio >10, so the sample size increases from 134,174 loans to 135,119 loans.

More than 95% of the loans have income ratios lower than 3.5, which we will then use as the first cutoff in income ratio range analysis. Table 5 and Figure 4 show the statistics for the two income ratio ranges: [0, 3.5) and [3.5, ∞). We focus on the three most important variables, default rate, original LTV, and credit score. With respect to default risk, from income ratio range [0,3.5) to range [3.5, ∞), full-doc loans' mean default rate has a modest increase (from 15.93% to 17.99%), while the stated-doc loans' mean default rate jumps more dramatically (from 20.89% to 29.95%). With respect to original LTV, the two groups' mean LTV ratios not only increase but do so roughly proportionately (62.18% to 72.43% for full-doc loans, and 61.31% to 73.16% for the stated-doc loans). This suggests that LTV may not be a reason for the cross-doc type variation in default rate. Finally, with respect to credit score, full-doc loans' average credit score has a noticeable increase from 659 to 688, while mean credit score for stated-doc loans' shows a smaller increase from 683 to 699. This suggests that the association between the stated-doc loan default risk and the inferred income-MSA median ratio may be due to income exaggeration by loan applicants.

We further investigate these patterns by dividing the sample into ten income ratio ranges. Results are shown in Table 6 and Figure 5. Visual inspection suggests that an income ratio 1.5 is an important inflection point. Above that point, default risk will increase dramatically especially for stated-doc loans; and in addition, original LTV also increases dramatically. In contrast, credit

score is stable, until the income ratio exceeds 2.5. This divergence between again suggests that the relationship between stated-doc default risk and income ratio may be due to income exaggeration.

In summary, our analysis suggests that by setting maximum allowable income ratio ranges, lenders might have been able to reduce default risk among stated income borrowers. When a borrower declares usually high income (relative to the local average income level) they could be either rejected for the stated-doc loan or required to switch to a full-doc loan, potentially requiring a smaller loan amount. Anecdotally, some lenders employed analogous techniques by comparing stated income to average reported income within the borrower's occupation. An interesting extension would be to compare default rates across stated-income lenders where some employed such a screening mechanism and others did not. We do not have this degree of detail available with our current data set.

3.2 Regression results

3.2.1 Documentation type regressions

We begin by investigating which loan types are more likely to have certain documentation types. Using the full sample, we run a logit regression for each doc type dummy variable on major loan characteristics including original LTV, credit score, whether it is a FRM and whether it is an AMP. We employ two model specifications to control for multi-co linearity among explanatory variables. As shown in Table 7, a loan is more likely to be stated-doc when it is ARM and/or AMPS, as the stated-doc logit regression has a large and negative coefficient for the FRM dummy (-2.406), and the largest positive coefficient for the AMPS dummy (0.611, as versus -0.650 with full-doc, and 0.536 with no-doc). In addition, a relatively higher LTV combined with a relatively higher credit score is also related to use of stated-doc. These results suggest that stated-doc loans are

used by borrowers with moderate-level LTVs and credit scores, but widely used in conjunction with aggressive lending instruments such as ARMs and AMPS.

3.2.2 Default regressions (for Hypothesis 1)

Using the full sample, we run a default probability regressions 1, 2 and 3 to test Hypothesis 1 that stated-doc and no-doc loans are more likely to default than full-doc loans. Comparing the regressions, we will examine whether the stated-doc dummy and the no-doc dummy positively affect the default dummy, while the full-doc dummy negatively affects the default dummy, that is, $\beta_f < 0$, $\beta_s > 0$ and $\beta_n > 0$. We develop four model specifications to explore the issue, with the results shown in Table 8. In Specification 1, after controlling multi-co linearity, we include original LTV, credit score, documentation type dummies and loan origination year dummies as explanatory variables. The results are in consistent with Hypothesis 1, with the full-doc dummy coefficient β_f negative (-0.527), and the stated-doc dummy coefficient β_s and the no-doc dummy coefficient both positive (0.352 and 0.139). Interestingly, the coefficient on stated-doc is larger than that on no-doc. The full-doc coefficient is negative, confirming that greater loan documentation reduces default risk. With respect to other factors, we find that as expected, default rates are increasing in original LTV and decreasing in credit score. In Specification 2, we add more risk factors such as loan balance, a dummy variable for FRM and AMP contract type. Previous results continue to hold, with β_f negative, while β_s and β_n both positive and β_s larger in magnitude than β_n. Collectively results strongly support Hypothesis 1: lack of full documentation boosts default risk. As a robustness test, we replace the original year dummies with a loan age dummy, forming Specifications 3 and 4, which generate essentially similar results as the previous two specifications.

Given these results, we can simulate default rates. For instance, at the sample mean of borrower characteristics (original LTV of 63%, credit score of 676, and loan age of 29 months and using

Specification 3 coefficients) we find that the average loan will have a 11.10% default rate if fully documented, while 18.35% if stated documented, or 17.83% if without documentation. This indicates a 65% increase in default rate by switching from full-doc to stated-doc, while a 61% increase in default rate by switching from full-doc to no-doc. A similar simulation using coefficients from Specification 4 indicates a 56% (full-to-stated) increase and a 46% (full-to-no) increase in default rate. Note these results are slightly higher than the estimates in Pennington-Cross and Ho (2010) discussed previously.

3.2.3 Inferred-MSA median income ratio regression (for Hypothesis 2)

We then use the restricted sample to run the inferred income-MSA median ratio regressions 4 and 5, to test Hypothesis 2 that stated-doc loans have higher ratio between borrower income and local average income compared to full-doc loans. Across the two regressions we will examine whether the stated-doc dummy positively affects the income ratio while the full-doc dummy negatively affects the income ratio, that is, $\rho_f < 0$ and $\rho_s > 0$. Table 9 reports results, with the dependent variable defined as the natural logarithm of inferred income-MSA median income ratio. It shows a significantly positive coefficient for the stated-doc dummy, ρ_s (0.203), while a significantly negative coefficient for the full-doc dummy, ρ_f (-0.242). Results confirm results in Table 4 that stated-doc loans are used by ostensibly higher income households (compared to full-doc loans); however, that higher income may be exaggerated.

3.2.4 Cross-subsample difference in default regression (for Hypothesis 3)

From the restricted sample, we can extract subsamples of full-doc and stated-doc loans. Using these subsamples we run regressions 6 to test Hypothesis 3 that default rate is increasing in the inferred-MSA median income ratio for the stated-doc subsample, but that any such effect is not

as large in the full-doc subsample. Results appear in Table 10. Again we include a comprehensive set of risk factors including loan-level characteristics, MSA-level housing market conditions, and national-level capital market factors, and develop five model specifications which control for multi-co linearity differently. Across all specifications results support Hypothesis 3, with the coefficient of the natural logarithm of inferred-MSA median income ratio positive for both stated-doc loan and full-doc loan subsamples, but the magnitude of the coefficient is over 35% greater for stated-doc loans.

Examining other risk factors, compared to full-doc loans, stated-doc loan default rates are more sensitive to housing cost (+), housing price appreciation (+) and affordability (-), while less sensitive to credit score (-), suggesting that relaxing documentation constraints will increase risk during difficult housing market conditions.

Again calculating implied default rates at sample means using Specification 6 (the only one that includes loan age), we find that an average loan will have a 16% default rate if fully documented, while 24% if stated documented, indicating a 51% increase in default rate by switching from full-doc to stated-doc.

