Taking the Lie Out of Liar Loans: The Effect of Reduced Documentation on the Performance and Pricing of Alt-A and Subprime Mortgages

Authors Michael LaCour-Little and Jing Yang

Abstract

We present a simple theoretical model of adverse selection when lenders allow reduced documentation. The model shows how reduced documentation attracts both riskier borrowers and larger size loans. We then empirically test implications of the model using stated income loans originated during the recent housing market run-up and collapse. After estimating the extent to which these loans have higher default rates than do fully documented loans, we develop a measure for the extent of income over-statement, providing results for both the Alt-A and subprime market segments. We also estimate that the incremental risk in these mortgages was priced at less than ten basis points.

The surge in default and foreclosure rates in residential mortgages in the United States during the recent financial crisis 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 mortgages (ARMs), subprime mortgages, and Alt-A quality credit (Campbell and Cocco, 2003; Wheaton and Nechayev, 2008; LaCour-Little and Yang, 2010; Berkovec, Chang, and McManus, 2011; Ding, Quercia, Li, and Ratcliffe, 2011; Pavlov and Wachter, 2011; and MacDonald and Winson-Geideman, 2012). While lax underwriting and misaligned incentives in the mortgage securitization process have been blamed for the market meltdown, the explicit effect of reduced loan documentation has not been widely studied. In contrast, a 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 home purchase loans originated during 2000–2007 and securitized by Bear Stearns with loan performance information observed as of May 2009. Tables in Appendix A compare our data to aggregate measures reported from the 2011 Mortgage Market Statistical Annual.[2] We find that the percentage of loans with full documentation dropped significantly from 42% to 26% 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 12% to 73%. 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 53% are classified as Alt-A. Among loans omitting information on traditional underwriting variables, including income, assets, and front-end and back-end ratios, which we call ‘‘no-doc’’ loans, almost all are ARMs and 90% are Alt-A.

Stated-doc loans also became an important component of the subprime segment. According to the 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] This may have stemmed from increased competition in the banking industry as interest rates fell to unprecedented low levels.[4] We believe our study is among the first to explore in detail the effects of loan documentation on default risk and analyze the underlying mechanisms that drive these results. We discuss the related work of Jiang, Nelson, and Vytlacil (2009) later in the text.

Residential loan applicants encounter varied documentation requirements with subtle differences across lenders and loan programs.[5] The following are some major categories: (1) ‘‘full doc,’’ under which the borrower must prove income with income tax returns, W-2s, paycheck with YTD earnings information, verification of employment, and evidence of assets; (2) ‘‘lite doc’’ in which the borrower provides bank statements to document income in lieu of paycheck stubs, W-2s, and tax returns, as well as a proof for employment; (3) SIVA (stated income/verified assets) in which borrower income is stated with only employment and assets verified; (4) SISA (stated income/stated assets), where both income and assets are stated with only employment verified; (5) NORA (no ratio) in which employment and assets are verified, income appears but debt ratios are not calculated; (6) NINA (no income/no asset) in which neither income nor assets are listed but employment is still verified; and (7) ‘‘no doc,’’ for which assets, income, and employment are all omitted from the loan application and the lender generally relies solely on credit score and loan-to-value (LTV) ratio.

Our focus here is the stated-income categories (SIVA and SISA). These loans were originally developed for self-employed borrowers or those with seasonal income for which income is hard to document or verify. We speculate that the low default rates experienced during the housing market run-up of 2000–2005 together with the demand for production to fuel private label MBS issuance caused lenders to make reduced documentation features more widely available to otherwise risky borrowers. In addition, mortgage brokers may have solicited stated-income loans because they produce ‘‘excessive rates and penalties’’ (Harney, 2009).

Academic research on the topic of loan documentation has been limited. LaCour-Little (2007) confirms the traditional relationship posited between selfemployment 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. 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) examine the performance of both fixed and adjustable-rate subprime mortgages using multi-lender data on securitized loans and report 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. Gerardi, Lehnert, Sherlund, and Willen (2009) report an increasing use of low documentation in the subprime segment, and a more rapid increase in default rates over time for loans with reduced documentation.

In a more recent study, Paley and Tzioumis (2011) argue that stated documentation loan performance varies based on whether the borrower, or the lender, initiated the reduced documentation. They employ a detailed dataset developed for the Office of the Comptroller of the Currency (OCC) called OCC Mortgage Metrics. This highly-detailed dataset allows the user to distinguish whether the borrower or the lender initiated the reduced documentation. Paley and Tzioumis argue that lenders may have superior private information about borrowers that may allow them to selectively reduce documentation requirements for better customers. In their empirical analysis, low doc loans initiated by lenders have loan performance that is comparable to full doc loans. In contrast, when borrowers initiated the reduced documentation, loan performance is dramatically worse. They estimate an average 10% income over-statement and also report 0.12% higher APR for such loans, a level of risk under-pricing that is similar to the estimates we later report here.

Building on these ideas, we begin with a theoretical model that differentiates on borrower quality dimensions, some of which are unobservable. We show that reduced documentation will worsen the loan performances of only those we call ‘‘weak’’ borrowers, while having little effect on ‘‘strong’’ borrowers. Our data, unlike OCC Mortgage Metrics, does not allow us to identify who initiated the reduced documentation, borrower or lender. However, since Bear Stearns was not a depository institution and originated its loans through wholesale channels, we think it unlikely they had superior private information about borrower credit quality.

Another related study is Rajan, Seru, and Vig (2008), who focus on what they term ‘‘hard versus soft information’’ in the context of asset securitization. Credit score is a good example of hard information, 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., expected future income for the borrower). Rajan et al. 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 borrowers who have the same hard information but heterogeneous soft information, increasing default risk. Our study shows that under mortgage securitization, lenders may have a tendency to reduce their reliance on even hard information, further weakening screening efficiency and exacerbating default risk.

Jiang, Nelson, and Vytlacil (2009) also examine the topic of stated income loans, as well as other issues, using data from a single large lender over the period 2004–2009. Similar to our work, they develop measures of likely income over-statement. They employ two methods: (1) a comparison of stated income to local area averages based on IRS data at the ZIP Code level; and (2) a regression approach in which income is estimated based on borrower and neighborhood characteristics using full doc loans, with regression coefficients then applied to low doc loans. The difference between predicted and stated income then becomes a measure of income over-statement. Jiang et al. estimate that the median income over-statement is $753 per month or about 20%.

Another paper that helps motivate our analysis here is Nichols, Pennington-Cross, and Yezer (2005). These authors argue that the credit supplied in the mortgage market is discrete as lenders specialize in different borrower risk categories. Focusing on the subprime segment, they argue that lenders have incentives to reduce credit standards if loan underwriting costs exceed expected credit losses. It would be straightforward to extend this theory to our study: lenders targeting medium- to high-quality borrowers have incentives to reduce documentation requirements when documentation costs exceed credit losses. This tendency is stronger when lenders can transfer away risk through securitization.[6] However, use of stated documentation produces adverse selection.

In summary, our study contributes to the broader literature on residential mortgage loan performance by providing the first comprehensive study on the effects of loan documentation levels. In addition, we find only a small pricing premium, a fact that may have attracted weak borrowers to the low documentation option. One implication of our work is that higher pricing and/or enhanced underwriting could have reduced the volume of ''liar loans,'' as well as the elevated default rates they produced.