3.2.5 Effect of inferred-MSA median income ratio cap (for Hypothesis 4)

Finally and again using the two doc type subsamples, we run a regression 7 to test Hypothesis 4 that more restrictive policy on income exaggeration will reduce stated-doc default risk, and this effect is stronger for stated-doc than for the full-doc loans. In other words, we will examine whether average default rates are negatively affected by the income ratio cap (that is, the upper boundary of the allowed range of inferred income-MSA median ratio) for the stated-doc loans and whether any such effect is stronger than is the case for full-doc loans.

We first create a data set for this regression. For each possible income ratio cap we calculate the loan counts, mean default rate, and mean inferred income-MSA median ratio for the all-loan sample, the full-doc and the stated-doc subsamples. We then vary the cap level, to generate a series of loan counts, mean default rates, and mean income ratios for each sample, which will be used later for running the regression 7. The data thus created are summarized in Table 11. As the income ratio cap increases, all three variables (loan count, mean default rate, and mean income ratio) in general increase faster for the stated-doc subsample than for the full-doc subsample. These effects are further highlighted in Figure 6. Interestingly, when the cap is higher, the default rate difference between the two documentation types becomes more significant. For the stated-doc loans, an income ratio cap of 1.5 will reduce the average default rate from 21.37% to 18.09% (a 18% drop), a cap of 2 will reduce the default rate to 19.70% (a 8% drop), and a cap of 2.5 will reduce to the default rate to 20.32% (a 5% drop). While these absolute levels of default are still unacceptably high, results help illustrate the effect of controlling the reliance on stated income that is likely exaggerated.

We then use these data to run regression 7 for stated-doc and full-doc loans. As shown in Table 12-Specification 2, we confirm that the cross-documentation type difference in the effect of inferred-MSA median income ratio cap on default risk is statistically significant, with the cap coefficient as 0.005 for the stated-doc loans while only 0.001 for the full-doc loans, supporting Hypothesis 4 that applying a restrictive policy on likely income exaggeration will help reduce default risk particularly for stated-doc loans. Similar results are apparent from Specifications 1 and 3.

In summary, regression results provide strong support for our four major hypotheses, confirming the importance of loan documentation discipline and management to reduce default risk.

4 Conclusions

In this paper we have examined the effects of documentation type on default risk. Although loan documentation requirements have changed dramatically in recent years, their contribution to increasing rates of residential mortgage default has not been rigorously analyzed. We believe our study is the first to focus on this issue. We do so using a large database on home purchase loans securitized by Bear Stearns and affiliates over the recent period 2000-2007.

We find that reduced levels of documentation do increase the likelihood of default after controlling for other risk factors. The problem is particularly acute for stated-doc loans, which are offered to lower quality borrowers (as measured by credit score and LTV) compared to no-doc loans, though they are higher in quality than full-doc loans. Simulation based on our default regression models suggests an over 50% increase in default rate when a loan with average characteristics switches from full-doc to stated-doc, a result that is slightly higher than the 40% increase in default rate for subprime loans reported by Pennington-Cross and Ho (2010), though our methods are not exactly comparable. The reason that these mid-quality borrowers perform worse than objectively worse borrowers because lenders allowed them simply state, as opposed to verify, income or assets, while not allowing objectively lower quality borrowers to do so. We find evidence of income exaggeration in the stated doc category and show that the degree of likely exaggeration is related to default risk. We also show that limiting the ratio between stated income and local average household income may reduce default risk. Simulation suggests that when the ratio is limited to 2.5, stated-doc loans should be less risky than full-doc with objectively worse credit score and LTV characteristics. Given these results, we think stated income lending can be a viable mortgage product; however, careful risk management is essential to mitigate inherent risks.

Further research efforts on this topic may involve replicating the current analysis using data through at least year-end 2008, potentially measuring local area income at a finer level of geography and replicating the analysis using multi-lender data, where available.

References

Campbell, John Y. and Joao Cocco. 2003. Household Risk Management and Optimal Mortgage Choice . Quarterly Journal of Economics 118(4): 1449-1494.

Cohen, Norma. 2009. Banks Still Lending at Many Times Earnings. Financial Times, January 31, 2009.

Courchane, Marcia. 2007. The Pricing of Home Mortgage Loans to Minority Borrowers: How Much of the APR Differential Can We Explain? Journal of Real Estate Research 29(4): 399-440.

Inside Mortgage Finance Publications. 2007. Regulators Keep Pressing for Tougher Standards In Subprime Market; Stated Income Under Fire. Inside Mortgage Finance, May 25, 2007.

Inside Mortgage Finance Publications. 2009. Default Rates Still Rising. Inside Mortgage Finance, June 5, 2009.

General Accountability Office. 2005. Alternative Mortgage Products.

Harney, Kenneth. 2009. Congress Takes A Series Look At Reforming the Mortgage Market. Washington Post, April 4, 2009.

LaCour-Little, Michael. 2007. The Home Purchase Mortgage Preferences of Low-and-Moderate Income Households". Real Estate Economics 35(4):265-290.

LaCour-Little, M., and J. Yang. 2009. Pay Me Now or Pay Me Later: Alternative Mortgage Products and the Mortgage Crisis. California State University at Fullerton Working Paper.

Pavlov, A. and S. Wachter. 2006. Aggressive Lending and Real Estate Markets. University of Pennsylvania Working Paper.

Pennington-Cross, Anthony and Giang Ho. 2010. The Termination of Subprime Hybrid and Fixed Rate Mortgages. Real Estate Economics 38(1), forthcoming.

Rajan, Uday, Amit Seru and Vikrant Vig. 2008. The Failure of Models That Predict Failure: Distance, Incentives and Defaults. University of Chicago Working Paper.

Turnbull, Stuart M., Crouhy, Michel and Jarrow, Robert A.,The Subprime Credit Crisis of 07(July 9, 2008). Available at SSRN: http://ssrn.com/abstract=1112467

United States Government Accountability Office. 2006. Alternative Mortgage Products . Report to the Chairman, Subcommittee on Housing and Transportation, Committee on Banking, Housing, and Urban Affairs, U.S. Senate. GAO-06-1021.

Figure 1 Time Trend of Loan Characteristics (Full Sample)

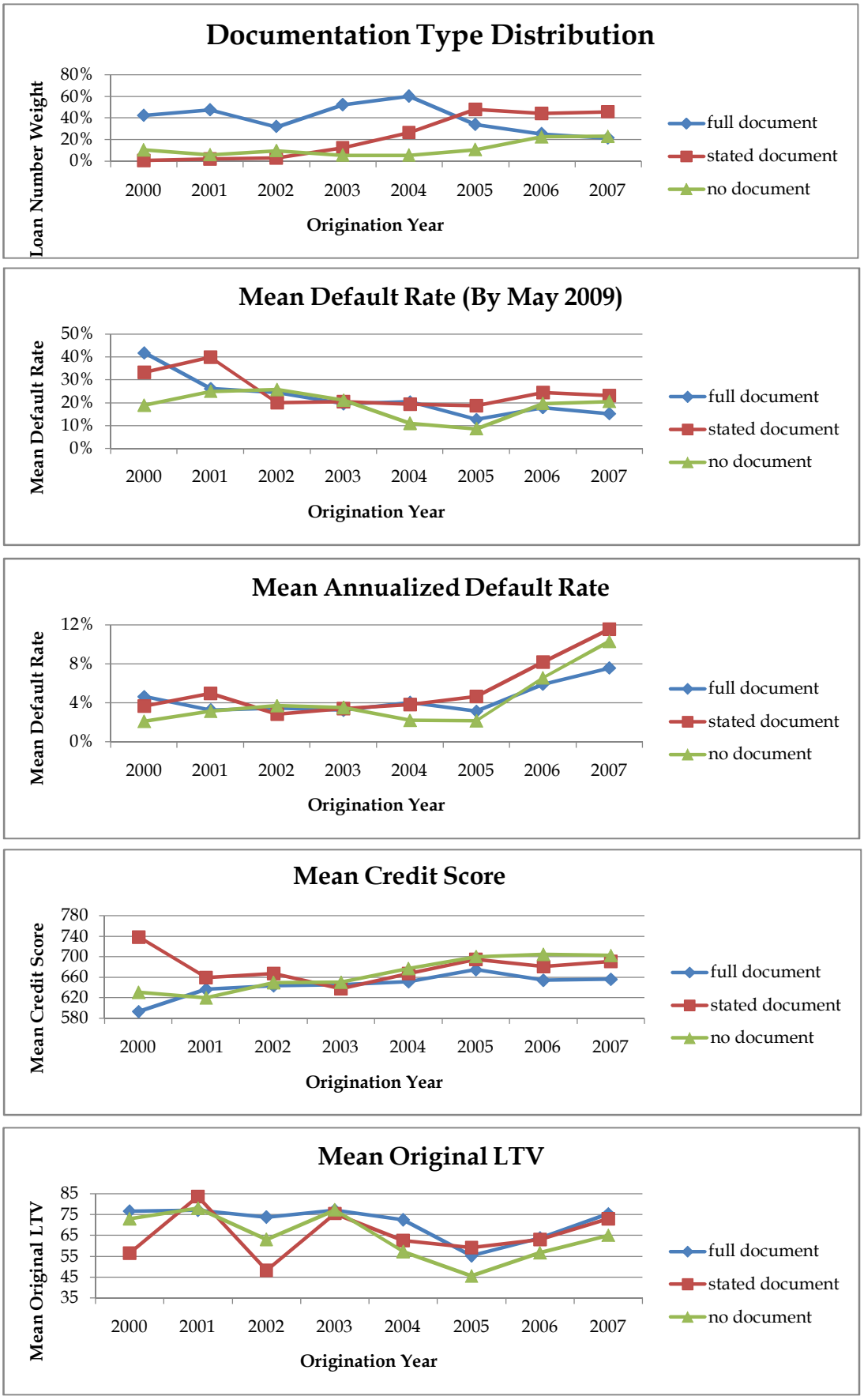
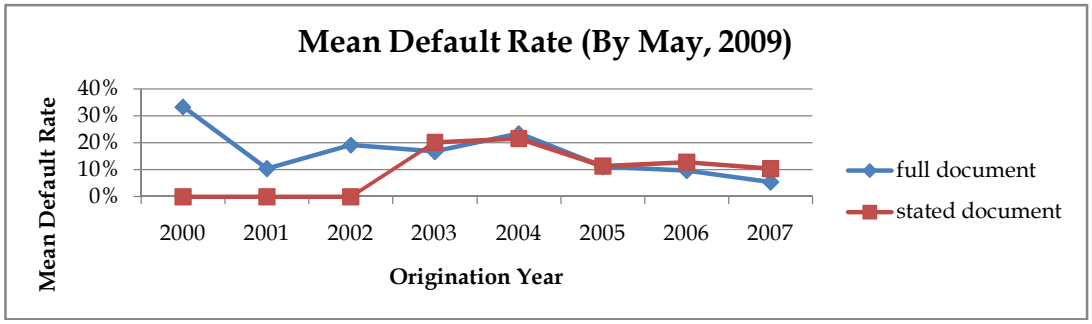
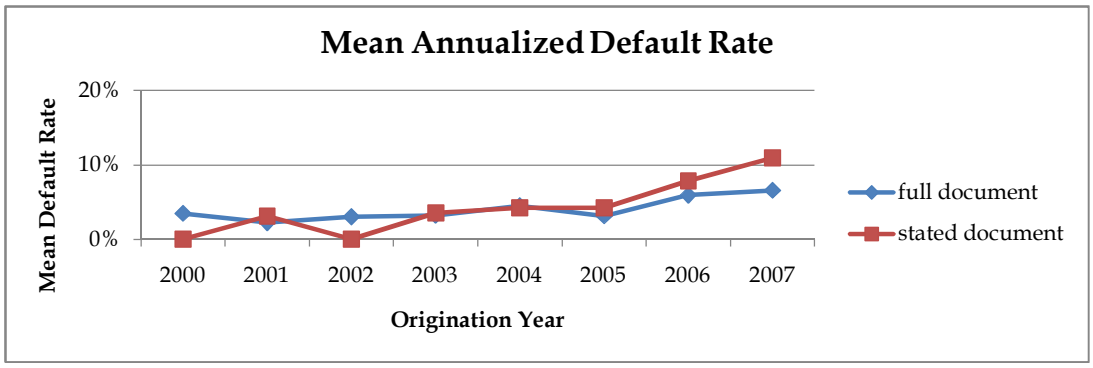
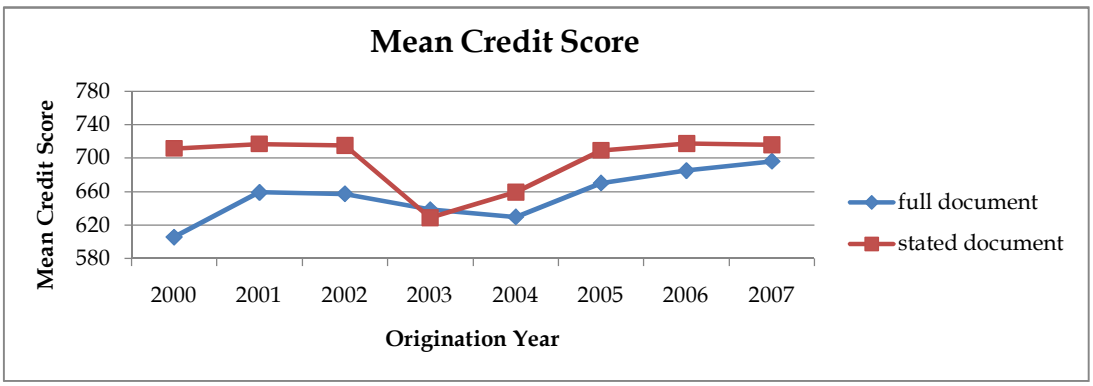
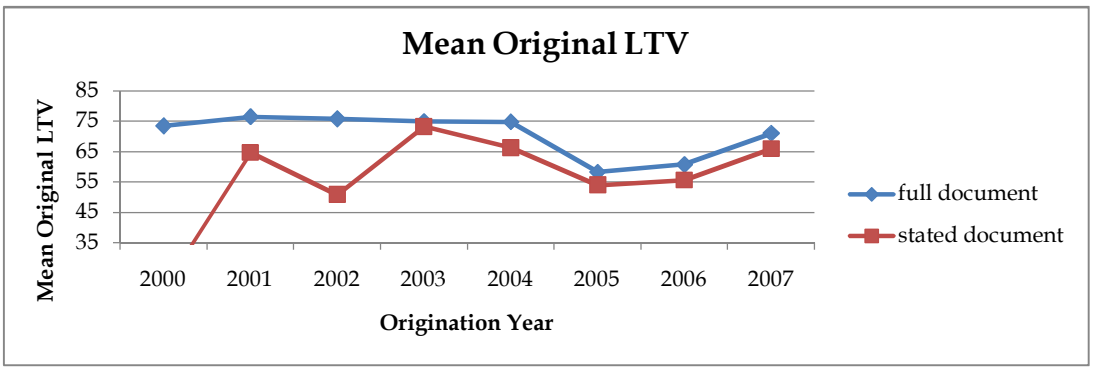