The structure of the paper is as follows. We first present the theoretical model of documentation type and risk pricing. We then describe the data and the empirical methodology, which is followed by a discussion of the results. We close with concluding remarks.

A Simple Theoretical Model

We start with a simple one-period, three-party game-theoretic model to explore the relationship between reduced documentation and default risk. The three parties are one mortgage lender and two borrowers. The two borrowers differ in their borrowing capacities, that is, the maximum loan size that the borrowers can service. The strong borrower has a high borrowing capacity B, while the weak borrower has a low borrowing capacity gB (with $g \in [0,1)$). This difference may result from exogenous factors such as income or equity available. The lender can require either full documentation or stated documentation. Full-doc reveals borrowing capacity, so the lender can limit the loan amount. Stated-doc, however, implies information asymmetry as to borrowing capacity, and hence we assume that the lender will offer each borrower the same loan amount B. In other words, we presume the existence of the loan-size adverse selection problem—the weak borrower can take advantage of the loose documentation requirements to get a loan that is larger than her/his borrowing capacity can support. In the real world, a weak borrower might borrow too much due to myopia about future income prospects. She/he may also borrow a lot due to rational incentives. For instance, in a rising market, the borrower may anticipate easy resale or refinancing, which will reduce mortgage default risk. For an aggressive borrower, she/he may want to buy a larger house anticipating a larger capital gain from resale.[7] We also assume that full documentation will bring additional cost to the lender, which is assumed to be a proportion u of the loan amount.[8] Full-doc will also impose a cost d to the borrower.

Unlike borrowing capacity, credit score is hard information that is readily available to the lender at low cost. For simplicity, but without the loss of generality, we assume that the strong borrower has a higher credit score than the weak borrower, and the corresponding credit risk premium in loan pricing is exogenously set at $e > 0$. At $t = 0$, the lender chooses documentation risk premium x, that is, the loan rate spread between the two documentation types, which we assume takes the same value across borrowers. Each borrower will then choose a documentation type. At $t = 1$, loans are repaid. We assume that the strong borrower will never default, while the weak borrower will default with a probability P under full-doc and $P + k$ under stated-doc. Note that $k \geq 0$ is the incremental default risk arising from stated documentation, which is probably associated with over-borrowing.[9]

Appendix B illustrates the game process. At $t = 0$, knowing the full-doc cost u, the lender decides on the documentation risk premium, x. Observing x, the base loan rate r and the borrower's credit risk premium e, each borrower chooses a loan documentation type. The strong borrower can either incur a documentation cost d to get a full-doc loan at rate r, or obtain a stated-doc loan at rate $r + x$ without incurring any documentation cost. Similarly, the weak borrower can either incur cost d to get a full doc loan at rate $r + e$, or incur no cost and get a stated-doc loan at rate $r + e + x$. At time $t = 1$, the strong borrower will pay off the loan principal and interest regardless of the documentation type chosen; the weak borrower, however, may default, producing a loss S to the lender.

Proposition 1: The lender's optimal documentation risk premium x and two borrowers' optimal documentation type choices under condition $k > (1 - g)(1 - P)$ are summarized in the Exhibit 1.

Exhibit 1 | Model Results

Range of x	Strong Borrower's Choice	Weak Borrower's Choice	Local optimum for x	Global optimum?
$(-\infty, x^S]$	Stated-doc	Stated-doc	x^S	No
$(x^S, x^W]$	Full-doc	Stated-doc	x^W	Yes if $\phi \geq 0$
$(x^W, +\infty]$	Full-doc	Full-doc	Any $x > x^W$	Yes if $\phi < 0$

$$x^S = \frac{d}{B} > 0, \tag{1}$$

$$x^W = \frac{d - kS + [k - (1 - g)(1 - P)](1 + r + e)B}{(1 - P - k)B} > 0, \tag{2}$$

$$\phi = r + g(1 - P)(1 + r + e) - u - 1. \tag{3}$$

x^S and x^W are the documentation risk premiums that make the strong and weak borrowers indifferent to the documentation type choice, respectively. See Appendix C for the proof.

Proposition 1 proposes four interesting outcomes for loan documentation type choice and pricing: (1) The weak borrower is more likely to choose stated-doc than the strong borrower, as long as $x^W > x^S$ holds (one sufficient condition of which is $k > (1 - g)(1 - P)$, which is generally true as a weak borrower is often more willing to accept a high doc risk premium than a strong borrower). (2) Reduced documentation will increase the mortgage loan rate, in other words, the doc-risk premium is positive. Paley and Tzioumis (2011) suggest that the rate of a reduced-doc loan will be higher only if this relaxed documentation requirement is a borrower choice (indicating that the borrower is very likely to be ''bad''), while not so if the reduced documentation is a lender choice (indicating that the borrower is very likely to be ''good''). Our results show that when the lender has no such way to observe borrowing capacity, under information asymmetry, he will have to charge a higher loan rate for stated-doc than for full-doc, regardless of borrower type. (3) Under condition $\phi \geq 0$, the lender's optimal doc risk premium will lead to a borrower-type adverse selection—the stated-doc loans will attract the weak borrower while repelling the strong borrower, raising the average default risk of stated-doc loans. (4) Under condition $\phi < 0$, the lender should price the documentation risk premium sufficiently high, forcing every borrower to choose full-doc to eliminate adverse selection. However, if the lender underprices the doc risk, an adverse selection will still appear, generating a positive relation between low documentation and default risk.

From Equation (3), condition $\phi \geq 0$ is more likely to hold when r, e, or g is larger, or when P is lower. To interpret these, when a weak borrower faces tight mortgage conditions such as a high loan rate r and/or a high credit risk premium e, the borrower might want to switch from full-doc to stated-doc, because reduced documentation can help relax mortgage constraints; the lender will also be willing to allow stated documentation when this weak borrower's default risk P is not too high, and/or when stated-doc will not inflate the weak borrower's choice of loan size too much (that is, when g is big). These, however, do not usually apply to a strong borrower. Therefore, a borrower-type adverse selection problem may appear: the stated-doc loans will attract the weak borrower but not the strong borrower.

Note that in the real world, lenders might initially set their loan documentation premium based on prior experience and that prior experience might involve only strong borrowers. As a result, lenders might set the documentation risk premium too low. In our model, this means that the documentation risk premium could be as low as x^S, so the low-doc loans will attract both strong and weak borrowers. This will alleviate the borrower-type adverse selection problem, but intensify the loan-size adverse selection problem, especially in a rising property market by inducing the weak borrowers to borrow even more. In our model, we assume an exogenous loan size factor g, which in reality will be affected by the documentation risk premium x.

Unlike earlier adverse selection models such as Rothschild and Stiglitz (1976) and its mortgage application in Brueckner (2000), in our model information asymmetry is endogenous. Information asymmetry, and hence adverse selection, is determined by the lender's doc risk pricing and the borrower's documentation choice response. Information asymmetry and adverse selection will appear only if the doc-risk premium is priced as such, that a borrower chooses stated-doc (which gives a bad borrower a chance to overstate income), but will not if a full-doc is chosen.

Thus, our model demonstrates the relationship between the doc-risk premium (pricing) and adverse selection (and, hence, changes in default risk). If pricing is high, both types of borrowers will choose full-doc, which produces no adverse selection. If pricing is low, adverse selection may appear: a good borrower will choose full-doc, while a bad borrower will choose low-doc. If pricing is extremely low, both borrowers will choose stated-doc, and information asymmetry will still appear. So, if we observe a positive relationship between default risk and stated-doc, a possibility is that the pricing of stated-doc is too low. Our model helps us explain why we care about pricing when analyzing the relationship between default risk and reduced documentation.

Our model also shows why lenders may be willing to accept adverse selection risk and offer stated-doc at a low price. They may do so either to generate doc type risk premiums or to avoid a high documentation cost.

It can be seen that our model contains quite rich information. We cannot, of course, test the model directly. However, we can test the model's implications indirectly. For instance, if documentation risk is underpriced, we may see a positive correlation between default risk and low documentation; if low documentation is associated with income overstatement, we may see a higher default risk, and so on. In our empirical tests, we focus on the following hypotheses:

Hypothesis 1: Overall, low documentation creates additional default risk.

Hypothesis 2: Overall, low documentation combined with income overstatement leads to a more pronounced default risk increase.

Hypothesis 3: Overall, low documentation risk and income overstatement risk are priced, but risk premiums are small.

Data and Methodology

Data

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

The loan-level data consist of loans securitized by Bear Stearns during 2000–2007, restricting the sample to home purchases, single-family dwelling units, with loan terms of thirty years. After deleting observations with missing data, we have a dataset of 92,771 loans, henceforth the ''full sample.''[10] As income per se is not a data element, we infer it from loan payment (PITI) and front-end ratio.[11] We measure ''income exaggeration'' by the ratio of the inferred borrower income to MSA median household income.[12] Alternatively, we measure income overstatement of a stated-doc loan by the ratio of the inferred borrower income to the borrower's ''predicted income'' as generated by a regression using the full-doc data, as explained later. For the ''no ratio'' and ''no doc'' categories, there is no front-end ratio from which to infer borrower income, so we discard those observations. There are also a small number of loans for which the ratio of borrower income to MSA median household income is extremely high. We treat these as outliers that may be inaccurately recorded. After exclusions, the final subsample consists of 60,465 loans or about 65% of the full sample. We call this the ''restricted sample.'' We calculate loan age (in months) as of May 2009 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 is reported in different ways 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.[13] In the full sample, these three categories comprise 36%, 39%, and 10% of the loans, respectively. We define a loan to be ‘‘low documentation’’ if either stated-doc or no-doc. In the restricted sample, the full-doc and stated-doc comprise 40% and 45% of all loans, respectively, and given that no-doc loans do not have income information, our major comparison within the restricted sample data is between the stated-doc and full-doc loans. Documentation type was missing in 15% of the loans in both the full sample and the restricted sample.

We also include local variables to control for market-specific factors. We include local housing price levels, which we measure using the publicly available MSA-level FHFA HPI; the five-year average annual growth rate in MSA HPI; the MSA-median household income; the local wealth level measured as the interaction of the ZIP Code median household income and the ZIP Code median age [as in LaCour-Little and Yang (2010)]; 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.[14] 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 for the S&P 500 Index. The level of mortgage rate is measured by the contract rate on 30-year, fixed-rate conventional home mortgages, based on the Freddie Mac Primary Mortgage Market Survey data.

Methodology

We address the three hypotheses previously mentioned. We also explore related issues such as whether low documentation and income exaggeration risk are more pronounced among subprime or Alt-A loan types.

Does Low Documentation Create Additional Default Risk? To examine this question, we need to be aware that an effect of documentation type on default risk might be caused by the high correlation of documentation type and other default risk determinants. To isolate the individual effect of documentation type on default risk, we develop the following multi-stage regressions. Using the full sample, the first stage is a logit regression of loan default:[15]

$$\text{(First stage)} \quad D_{default} = a + \Sigma_{j=1}^{v} \gamma_j V_j + \vartheta, \tag{4}$$

where D_{default} is the default dummy, which takes on the value of 1 if the data show that the loan has defaulted; V contains control variables (other than loan documentation type) that are expected to affect default probability, including loan factors such as credit score and LTV, MSA-level factors such as local housing market affordability, and capital market conditions such as one-year return in the S&P 500 Index; α is the intercept; γ_j ($j = 1,..., v$) is the coefficient for the jth control variables; and finally, ϑ is the error term. Loan origination year dummies are included to control for vintage fixed-effects.

In the second stage, we estimated the following two-specification models:

$$\text{(Second stage)} \quad P_{default} = \alpha_b + \beta_s D_s + \beta_n D_n + \vartheta_b, \tag{5}$$

$$P_{default} = \alpha_l + \beta_l D_l + \vartheta_l, \tag{6}$$

where P_{default} is the residual from the first-stage logit regression, after conversion into default probability; D_s, D_n, and D_l are the dummies for stated-doc, no-doc, and low-doc loans; α_b and α_l are the intercepts; β_s, β_n, and β_l are the coefficients for the documentation type dummies; and ϑ_b and ϑ_l are the error terms. In both specifications, the reference documentation type is the full documentation.

This two-stage regression can help separate out the individual effects of documentation types on loan default risk from the effects of other variables.[16] Hypothesis 1 is specified as:

Hypothesis 1: Stated-doc and no-doc loans are more likely to default than full-doc loans. In regressions (5) and (6), this means that the stated-doc dummy, no-doc dummy, and low-doc dummy positively affect the default probability, that is, β_s, β_n, and $\beta_l > 0$.

Does Low Documentation Combined with Income Overstatement Lead to a More Pronounced Default Risk Increase? Here we also need to consider the possible correlations between income exaggeration and other default risk determinants.[17] To separate out the individual effect of income exaggeration on default risk from the effects of other default risk determinants, we again use a two-stage strategy. In the first stage, we estimate the following logit model for the full-doc subsample and the stated-doc subsample in the restricted sample:

$$\text{(First stage)} \quad D_{default} = \pi + \Sigma_{j=1}^{k} \omega_j S_j + \xi, \tag{7}$$

where D_{default} is the default dummy; S contains control variables (other than income exaggeration) that are also expected to affect default probability; π is the intercept; $\omega_j(j = 1,..., \upsilon)$ are the coefficients for control variables; and ξ is the error term. Fixed effects for loan origination year are included.

In the second stage, we estimate the following model, for both the full-doc subsample and the stated-doc subsample in the restricted sample:

$$\text{(Second stage)} \quad P_{default} = \pi_r + \mu_r \, Incratio + \xi_r, \tag{8}$$

where P_{default} is the residual from the first-stage logit regression, after conversion into default probability; *Incratio* is the measure of income exaggeration; π_r is the intercept; μ_r is the coefficient for *Incratio*; and ξ_r is the error term. Hypothesis 2 is specified as:

Hypothesis 2: The default rate is increasing in the income ratio for the stated-doc subsample, with a stronger effect than that for the full-doc subsample. In regression (8), this means that with the stated-doc subsample data, $\mu_r > 0$, and in addition, μ_r is larger in positive magnitude than when estimated using the full-doc subsample data.

Are Low Documentation Risk and Income Overstatement Risk Appropriately Priced? Research has suggested that under-pricing of the default risk inherent in alternative mortgage products contributed to the mortgage crisis (LaCour-Little and Yang, 2010; Pavlov and Wachter; 2010). Analogously, if lenders earned inadequate risk premiums for reduced loan documentation, this may help explain the proliferation of the product and the unfortunate ultimate outcome. In our analysis, we focus on a limited subset of data consisting of what we believe to be otherwise similar loans, with and without reduced documentation features. We can also test the effects of income exaggeration on pricing. Given that nearly all loans in our samples are ARMs, we follow the literature (see, for instance, Sa-Aadu and Sirmans, 1989, and Pennington-Cross and Ho, 2008) to measure loan pricing with the loan margin.