Figure 2 Time Trend of Loan Characteristics (Restricted Sample)









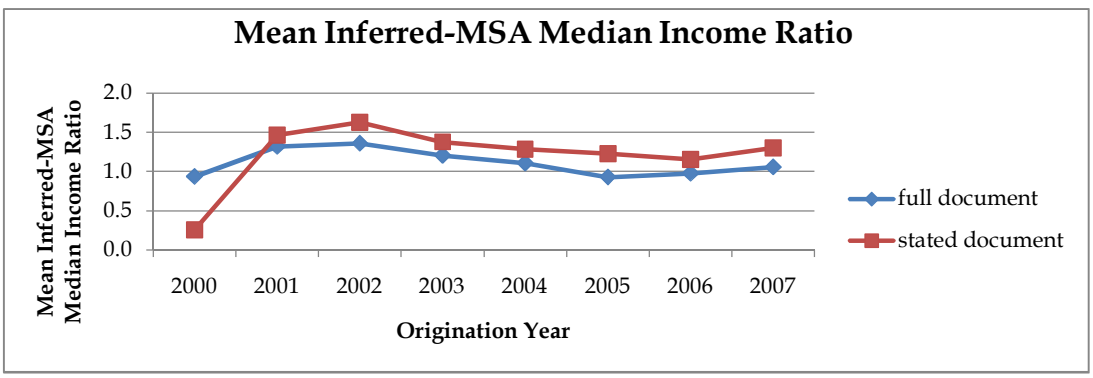


Table 1 Descriptive statistics for the full sample as a whole

Variable	N	Mean	Median	Minimum	Maximum	Std Dev
loan balance	218,589	209612	155072	4050	6375000	196972
original LTV	218,589	63.06	80	1	100	28.40
origination year	218,589	2005.6	2006	2000	2007	0.98
credit score	218,589	676.07	683	356	843	68.00
default	218,589	0.20	0	0	1	0.40
FRM	218,589	0.003	0	0	1	0.056
prime	218,589	0.04	0	0	1	0.20
subprime	218,589	0.31	0	0	1	0.46
ALT-A	218,589	0.63	1	0	1	0.48
AMPS	218,589	0.43	0	0	1	0.49
full-document	218,589	0.30	0	0	1	0.46
stated document	218,589	0.44	0	0	1	0.50
no document	218,589	0.18	0	0	1	0.38
loan age	218,589	28.61	25	4	1299	50.56

Figure 3 Documentation Type Distribution in the Full Sample

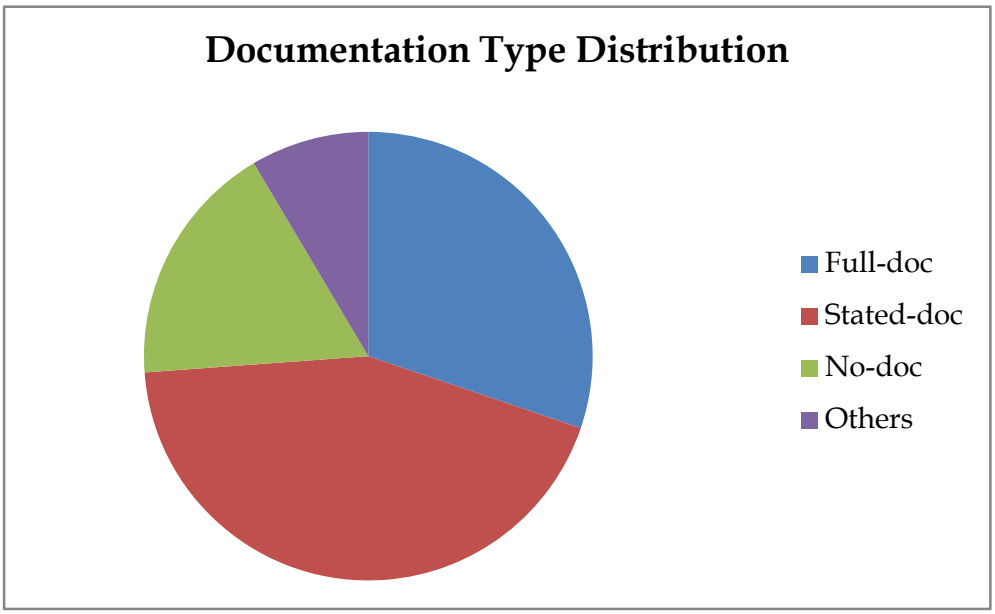


Table 2 Descriptive statistics for the subsamples of the full sample

Panel A

	Full-Document Subsample				Stated-Document Subsample						
Variable	**N**	**Mean**	**Median**	**Std Dev**	**N**	**Mean**	**Median**	**Std Dev**	**Mean Diff.**	**t-stat**	**p-value**
loan balance	66,032	171136	126810	161535	95377	234185	184800	214504	63049	67.30	<.0001
original LTV	66,032	63.67	80	28.93	95,377	63.47	80	27.75	-0.20	-1.40	0.1606
origination year	66,032	2005.4	2005	1.07	95,377	2005.7	2006	0.79	0.37	75.48	<.0001
credit score	66,032	660.70	663	72.39	95,377	686.63	693	62.84	25.93	74.62	<.0001
default	66,032	0.16	0	0.37	95,377	0.22	0	0.42	0.06	30.18	<.0001
FRM	66,032	0.0005	0	0.02	95,377	0.0001	0	0.01	-0.0004	-4.31	<.0001
prime	66,032	0.06	0	0.24	95,377	0.02	0	0.14	-0.04	-40.30	<.0001
subprime	66,032	0.47	0	0.50	95,377	0.23	0	0.42	-0.24	-100.72	<.0001
ALT-A	66,032	0.42	0	0.49	95,377	0.74	1	0.44	0.32	132.83	<.0001
AMPS	66,032	0.30	0	0.46	95,377	0.52	1	0.50	0.22	92.63	<.0001
loan age	66,032	32.23	28	60.24	95,377	27.26	25	48.58	-4.97	-17.59	<.0001

Panel B

	Stated-Document Subsample				No-Document Subsample						
Variable	**N**	**Mean**	**Median**	**Std Dev**	**N**	**Mean**	**Median**	**Std Dev**	**Mean Diffe**	**t-stat**	**p-value**
loan balance	95,377	234185	184800	214504	38559	227197	171200	211506	-6988	-5.45	<.0001
original LTV	95,377	63.47	80	27.75	38,559	56.50	75	29.34	-6.97	-39.99	<.0001
origination year	95,377	2005.7	2006	0.79	38,559	2005.9	2006	0.79	0.20	41.12	<.0001
credit score	95,377	686.63	693	62.84	38,559	702.54	701	48.23	15.91	49.88	<.0001
default	95,377	0.22	0	0.42	38,559	0.18	0	0.38	-0.04	-18.99	<.0001
FRM	95,377	0.0001	0	0.01	38,559	0.0002	0	0.01	0.0001	1.02	0.3062
prime	95,377	0.02	0	0.14	38,559	0.02	0	0.13	0.00	-1.92	0.0546
subprime	95,377	0.23	0	0.42	38,559	0.02	0	0.12	-0.22	-144.24	<.0001
ALT-A	95,377	0.74	1	0.44	38,559	0.97	1	0.18	0.23	135.56	<.0001
AMPS	95,377	0.52	1	0.50	38,559	0.56	1	0.50	0.03	11.60	<.0001
loan age	95,377	27.26	25	48.58	38,559	23.56	21	26.69	-3.69	-17.76	<.0001