To isolate the influence of loan documentation type and income exaggeration from that of other factors, we employ a multiple-stage regression approach. Starting with the full sample data, in the first stage, we estimate the following loan margin regression:

$$\text{(First stage)} \quad r = \psi + \Sigma_{j=1}^{g} \, \upsilon_j Y_j + \epsilon, \tag{9}$$

where r is the loan margin; Y is a set of control variables that are expected to affect loan pricing; ψ is the intercept; $\upsilon_j(j = 1, ..., g)$ are the coefficients for control variables; and ε is the error term. Fixed effects for loan origination year are included.

In the second stage, we estimate the following two-specification models:

$$\text{(Second stage: full sample) } Rr = \psi_b + \xi_s D_s + \xi_n D_n + \varepsilon_b, \quad (10)$$

$$Rr = \psi_l + \xi_l D_l + \varepsilon_l, \quad (11)$$

where Rr is the loan margin residual from the first-stage regression; D_s, D_n, and D_l are the dummies for stated documentation, no documentation, and low documentation; ψ_b and ψ_l are the intercepts; ξ_s, ξ_n, and ξ_l are the coefficients for the documentation type dummies; and ε_b and ε_l are the error terms. In both specifications, the reference category is full documentation.

For the restricted sample, there is information on inferred borrowers' income, and we estimate three-stage regressions. The first stage is similar to regression (9). The second-stage regression is slightly modified as:

$$\text{(Second stage: restricted sample) } Rr_i = \psi_i + \chi_i Incratio + \varepsilon_i, \quad (12)$$

where Rr_i is the loan rate residual from the first-stage regression (9); *Incratio* is the inferred-MSA median income ratio; ψ_i is the intercept; χ_i is the coefficient of Incratio; and ε_i is the error term. The third stage is the following regression:

$$\text{(Third stage: restricted sample) } Rr_s = \psi_s + \xi_s D_s + \varepsilon_s, \quad (13)$$

where Rr_s is the interest rate residual from the second-stage regression (12); ψ_s is the intercept; and ε_s is the error term. The reference category is again the full documentation. Hypothesis 3 is specified as:

Hypothesis 3: Loan margin is increasing in the stated-doc, no-doc, and low-doc dummies, as well as increasing in the inferred-MSA median income ratio. In regressions (10), (11), (13), and (12), this means that ξ_s, ξ_n, $\xi_l > 0$, and $\chi_i > 0$.

This provides an interesting test for the existence of two risk premiums in loan pricing, which may help explain the motivation for allowing low documentation and income overstatement. We can measure the magnitudes of these coefficients to assess whether risk premiums were sufficient to compensate lenders for these two risks.

Testing these three hypotheses is our empirical focus. We also explore related issues such as the relationship between documentation type and other loan characteristics, the interaction between product type and documentation type, and variation across housing markets and capital market conditions.

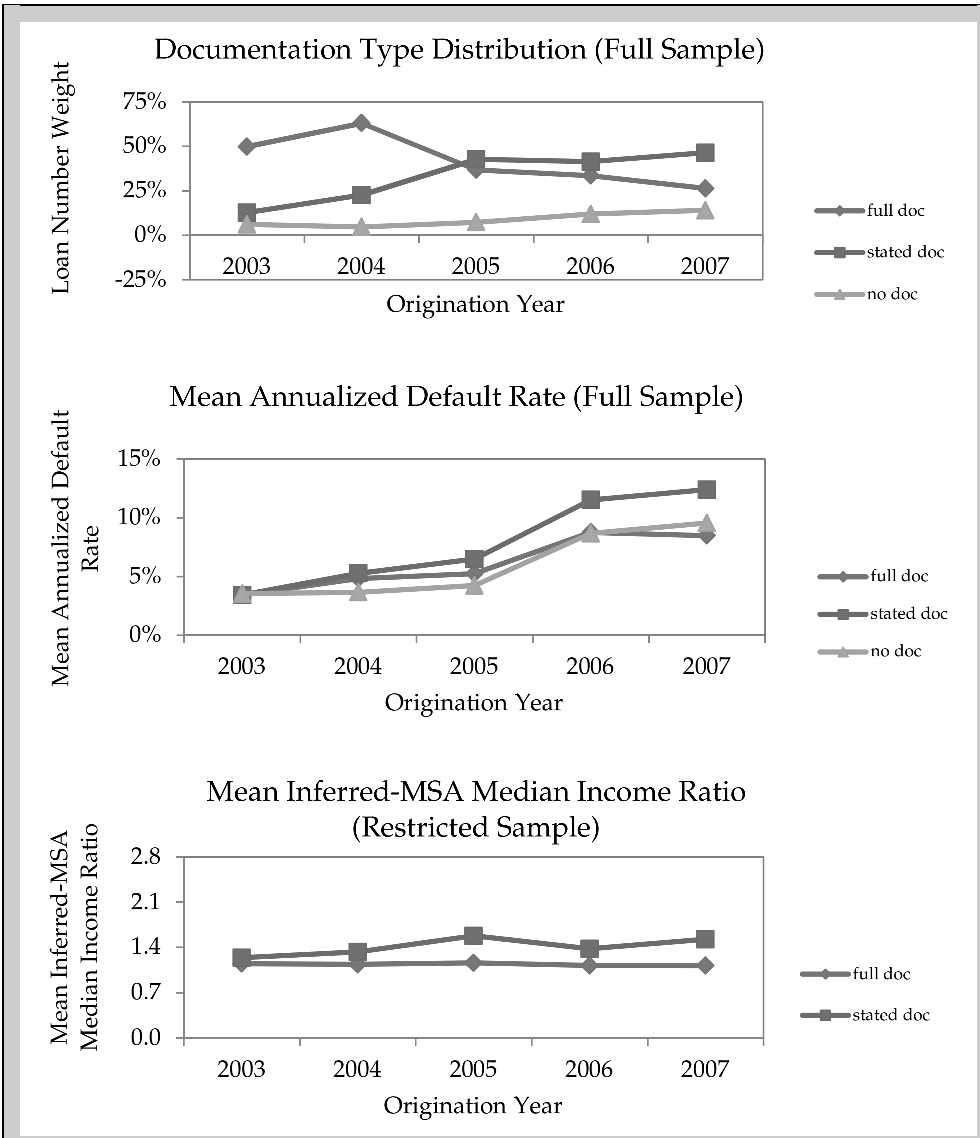
Empirical Results

Descriptive Statistic Results

We first examine time trends in the data for loans originated during our sample period. Since the years of 2000, 2001, and 2002 have fewer than 1,000 observations, we focus on the years of 2003–2007. As illustrated in Exhibit 2, 2005 is a turning point for loan origination and is when stated-doc started to replace full-doc to become the dominant category, together with the rapidly growing no-doc category. Meanwhile, stated-doc loans had become the riskiest category with mean default rates exceeding those of the other two documentation types. Our data show that 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, since stated-doc and no-doc loans had higher credit scores and lower LTVs, on average, than full-doc loans. However, our data does suggest an interaction between the growth of low documentation practices and high risk product types, including subprime loans, interest-only loans, and negative amortization ARMs. Exhibit 2 also shows that since 2001, stated-doc loans had higher ratios between borrower income and MSA median household income than full-doc loans, a pattern consistent with the possibility of income exaggeration.

As shown in Exhibit 3, based on our full sample, stated-doc is the most common loan category (39%), followed by full-doc (36%) and then no-doc (10%). Comparing different documentation types, no-doc loans have, on average, the highest credit scores, the lowest original LTV, the lowest subprime percentage, the highest ALT-A and AMP percentages, the most recent origination years, and relatively large loan size. This pattern suggests stricter underwriting standards for such loans, consistent with their lower mean default rates. In contrast, full-doc loans have the lowest borrower credit score, the highest original LTV, the highest subprime percentage, the lowest ALT-A and alternative mortgage product (“AMP”) percentages, and tend to be older and smaller in loan size. This suggests that underwriting standards are endogenous: lower quality borrowers are required to provide more documentation.[18] Perhaps as a result, this category has relatively low mean default rate as of May 2009. Finally, stated-doc loans have features in between the other two categories, except that the default rate is the highest across the three groups. These patterns suggest that lenders were able to successfully screen high- and low-quality borrowers but were less effective in underwriting the middle-range, particularly since those borrowers had the opportunity to misrepresent their income.

Exhibit 2 | Time Trend



The restricted sample provides additional information. Stated-doc loans have much higher mean income ratios to MSA than do full-doc loans, and are more concentrated in areas with higher and more rapidly increasing housing costs and areas with lower housing affordability. As expected, income ratios are higher in the fast appreciation subsample than in the slow appreciation subsample, when we divide the full sample based on whether the five-year historical growth rate in MSA HPI is at/above the sample median of 70% or not. Interestingly, faster

Exhibit 3 | Descriptive Statistics of Loans

	All Loans	Full-Doc	Stated-Doc	No-Doc	Stated vs. Full	No vs. Stated	No vs. Full
Variable	Mean	Mean	Mean	Mean	Mean Diff.	Mean Diff.	Mean Diff.
Panel A: Full sample							
Loan balance	$249,140	$209,005	$287,866	$312,291	$78,862**	$24,425**	$103,286***
Original LTV (%)	79.01	79.84	78.73	74.94	−1.11***	−3.78***	−4.90***
Origination year	2005.6	2005.4	2005.9	2005.9	0.50***	0.04***	0.54***
Cedit score	648.89	634.72	664.08	698.33	29.37***	34.25***	63.62***
Default	0.26	0.23	0.29	0.22	0.06***	−0.07***	−0.01
1-year default	0.0211	0.0174	0.0243	0.0106	0.0069***	−0.0137***	−0.0068***
Current LTV (%) (May, 2009)	77.0557	77.1969	77.5800	74.9386	0.3831***	−2.6414***	−2.2583***
FRM	0.006	0.0005	0.0001	0.0005	−0.0004***	0.0004*	0.0000
Subprime	0.52	0.66	0.41	0.05	−0.25***	−0.37***	−0.61***
ALT-A	0.38	0.20	0.53	0.90	0.33***	0.36***	0.70***
AMPS	0.36	0.21	0.50	0.70	0.29***	0.19***	0.48***
Full-doc dummy	0.36						
Stated-doc dummy	0.39						
No-doc dummy	0.10						
Loan age	38.69	41.86	35.52	34.73	−6.34***	−0.79**	−7.13***
Interest-only dummy	0.21	0.14	0.25	0.46	0.11***	0.21***	0.32***
Negative amortization ARM dummy	0.22	0.10	0.36	0.39	0.26***	0.04***	0.29***
Original LTV>80% dummy	0.40	0.46	0.36	0.17	−0.10***	−0.18***	−0.29***
Prepayment penalty dummy	0.65	0.63	0.69	0.59	0.05***	−0.10***	−0.05***
Number of loan observations	92,771	33,211	36,461	9,633			

Exhibit 3 | (continued)

Descriptive Statistics of Loans

Variable	All Loans Mean	Full-Doc Mean	Stated-Doc Mean	Stated vs. Full Mean Diff.	Loans in Fast Appreci. Area Mean	Loans in Slow Appreci. Area Mean	Fast vs. Slow Mean Diff.
Panel B: Restricted sample							
Origination year	2005.7	2005.4	2005.9	0.44***	2005.8	2005.6	−0.20***
Ln(loan balance)	12.38	12.23	12.53	0.30***	12.50	12.23	−0.27***
Original LTV (%)	79.67	79.98	78.87	−1.11***	79.20	80.21	1.02***
Credit score	642.10	634.42	658.95	24.53***	645.50	638.29	−7.21***
Default	0.27	0.23	0.30	0.07***	0.29	0.25	−0.04***
1-year default	0.0200	0.0159	0.0215	0.01***	0.0227	0.0170	−0.01***
Inferred-MSA median-income ratio	1.29	1.14	1.46	0.32***	1.39	1.18	−0.21***
Current LTV (%) (May, 2009)	77.15	77.08	77.16	0.08	76.92	77.41	0.48***
Debt-to-income ratio (front end)	25.06	25.70	23.67	−2.04***	25.60	24.45	−1.16***
MSA median household income	$53,202	$52,599	$54,154	$1,555***	$53,594	$52,762	831.70***
MSA HPI	239.06	226.12	253.31	27.19***	279.73	193.33	−86.40***
MSA ratio of household income to HPI	235.65	245.85	225.24	−20.61***	196.28	279.90	83.62***
Ln(wealth = zip code median household income * median age)	14.29	14.29	14.31	0.02***	14.29	14.29	0.00
5-year historical growth in MSA median household income	0.14	0.13	0.16	0.03***	0.18	0.10	−0.07***
5-year historical growth rate in MSA HPI	0.70	0.65	0.76	0.11***	1.01	0.36	−0.65***
Level of 30 year FRM	6.19	6.15	6.23	0.08***	6.22	6.16	−0.06***

Exhibit 3 | (continued)

Descriptive Statistics of Loans

	All Loans	Full-Doc	Stated-Doc	Stated vs. Full	Loans in Fast Appreci. Area	Loans in Slow Appreci. Area	Fast vs. Slow
Variable	Mean	Mean	Mean	Mean Diff.	Mean	Mean	Mean Diff.
Yield curve slope	1.11	1.17	1.06	−0.11***	1.06	1.16	0.10***
SP500 1-year return	0.10	0.09	0.10	0.01***	0.10	0.10	0.00***
FRM (%)	0.008%	0.008%	0.000%	−0.008%	0.000%	0.018%	0.018%**
Subprime	0.64	0.74	0.47	−0.27***	0.62	0.67	0.05***
ALT-A	0.31	0.20	0.50	0.30***	0.34	0.28	−0.07***
AMPS	0.31	0.20	0.47	0.26***	0.36	0.25	−0.11***
Full-doc dummy	0.40	1	0	−1.00***	0.36	0.45	0.09***
Stated-doc dummy	0.45	0	1	1.00***	0.50	0.40	−0.10***
Loan age	35.46	38.00	33.02	−4.99***	33.85	37.27	3.42***
Interest-only dummy	0.18	0.15	0.24	0.09***	0.21	0.14	−0.07***
Negative amortization ARM dummy	0.17	0.08	0.30	0.22***	0.20	0.14	−0.05***
Original LTV>80% dummy	0.45	0.48	0.38	−0.10***	0.42	0.47	0.05***
Prepayment penalty dummy	0.67	0.66	0.68	0.02***	0.73	0.59	−0.14***
Number of observations	60,465	24,157	27,360		31,999	28,466	