Panel C

	Full-Document Subsample				No-Document Subsample						
Variable	**N**	**Mean**	**Median**	**Std Dev**	**N**	**Mean**	**Median**	**Std Dev**	**Mean Diffe**	**t-stat**	**p-value**
loan balance	66,032	171136	126810	161535	38559	227197	171200	211506	56061	44.95	<.0001
original LTV	66,032	63.67	80	28.93	38,559	56.50	75	29.34	-7.17	-38.35	<.0001
origination year	66,032	2005.4	2005	1.07	38,559	2005.9	2006	0.79	0.57	97.50	<.0001
credit score	66,032	660.70	663	72.39	38,559	702.54	701	48.23	41.84	111.95	<.0001
default	66,032	0.16	0	0.37	38,559	0.18	0	0.38	0.01	6.00	<.0001
FRM	66,032	0.0005	0	0.02	38,559	0.0002	0	0.01	-0.0003	-2.97	0.0029
prime	66,032	0.06	0	0.24	38,559	0.02	0	0.13	-0.04	-37.61	<.0001
subprime	66,032	0.47	0	0.50	38,559	0.02	0	0.12	-0.46	-223.73	<.0001
ALT-A	66,032	0.42	0	0.49	38,559	0.97	1	0.18	0.55	257.34	<.0001
AMPS	66,032	0.30	0	0.46	38,559	0.56	1	0.50	0.26	83.34	<.0001
loan age	66,032	32.23	28	60.24	38,559	23.56	21	26.69	-8.66	-31.96	<.0001

Table 3 Descriptive statistics for the restricted sample as a whole

Variable	N	Mean	Median	Minimum	Maximum	Std Dev
Ln(loan balance)	134,174	11.78	12	9	15	0.98
original LTV	134,174	63.09	80	2	100	28.83
origination year	134,174	2005.7	2006	2000	2007	0.83
credit score	134,174	669.18	675	411	843	68.67
default	134,174	0.20	0	0	1	0.40
inferred-MSA median-income ratio	134,174	1.10	0.81	0.00	9.99	1.09
MSA median household income	134,174	53817.89	53336	25701	84318	9184.99
MSA HPI	134,174	238.70	237	87	366	60.39
MSA ratio of household income to HPI	134,174	238.90	233	82	664	66.61
5-year historical growth in MSA median household income	134,174	0.14	0	0	0	0.08
5-year historical growth rate in MSA HPI	134,174	0.70	1	0	1	0.39
level of 30 Year FRM	134,174	6.19	6	0	9	0.36
yield curve slope	134,174	1.08	1	0	3	0.23
SP500 1-year return	134,174	0.09	0	0	0	0.05
FRM	134,174	0.00004	0	0	1	0.01
prime	134,174	0.03	0	0	1	0.16
subprime	134,174	0.42	0	0	1	0.49
ALT-A	134,174	0.55	1	0	1	0.50
AMPS	134,174	0.37	0	0	1	0.48
full-document	134,174	0.36	0	0	1	0.48
stated document	134,174	0.54	1	0	1	0.50
loan age	134,174	26.07	25	4	133	10.14

Table 4 Descriptive statistics for the subsamples of the restricted sample

Variable	Full-Document Subsample				Stated-Document Subsample				Comparisons		
	N	Mean	Median	Std Dev	N	Mean	Median	Std Dev	Mean Diff.	t-stat	p-value
Ln(loan balance)	48,819	11.62	12	0.94	72,505	11.87	12	1.01	0.25	43.5	<.0001
original LTV	48,819	62.42	80	29.62	72,505	61.86	80	28.68	-0.56	-3.26	0.0011
origination year	48,819	2005.5	2006	0.91	72,505	2005.8	2006	0.72	0.33	67.89	<.0001
credit score	48,819	659.93	662	70.84	72,505	683.33	689	63.04	23.41	58.96	<.0001
default	48,819	0.16	0	0.37	72,505	0.21	0	0.41	0.05	23.72	<.0001
inferred-MSA median-income ratio	48,819	0.94	1	0.93	72,505	1.20	1	1.16	0.26	43.57	<.0001
MSA median household income	48,819	52926.04	52713	8836.05	72,505	54714.79	53708	9465.45	1788.75	33.59	<.0001
MSA HPI	48,819	224.84	217	57.46	72,505	249.63	259	60.02	24.79	72.37	<.0001
MSA ratio of household income to HPI	48,819	248.82	248	66.73	72,505	231.73	220	65.45	-17.09	-44.08	<.0001
5-year historical growth in MSA median household inc	48,819	0.12	0	0.08	72,505	0.15	0	0.08	0.03	52.95	<.0001
5-year historical growth rate in MSA HPI	48,819	0.63	1	0.38	72,505	0.74	1	0.39	0.11	48.66	<.0001
level of 30 Year FRM	48,819	6.14	6	0.40	72,505	6.23	6	0.30	0.09	42.07	<.0001
yield curve slope	48,819	1.13	1	0.30	72,505	1.04	1	0.15	-0.09	-58.56	<.0001
SP500 1-year return	48,819	0.09	0	0.05	72,505	0.10	0	0.04	0.01	19.25	<.0001
FRM	48,819	0.00004	0	0.01	72,505	0.00000	0	0.00	-0.00004	-1.41	0.1573
prime	48,819	0.04	0	0.20	72,505	0.01	0	0.12	-0.03	-26.96	<.0001
subprime	48,819	0.53	1	0.50	72,505	0.26	0	0.44	-0.26	-93.97	<.0001
ALT-A	48,819	0.43	0	0.50	72,505	0.72	1	0.45	0.29	104.13	<.0001
AMPS	48,819	0.28	0	0.45	72,505	0.49	0	0.50	0.21	76.55	<.0001
loan age	48,819	28.14	27	11.22	72,505	24.61	24	8.65	-3.54	-58.81	<.0001

Table 5 Default analysis in two ranges of the inferred-MSA income ratio

Range Index	Range of Inferred-MSA Income Ratio	Loan Counts			Default Rate			Defaulted Loan Counts			Original LTV			Credit Score		
		All	Full-doc	Stated-doc	All	Full-doc	Stated-doc	All	Full-doc	Stated-doc	All	Full-doc	Stated-doc	All	Full-doc	Stated-doc
1	[0, 3.5)	129,275	47,716	69,211	19.55%	15.93%	20.89%	25,273	7,601	14,458	62.69	62.18	61.31	668	659	683
2	[3.5, ∞)	5,844	1,401	3,841	27.64%	17.99%	29.95%	1,615	252	1,150	73.21	72.43	73.16	690	688	699
	Total	**135,119**	**49,117**	**73,052**				**26,889**	**7,853**	**15,609**						