Notes:
*Significant at the 10% level.
**Significant at the 5% level.
***Significant at the 1% level.

appreciation areas also have a higher loan default rate (29% vs. 25%), as well as a higher early payment default rate (2.27% vs. 1.70%), which is apparently associated with greater usage of stated-doc loan types (50% vs. 40%).[19]

What Affects Documentation Types

We explore the relationship among doc type, loan type, and borrower characteristics. Using the full sample, we first delete loans without documentation type information, and then estimate a multinomial logit regression of doc type (stated-doc vs. full-doc and no-doc vs. full-doc) on major loan characteristics including original LTV, credit score, and whether it is an AMP. We also estimate a logit regression of low-doc on the same set of explanatory variables. For each regression, we employ two model specifications that include different sets of explanatory variables, to avoid multicollinearity. Results using all loans and non-subprime loans only are shown in Exhibit 4. As expected, borrowers with higher credit scores or lower original LTV are subjected to looser loan documentation requirements (either stated-doc or no-doc). Furthermore, low documentation requirements are also more frequent for AMPS than for traditional amortizing loans.

Main Test Results

Hypothesis 1: Low documentation creates additional default risk: Using the non-subprime loan full sample, we employ multi-stage default probability regressions (4), (5), and (6) to test Hypothesis 1. We examine whether the stated-doc, no-doc, and low-doc dummies positively affect default, that is, β_s, β_n, and $\beta_l > 0$. We develop four model specifications for the first-stage default logit regression, with the corresponding two-stage results shown in Panel A of Exhibit 5. In Specification (1), we include original LTV, credit score, loan origination year dummies, and documentation type dummies as explanatory variables. Consistent with Hypothesis 1, in the second stage, the coefficients of stated-doc, no-doc dummy, and low-doc dummy are all positive (0.019, 0.016, and 0.023). Interestingly, when we estimate the second-stage regression with stated-doc dummy only or with no-doc dummy only, the stated-doc dummy has a larger coefficient than the no-doc dummy (0.013 vs. 0.005) at a 1% significance level. The low-documentation coefficients are consistently positive, confirming that reduced loan documentation raises default risk. Other variables have the correct signs: default rates are increasing in original LTV and decreasing in credit score. In Specification (2), as a robustness test, at the first stage, we replace original LTV with a high LTV dummy (defined as original LTV $> 80\%$); in addition, we introduce additional factors such as a loan size (the natural logarithm of loan balance), a dummy variable for FRM, and a dummy variable for IO (interest-only) contract type. Previous second-stage results are unchanged, with β_s, β_n, and β_l all positive, and when we employ the second-stage regression with stated-doc dummy only or with no-doc dummy only, the coefficient on stated-doc is larger than that

Exhibit 4 | Documentation Type Regression Results: For Loans in the Full Sample (with documentation type identifiable)

	All Loans			Non-subprime Loans		
	Multinomial Logit Model		Logit Model	Multinomial Logit Model		Logit Model
	Stated-Doc versus Full-Doc	No-Doc versus Full-Doc	Low-Doc	Stated-Doc versus Full-Doc	No-Doc versus Full-Doc	Low-Doc
Variable	Coeff.	Coeff.	Coeff.	Coeff.	Coeff.	Coeff.
Specification 1						
Intercept	−6.918***	−2.428***	0.079***	−7.376***	−3.744***	0.774***
AMPS dummy	1.141***	2.082***	1.321***	1.014***	0.772***	0.938***
Origination year dummies	Yes	Yes	Yes	Yes	Yes	Yes
Specification 2						
Intercept	−10.268***	−9.839***	−2.628***	−10.215***	−4.991***	−0.117
Original LTV (%)	−0.005***	−0.021***	−0.008***	−0.009***	−0.024***	−0.014***
Credit score	0.005***	0.012***	0.006***	0.004***	0.004***	0.004***
Origination year dummies	Yes	Yes	Yes	Yes	Yes	Yes

Note: In All Loans, the number of observations is 79,305. In Subprime Loans, the number of observations is 41,918.
***Significant at the 1% level.

Exhibit 5 | Two-stage Default Regression Results: Full Sample Excluding Subprime Loans

	Specification 1	Specification 2	Specification 3	Specification 4
Variable	Coeff.	Coeff.	Coeff.	Coeff.
Panel A: All non-subprime loans				
Default Dummy Logit Regression				
Intercept	−0.080	−1.663***	3.654***	2.463***
ln (loan balance)		0.229***		
Original LTV (%)	0.016***			0.014***
Original LTV>80% dummy		0.364***	0.271***	
Credit score	−0.004***	−0.004***	−0.006***	−0.005***
FRM dummy		0.090		
IO dummy		−0.245***		
Neg. amort. ARM dummy			0.195***	
Prepayment penalty dummy				0.059**
Loan age			−0.035***	−0.037***
Origination year dummies	Yes	Yes	Yes	Yes
Default Probability Residual Regression				
Intercept	−0.021***	−0.020***	−0.017***	−0.018***
Low document dummy	0.023***	0.022***	0.018***	0.019***
Intercept	−0.016***	−0.016***	−0.009***	−0.009***
Stated document dummy	0.019***	0.018***	0.008***	0.008***
No document dummy	0.016***	0.017***	0.008***	0.008***
Intercept	−0.010***	−0.009***	−0.006***	−0.006***
Stated document dummy	0.013***	0.012***	0.004***	0.005***
Intercept	−0.005***	−0.005***	−0.004***	−0.004***
No document dummy	0.005***	0.006***	0.003***	0.003***
Stated doc–no doc	0.007***	0.006***	0.001***	0.002***
Number of observations	44,956	44,956	44,956	44,956

Panel B: By subsample

	Loans with Prepay Option Value Info	IOs	Non-IOs	NA ARMs	Non-NA ARMs
Variable	Coeff.	Coeff.	Coeff.	Coeff.	Coeff.
Default Dummy Logit Regression					
Intercept	−0.376	−0.047	−0.034	0.080	0.076***
Original LTV (%)	0.019***	0.016***	0.015***	0.021***	0.013***
Credit score	−0.004***	−0.004***	−0.004***	−0.005***	−0.005***
Prepayment option value	1.272***				
Origination year dummies	Yes	Yes	Yes	Yes	Yes

Exhibit 5 | (continued)

Two-stage Default Regression Results: Full Sample Excluding Subprime Loans

Variable	Loans with Prepay Option Value Info Coeff.	IOs Coeff.	Non-IOs Coeff.	NA ARMs Coeff.	Non-NA ARMs Coeff.
Default Probability Residual Regression					
Intercept	−0.007***	−0.007***	−0.010***	−0.010***	−0.006***
Stated documentation dummy	0.006***	0.008***	0.014***	0.010***	0.009***
Intercept	−0.006***	−0.006***	−0.004***	−0.006***	−0.005***
No documentation dummy	0.008***	0.011***	0.003**	0.009***	0.006***
Intercept	−0.023***	−0.022***	−0.020***	−0.026***	−0.016***
Low documentation dummy	0.023***	0.023***	0.023***	0.026***	0.018***
Number of observations	14,191	15,658	29,298	20,425	24,531

Panel C: By loan origination year

Variable	2003 Coeff.	2004 Coeff.	2005 Coeff.	2006 Coeff.	2007 Coeff.
Default Dummy Logit Regression					
Intercept					
Intercept	2.721***	1.259	−1.492***	0.564**	−0.023***
Original LTV (%)	0.005	0.010*	0.022***	0.010***	0.024***
Credit score	−0.007***	−0.006***	−0.003***	−0.004***	12.170
Default Probability Residual Regression					
Intercept	−0.015***	−0.010***	−0.019***	−0.024***	
Low documentation dummy	0.018***	0.016***	0.022***	0.026***	
Number of observations	2,333	2,089	11,279	15,780	

Notes: The first stage is a default dummy logit regression on explanatory variables excluding the documentation type dummies. In the second stage, the residual from this regression is converted into default probability residual, which is regressed on the documentation type dummies. By definition, "prepayment option value" equals one minus the matching loan rate on May 2009/initial loan rate.

*Significant at the 10% level.

**Significant at the 5% level.

***Significant at the 1% level.

on no-doc. In Specification (3), as an alternative robustness test, at the first stage, we replace the original year dummies with loan age and include an additional risk factor, a dummy for negative amortization (to capture pay-option ARM contracts). Finally, in Specification (4), we control a prepayment penalty dummy. These two specifications generate essentially identical second-stage results as those previously discussed. Collectively, the results strongly support Hypothesis 1: reduced documentation boosts default risk.

Given regression results, we can simulate default rates. For instance, at the non-subprime loan sample mean of borrower characteristics (original LTV of 77%, credit score of 696, prepayment penalty dummy 0.57, and loan age of 37 months and using Specification (4) coefficients), we find that the average loan will have a 16.08% default rate if fully documented, increasing to 17.79% if with stated documentation, and 17.76% if with no documentation. This implies an 11% (or 10%) increase in default risk by switching from full-doc to stated-doc (or no-doc). These results are lower than the estimates in Pennington-Cross and Ho (2010). One possible reason for this difference may be the sample difference. Their results are for subprime loans, while ours are for non-subprime loans. In addition, our sample excludes loans where piggyback seconds were used (where CLTV $>$ LTV at origination). Such borrowers may be financially stronger than those using piggyback loans. Also important could be our decomposition of documentation type effects from other effects.

It is widely accepted in the mortgage literature that default and prepayment are competing risks that should be jointly modeled (e.g., Deng, Quigley, and Van Order, 2000; Pennington-Cross and Ho, 2010). Hazard or conditional probability methods are often adopted to do so.[20] In our study, unfortunately, we have a cross-sectional data (rather than a panel data) that prevents us from directly observing prior prepayments. As a result, our data may have survivorship bias since the default rates we observe reflect the default risk of un-prepaid loans rather than for all loans originated. To address this issue, and given that the majority of the loans in our sample are ARMs, we retrieve a subsample of 14,191 five-year ARMs with stated doc features that were still in the fixed-rate periods as of May 2009. We use the five-year ARMs because other ARMs types had very few observations that were still in their fixed rate period. We calculate a proxy for each loan's prepayment option value as of May 2009 following the method in Firestone, Van Order, and Zorn (2007), where ''prepayment option value'' equals 1 minus the matching loan rate on May 2009 divided by the initial loan rate. We then add this prepayment option value into every specification of the default rate regression as a control variable. The results of Specification (1) are reported in Panel B of Exhibit 5, which confirm that reduced documentation is associated with a high default risk. Similar results are generated from other regression specifications.[21] In a robustness test for controlling the competing risk, we also follow Firestone, Van Order, and Zorn (2007) and segment loans into different groups based on the size of the prepayment option value proxy. We estimate a default rate regression for each segment and find that the documentation effect on default risk exists across segments.

Given the large percentage of AMPs in the stated-doc and no-doc categories, it is interesting to further explore contract type-documentation type relationship. To do so, we develop a two-stage default risk regressions for several AMPs subtype samples: IOs versus non-IOs, and negative amortization versus standard amortization ARMs. Panel B of Exhibit 5 shows that the default effects for stated-doc, no doc, and low-doc are larger for negative amortization loans than standard amortization ARMs. This result is consistent with our expectation that low documentation has a stronger effect when coupled with other risky loan contract features. This phenomenon has come to be known in the industry as risk-layering.

Finally, our large sample size allows us to explore temporal variation in the documentation-default relationship. To do this, we divide the full sample into subsamples by loan origination year and estimate a two-stage default regression for each subsample. Again, due to too few observations in the subsamples from 2000, 2001 and 2002, we focus on 2003–2007, reporting results in Panel C of Exhibit 5. From the magnitude of the positive coefficient of the low-documentation dummy, the effect of low documentation on default seemed to slightly decline from 2003 to 2004 (with the coefficient declining from 0.018 to 0.016), but rapidly increase in 2005 (0.022) and 2006 (0.026) and remain high at 0.024 in 2007. This indicates increasing risk associated with low-documentation loans leading up to the onset of the financial crisis.

Hypothesis 2: Low documentation combined with income overstatement leads to a more pronounced default risk increase: From the restricted sample we can extract subsamples of full-doc and stated-doc loans. We use these subsamples to estimate two-stage regressions (7) and (8) to test Hypothesis 2 that default rate is increasing in the income ratio for the stated-doc subsample, but that any such effect is not as large in the full-doc subsample. The results appear in Panel A of Exhibit 6. Again, at the first stage, 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 six model specifications, each including a different set of explanatory variables to avoid multicollinearity. Across all specifications, the second-stage results support Hypothesis 2, with the coefficient of the natural logarithm of the income ratio positive for the stated-doc loan subsample, while insignificant for the full-doc loan subsample, with the difference consistently significant at a 1% level.

Examining the other risk factors, stated-doc loan default rates appear generally to be less sensitive to the current LTV ratio (May, 2009) (+), the credit score (−), the negative amortization ARM dummy (+), and the existence of prepayment penalty (+).[22]

We also explore time variation in the income exaggeration-default relationship by employing a two-stage default regression for stated-doc loans originated in each of 2005, 2006, and 2007 (and excluding other years due to few observations). As shown in Panel B of Exhibit 6, the impact of the income ratio increases year over year.