Figure 4 Statistics for Two Ranges of Inferred-MSA Median Income Ratio

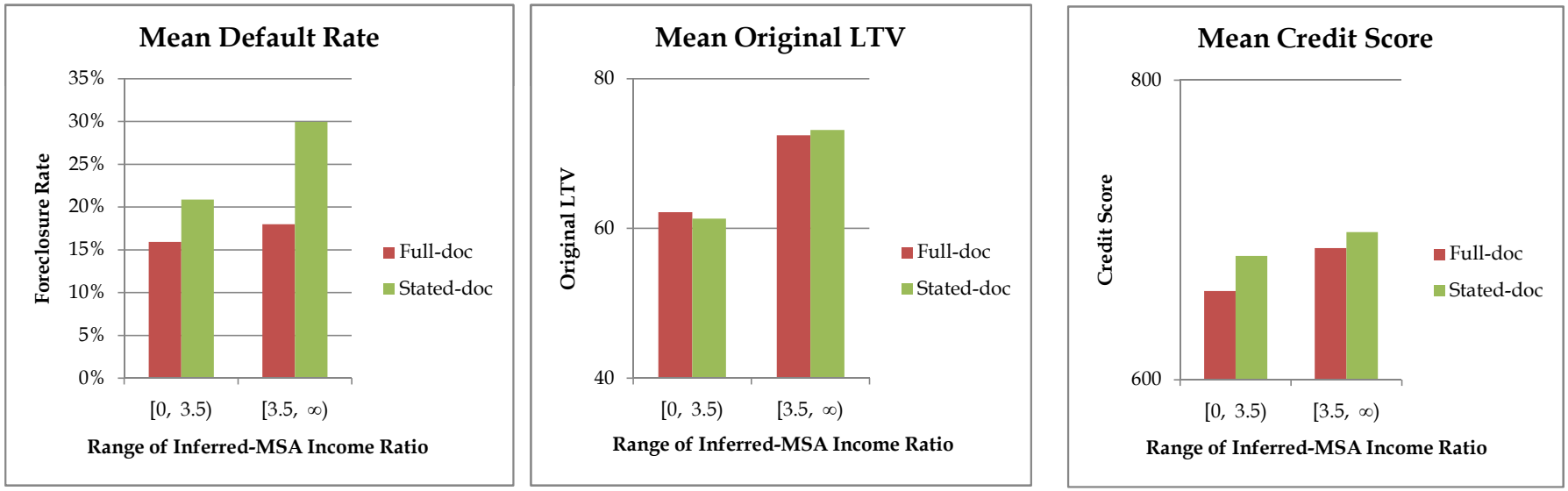


Table 6 Default analysis in ten ranges of the inferred-MSA income ratio

Range Index	Range of Inferred-MSA Income Ratio	Loan Counts			Default Rate			Defaulted Loan Number			Original LTV			Credit Score		
		All	Full-doc	Stated-doc	All	Full-doc	Stated-doc	All	Full-doc	Stated-doc	All	Full-doc	Stated-doc	All	Full-doc	Stated-doc
1	[0, 1.5)	106,444	41,247	55,109	17.59%	15.22%	18.09%	18,719	6,279	9,968	60.02	60.14	57.88	668	659	683
2	[1.5, 2.5)	17,590	5,137	10,679	28.75%	20.99%	31.85%	5,057	1,078	3,401	75.28	75.47	74.77	668	659	681
3	[2.5, 3.5)	5,241	1,332	3,423	28.56%	18.39%	31.81%	1,497	245	1,089	74.68	73.89	74.58	681	675	691
4	[3.5, 4.5)	2,131	486	1,447	27.69%	16.26%	30.27%	590	79	438	74.25	73.32	74.16	687	688	695
5	[4.5, 5.5)	1,086	230	744	28.55%	21.30%	31.72%	310	49	236	73.66	72.45	73.99	688	686	697
6	[5.5, 6.5)	661	166	439	27.08%	18.07%	29.61%	179	30	130	71.55	71.50	71.67	691	691	697
7	[6.5, 7.5)	439	97	288	28.47%	20.62%	27.78%	125	20	80	73.88	75.05	72.98	690	681	705
8	[7.5, 8.5)	305	72	189	27.21%	18.06%	28.57%	83	13	54	73.82	71.38	74.29	687	689	697
9	[8.5, 9.5)	195	34	132	31.28%	20.59%	33.33%	61	7	44	71.63	74.71	71.08	694	677	708
10	[9.5, ∞)	1,027	316	602	26.00%	17.09%	27.91%	267	54	168	71.49	70.72	71.01	695	690	704
	Total	**135,119**	**49,117**	**73,052**				**26,888**	**7,854**	**15,608**						

Figure 5 Statistics for Ten Ranges of Inferred-MSA Median Income Ratio

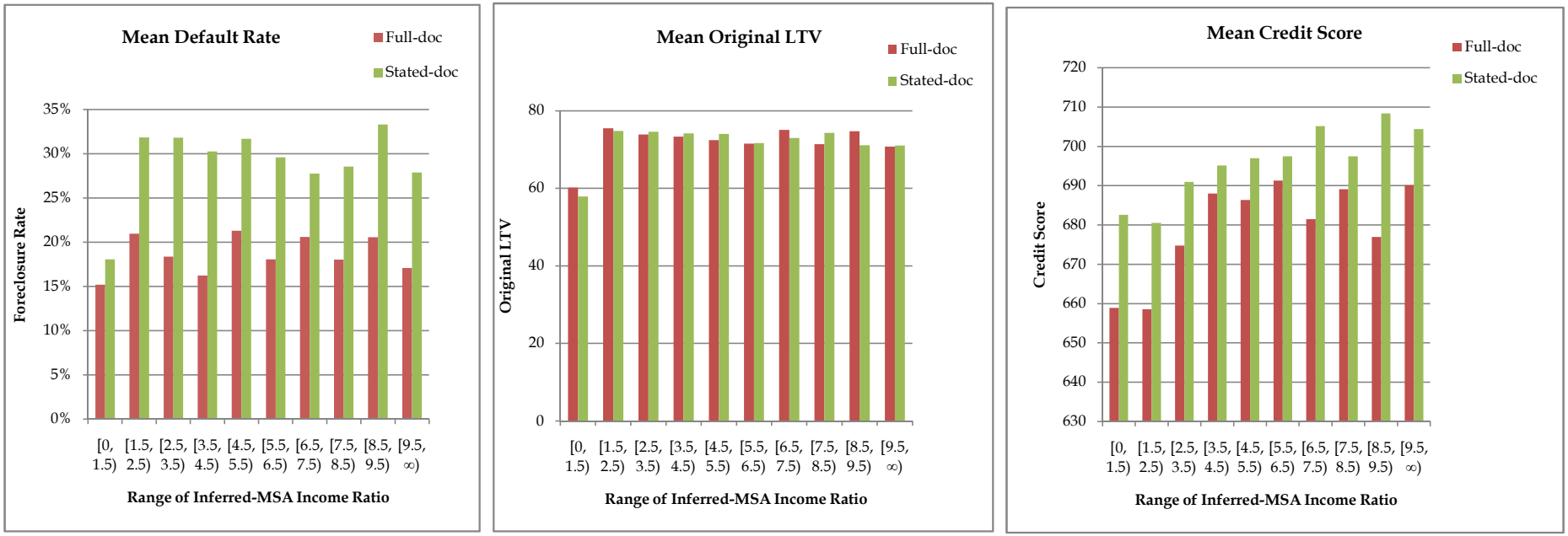


Table 7 Documentation type logit regression results: for the full sample

	Full-Documentation			Stated-Documentation			No-Documentation		
Variable	**coef.**	**p-value**		**coef.**	**p-value**		**coef.**	**p-value**	
[Specification 1]									
intercept	-0.928	<.0001	***	-0.533	<.0001	***	-1.529	<.0001	***
FRM dummy	-3.165	<.0001	***	-2.406	<.0001	***	-2.193	<.0001	***
AMPS dummy	-0.650	<.0001	***	0.611	<.0001	***	0.536	<.0001	***
origination year dummies	Yes			Yes			Yes		
[Specification 2]									
intercept	1.814	<.0001	***	-3.062	<.0001	***	-5.838	<.0001	***
original LTV	-0.001	<.0001	***	0.004	<.0001	***	-0.009	<.0001	***
credit score	-0.005	<.0001	***	0.004	<.0001	***	0.008	<.0001	***
origination year dummies	Yes			Yes			Yes		

Table 8 Default logit regression results: for the full sample

Variable	coef.	p-value	coef.	p-value	coef.	p-value
[Specification 1]						
intercept	-1.565	<.0001 ***	-1.206	<.0001 ***	-1.605	<.0001 ***
original LTV	0.034	<.0001 ***	0.035	<.0001 ***	0.035	<.0001 ***
credit score	-0.004	<.0001 ***	-0.004	<.0001 ***	-0.004	<.0001 ***
full-document dummy			-0.527	<.0001 ***		
stated document dummy	0.352	<.0001 ***				
no document dummy					0.139	<.0001 ***
origination year dummies	Yes		Yes		Yes	
[Specification 2]						
intercept	-1.027	<.0001 ***	-0.744	<.0001 ***	-1.018	<.0001 ***
Loan balance	0.000	<.0001 ***	0.000	<.0001 ***	0.000	<.0001 ***
original LTV	0.032	<.0001 ***	0.033	<.0001 ***	0.032	<.0001 ***
credit score	-0.005	<.0001 ***	-0.005	<.0001 ***	-0.005	<.0001 ***
FRM dummy	0.007	0.945	-0.232	0.028 **	-0.013	0.903
AMPS dummy	0.180	<.0001 ***	0.160	<.0001 ***	0.206	<.0001 ***
full-document dummy			-0.472	<.0001 ***		
stated document dummy	0.306	<.0001 ***				
no document dummy					0.079	<.0001 ***
origination year dummies	Yes		Yes		Yes	
[Specification 3]						
intercept	-1.458	<.0001 ***	-1.068	<.0001 ***	-1.520	<.0001 ***
original LTV	0.034	<.0001 ***	0.035	<.0001 ***	0.035	<.0001 ***
credit score	-0.004	<.0001 ***	-0.004	<.0001 ***	-0.004	<.0001 ***
full-document dummy			-0.546	<.0001 ***		
stated document dummy	0.359	<.0001 ***				
no document dummy					0.184	<.0001 ***
loan age	0.001	<.0001 ***	0.001	<.0001 ***	0.001	<.0001 ***
[Specification 4]						
intercept	-0.829	<.0001 ***	-0.539	<.0001 ***	-0.821	<.0001 ***
Loan balance	0.000	<.0001 ***	0.000	<.0001 ***	0.000	<.0001 ***
original LTV	0.032	<.0001 ***	0.033	<.0001 ***	0.032	<.0001 ***
credit score	-0.005	<.0001 ***	-0.005	<.0001 ***	-0.005	<.0001 ***
FRM dummy	-0.004	0.964	-0.283	0.005 ***	-0.092	0.358
AMPS dummy	0.171	<.0001 ***	0.145	<.0001 ***	0.202	<.0001 ***
full-document dummy			-0.479	<.0001 ***		
stated document dummy	0.304	<.0001 ***				
no document dummy					0.109	<.0001 ***
loan age	0.001	<.0001 ***	0.001	<.0001 ***	0.001	<.0001 ***

Table 9 Inferred-MSA median-income ratio regression result

Dependent variable is the natural logarithm of inferred-MSA median income ratio

Variable	coef.	p-value	coef.	p-value
intercept	-0.166	<.0001	0.033	<.0001
full-document dummy			-0.242	<.0001
stated document dummy	0.203	<.0001		
origination year dummies	Yes		Yes	

Table 10 Default logit regression results: restricted sample

	Full-Doc		Stated-Doc			
Variable	Estimate	p-value	Estimate	p-value	Difference	p-value
[Specification 1]						
intercept	3.388	<.0001 ***	3.589	<.0001 ***	0.201	<.0001 ***
credit score	-0.008	<.0001 ***	-0.006	<.0001 ***	0.001	<.0001 ***
Ln (inferred-MSA median-income ratio)	0.558	<.0001 ***	0.764	<.0001 ***	0.206	<.0001 ***
MSA ratio of household income to HPI	-0.001	<.0001 ***	-0.003	<.0001 ***	-0.002	<.0001 ***
SP500 1-year return	0.674	0.0153 **	0.989	0.0003 ***	0.315	0.4194
origination year dummies		Yes		Yes		
[Specification 2]						
intercept	3.166	<.0001 ***	3.434	<.0001 ***	0.268	<.0001 ***
credit score	-0.008	<.0001 ***	-0.006	<.0001 ***	0.001	<.0001 ***
Ln (inferred-MSA median-income ratio)	0.574	<.0001 ***	0.806	<.0001 ***	0.232	<.0001 ***
yield curve slope	-0.073	0.3448	-0.630	<.0001 ***	-0.558	0.0002 ***
SP500 1-year return	0.764	0.0063 ***	1.227	<.0001 ***	0.463	0.2380
origination year dummies		Yes		Yes		
[Specification 3]						
intercept	2.921	<.0001 ***	2.187	<.0001 ***	-0.734	<.0001 ***
credit score	-0.008	<.0001 ***	-0.007	<.0001 ***	0.001	<.0001 ***
Ln (inferred-MSA median-income ratio)	0.565	<.0001 ***	0.775	<.0001 ***	0.211	<.0001 ***
MSA HPI	0.001	0.0004 ***	0.003	<.0001 ***	0.002	<.0001 ***
SP500 1-year return	0.704	0.0112 **	0.961	0.0004 ***	0.257	0.5099
origination year dummies		Yes		Yes		
[Specification 4]						
intercept	3.024	<.0001 ***	2.626	<.0001 ***	-0.397	<.0001 ***
credit score	-0.008	<.0001 ***	-0.007	<.0001 ***	0.001	<.0001 ***
Ln (inferred-MSA median-income ratio)	0.563	<.0001 ***	0.776	<.0001 ***	0.213	<.0001 ***
FRM dummy	-8.725	0.9209		***		
5-year historical growth rate in MSA HPI	0.160	<.0001 ***	0.512	<.0001 ***	0.353	<.0001 ***
SP500 1-year return	0.702	0.0114 **	1.015	0.0002 ***	0.314	0.4199
origination year dummies		Yes		Yes		
[Specification 5]						
intercept	2.783	<.0001 ***	1.645	<.0001 ***	-1.138	<.0001 ***
credit score	-0.008	<.0001 ***	-0.006	<.0001 ***	0.001	<.0001 ***
Ln (inferred-MSA median-income ratio)	0.574	<.0001 ***	0.804	<.0001 ***	0.230	<.0001 ***
level of 30 Year FRM	0.048	0.2217	0.177	0.0004 ***	0.129	0.0433 **
SP500 1-year return	0.772	0.0056 ***	1.619	<.0001 ***	0.847	0.0352 **
origination year dummies		Yes		Yes		
[Specification 6]						
intercept	3.256	<.0001 ***	2.437	<.0001 ***	-0.818	<.0001 ***
credit score	-0.008	<.0001 ***	-0.006	<.0001 ***	0.001	<.0001 ***
Ln (inferred-MSA median-income ratio)	0.580	<.0001 ***	0.797	<.0001 ***	0.217	<.0001 ***
loan age	0.013	<.0001 ***	0.025	<.0001 ***	0.013	<.0001 ***

Table 11 The effects of inferred-MSA median income ratio cap

Inferred-MSA Income Ratio Cap	Loan Numbers				Mean Default Rate						Mean Inferred-MSA Median-Income Ratio					
	All	Full-doc	Stated-doc	Stated-Full Difference	All	Full-doc	Stated-doc	Stated-Full Difference	Difference P-value		All	Full-doc	Stated-doc	Stated-Full Difference	Difference P-value	
1.5	106,444	41,247	55,109	13,862	17.59%	15.22%	18.09%	2.86%	<.0001	***	0.695	0.649	0.725	0.076	<.0001	***
2	118,314	44,883	62,182	17,299	18.72%	15.68%	19.70%	4.02%	<.0001	***	0.798	0.736	0.839	0.103	<.0001	***
2.5	124,034	46,384	65,788	19,404	19.17%	15.86%	20.32%	4.46%	<.0001	***	0.864	0.784	0.914	0.131	<.0001	***
3	127,321	47,240	67,914	20,674	19.42%	15.92%	20.70%	4.78%	<.0001	***	0.912	0.819	0.971	0.152	<.0001	***
3.5	129,275	47,716	69,211	21,495	19.55%	15.93%	20.89%	4.96%	<.0001	***	0.947	0.843	1.014	0.171	<.0001	***
4	130,501	47,995	70,045	22,050	19.62%	15.92%	21.01%	5.09%	<.0001	***	0.973	0.860	1.046	0.186	<.0001	***
4.5	131,406	48,202	70,658	22,456	19.68%	15.94%	21.08%	5.15%	<.0001	***	0.996	0.874	1.074	0.199	<.0001	***
5	132,047	48,333	71,101	22,768	19.72%	15.94%	21.15%	5.21%	<.0001	***	1.014	0.885	1.096	0.212	<.0001	***
5.5	132,492	48,432	71,402	22,970	19.75%	15.96%	21.19%	5.23%	<.0001	***	1.028	0.894	1.114	0.220	<.0001	***
6	132,872	48,517	71,662	23,145	19.78%	15.97%	21.22%	5.25%	<.0001	***	1.041	0.902	1.131	0.229	<.0001	***
…				…				…						…		
10	134,174	48,819	72,505	23,686	19.86%	15.98%	21.32%	5.34%	<.0001	***	1.104	0.942	1.205	0.263	<.0001	***
100	135,015	49,070	73,001	23,931	19.90%	16.00%	21.36%	5.36%	<.0001	***	1.228	1.059	1.330	0.271	<.0001	***
1000	135,079	49,101	73,030	23,929	19.90%	15.99%	21.36%	5.36%	<.0001	***	1.411	1.294	1.502	0.208	<.0001	***
∞	135,119	49,117	73,052	23,935	19.90%	15.99%	21.37%	5.38%	<.0001	***	2.266	2.270	2.394	0.124	<.0001	***

Figure 6 The Illustrations of Effects of Inferred-MSA Income Ratio Cap

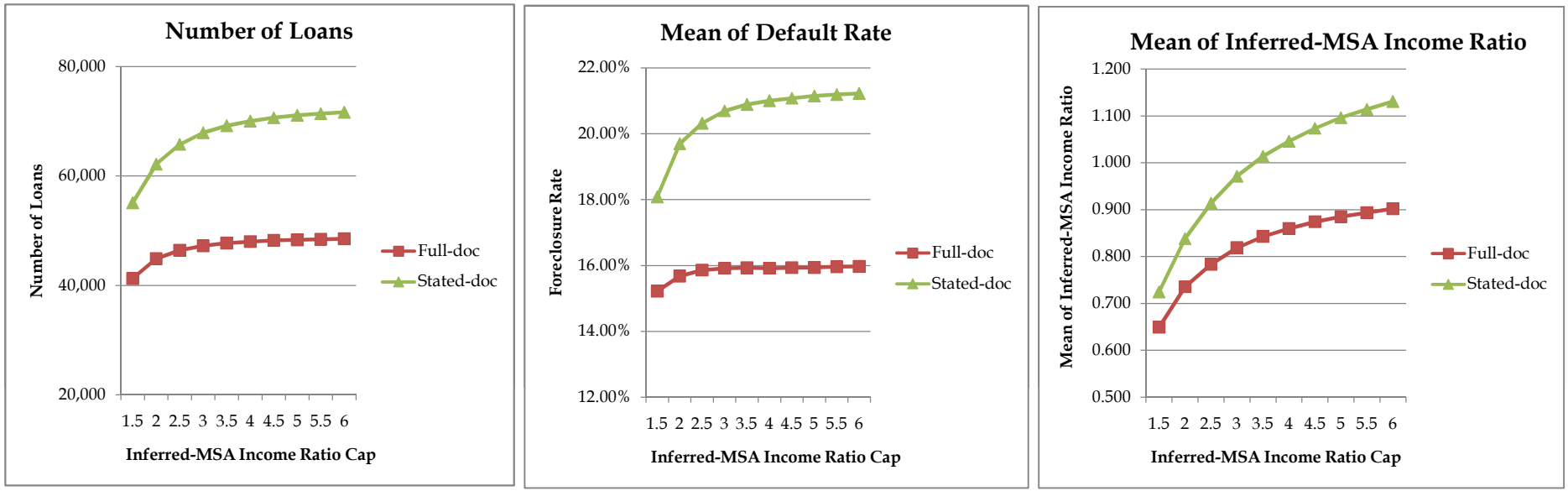


Table 12 Regressions results for the effects of inferred-MSA median income ratio cap

	All Loans		Full-Doc		Stated-Doc		Stated-Full Difference	
Variable	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value
[Specification 1] Regression of loan number sample mean								
intercept	108714	<.0001 ***	42205	<.0001 ***	56182	<.0001 ***	13978	<.0001 ***
Cap	4735.079	0.002 ***	1250.703	0.003 ***	3019.976	0.001 ***	1769.273	0.001 ***
Adj. R-square		0.699		0.648		0.722		
[Specification 2] Regression of default rate sample mean								
intercept	0.179	<.0001 ***	0.154	<.0001 ***	0.185	<.0001 ***	0.031	<.0001 ***
Cap	0.004	0.004 ***	0.001	0.017 ***	0.005	0.004 ***	0.004	0.002 ***
Adj. R-square		0.629		0.468		0.636		
[Specification 3] Regression of inferred-MSA median income ratio sample mean								
intercept	0.665	<.0001 *	0.640	<.0001 *	0.682	<.0001 *	0.043	0.002 ***
Cap	0.070	<.0001 ***	0.049	0.000 ***	0.083	<.0001 ***	0.033	<.0001 ***
Adj. R-square		0.876		0.835		0.892		