Exhibit 6 | Two-stage Default Regression Results: Restricted Sample Excluding Subprime Loans

	Specification 1			Specification 2			Specification 3		
	Full-Doc	Stated-Doc		Full-Doc	Stated-Doc		Full-Doc	Stated-Doc	
Variable	Coeff.	Coeff	Diff.	Coeff.	Coeff	Diff.	Coeff.	Coeff	Diff.
Panel A: Stated-doc vs. full-doc loans									
Default Dummy Logit Regression									
Intercept	1.220*	0.299	−0.922*	0.278	0.197	−0.081	−3.523***	−2.257***	1.266***
Credit score	−0.006***	−0.004***	0.002***	−0.005***	−0.003***	0.002**	−0.006***	−0.004***	0.001*
Original LTV (%)	0.021***	0.022***	0.001	0.020***	0.020***	0.001			
Current LTV (%) (May, 2009)				0.050***				0.035***	−0.015***
IO dummy				−0.317***	−0.231***	0.085			
Yield curve slope				−0.050	−0.751**	−0.701**			
MSA HPI							0.004***	0.003***	−0.001
MSA ratio of household income to HPI	−0.004***	−0.004***	0.000						
SP500 1-year return	−0.309	0.566	0.875	0.025	0.601	0.576	−0.512	0.517	1.029
Origination year dummies	Yes	Yes	Yes		Yes		Yes	Yes	
Default Probability Residual Regression									
Intercept	−0.002**	−0.00***	−0.005	−0.002**	−0.006***	−0.004	−0.002**	−0.006***	−0.004
Ln (inferred-MSA median-income ratio)	−0.001	0.007***	0.008***	0.000	0.009***	0.009***	−0.001	0.009***	0.010***
Number of observations	6,304	14,401		6,304	14,401		6,304	14,401	

Exhibit 6 | (continued)

Two-stage Default Regression Results: Restricted Sample Excluding Subprime Loans

	Specification 4			Specification 5			Specification 6		
	Full-Doc	Stated-Doc		Full-Doc	Stated-Doc		Full-Doc	Stated-Doc	
Variable	Coeff.	Coeff	Diff.	Coeff.	Coeff	Diff.	Coeff.	Coeff	Diff.
Default Dummy Logit Regression									
Intercept	−2.831***	−1.781***	1.050***	0.143	−1.498*	−1.641	2.535***	2.302***	−0.233***
Credit score	−0.006***	−0.004***	0.001*	−0.006***	−0.004***	0.002**	−0.007***	−0.003***	0.004***
Original LTV (%)				0.020***	0.022***	0.002	0.016***	0.012***	−0.005
Current LTV (%) (May, 2009)	0.049***	0.035***	−0.014***						
5-year historical growth rate in MSA HPI	0.660***	0.537***	−0.124						
Neg. amort. ARM dummy				0.812***	0.459***	−0.353***			
Prepayment penalty dummy							0.180**	0.010	−0.169*
Level of 30 Year FRM				−0.022	0.140	0.162			
SP500 1-year return	−0.421	0.624	1.044	−0.080	0.888	0.968			
Loan age							−0.035***	−0.070***	−0.035***
Origination year dummies	Yes	Yes		Yes	Yes				
Default Probability Residual Regression									
Intercept	−0.002**	−0.006***	−0.004	−0.002**	−0.006***	−0.005*	−0.002**	−0.006***	−0.004*
Ln (inferred-MSA median-income ratio)	0.000	0.009***	0.010***	0.001	0.011***	0.009***	0.001	0.010***	0.010***
Number of observations	6,304	14,401		6,304	14,401		6,304	14,401	

Exhibit 6 | (continued)

Two-stage Default Regression Results: Restricted Sample Excluding Subprime Loans

	Non-subprime Stated-Doc			Stated-Doc Loans		
	2005	2006	2007	Non-subprime	Subprime	
Variable	Coeff.	Coeff.	Coeff.	Coeff.	Coeff.	Diff.
Panel B: For stated-doc loans only						
Default Dummy Logit Regression						
Intercept	0.390	0.814	0.126	0.299	0.359	0.060
Credit score	−0.004***	−0.003***	−0.004***	−0.004***	0.000	0.003***
Original LTV (%)	0.022***	0.014***	0.031***	0.022***	0.001	−0.021***
MSA ratio of household income to HPI	−0.003***	−0.004***	−0.006***	−0.004***	−0.003***	0.002***
SP500 1-year return	−0.649	0.158	1.577*	0.566	−1.067**	−1.633**
Origination year dummies	Yes	Yes	Yes	Yes	Yes	
Default Probability Residual Regression						
Intercept	−0.005***	−0.005***	−0.006***	−0.005***	−0.006***	−0.001***
Ln (inferred-MSA median-income ratio)	0.005***	0.006***	0.008***	0.007***	0.035***	0.028***
Number of observations	3,491	5,561	5,103	14,401	12,959	

The first stage is a default dummy logit regression on explanatory variables excluding the natural logarithm of inferred-MSA median-income ratio. In the second stage, the residual from this regression is converted into default probability residual, which is regressed on the natural logarithm of inferred-MSA median-income ratio.

*Significant at the 10% level.

**Significant at the 5% level.

***Significant at the 1% level.

So far we have confined our analysis to non-subprime loans, the market segment customarily referred to as Alt-A, where reduced documentation was originally developed. We now turn to the subprime segment to see whether default sensitivity is more pronounced there. We replicate the default risk regressions in Panel A of Exhibit 6 for subprime stated-doc loans and compare results to those for non-subprime stated-doc loans. The last few columns of Panel B provide an obvious contrast: across model specifications, the sensitivity of default to income ratio is much greater for subprime than for non-subprime (with the coefficient of the income ratio in the default risk regression around 0.035 versus around 0.007 and the difference consistently significant at a 1% level). According to industry publications, the use of reduced documentation migrated from Alt-A to subprime starting roughly in 2005, helping to produce the market meltdown in 2007. Our results confirm the role of income exaggeration in the recent subprime market meltdown.

Robustness Checks

We conduct several robustness checks on our default regression results, some of which are reported in Exhibit 7. In one robustness test, we check the statistics and re-estimate all the default risk regressions in Exhibits 5 and 6, re-defining ‘‘default’’ as early payment default (‘‘EPD’’), a dummy equal to 1 if the borrower defaults within one year after the loan origination. The results are entirely consistent with prior results for default at any point in loan life.

As an additional robustness test, we use an alternative method to estimate income exaggeration. Borrowing ideas from Jiang, Nelson, and Vytlacil (2009), we estimate borrower income for each stated-doc loan borrower using the full-doc loan data. We then replace the income proxy we employ earlier (that is, the local MSA median income) with this predicted value, and use that ratio to measure income overstatement. Even with this substitution, previous results for stated-doc loans remain unchanged.

We also conducted further robustness tests to address the issue of competing risks between default and prepayment. One such test is to create and control for prepayment option value as of May 2009 as we did in Panel B of Exhibit 5, for a subsample of 7,294 five-year ARMs (with stated documentation features) from the restricted sample. As shown in Exhibit 7, the positive effect of income overstatement on default risk remains after this control variable is introduced.

In an alternative robustness test to control the prepayment risk, we re-estimate the default rate regressions on loans originated during only 2006–2007. Falling housing prices and tighter underwriting during 2007–2009 limited borrowers opportunities to prepay, especially after the subprime and Alt-A market segments collapsed in late 2007, so prepayment would have been a very rare event. Again, results for loans originated during this much shorter window are consistent with what we obtain using all originations during 2003–2007, indicating that our results are not affected by survivorship bias.

Exhibit 7 | Selected Results from Robustness Checks for the Loan Default Rate Regressions: Non-subprime Stated-Doc Loans in the Restricted Sample

	Reg. 1	Reg. 2	Reg. 3	Reg. 4	Reg. 5	Reg. 6
Variable	Coeff.	Coeff.	Coeff.	Coeff.	Coeff.	Coeff.
Default Dummy Logit Regression[a]						
Intercept	2.657*	0.401	0.232	0.318	0.173	0.209
Credit score	0.004**	−0.004***	−0.004***	−0.004***	−0.004***	−0.004***
Original LTV (%)	−0.029***	0.022***	0.026***	0.022***	0.023***	0.022**
MSA ratio of household income to HPI	0.008***	−0.005***	−0.005***	−0.005***	−0.003***	−0.004***
SP500 1-year return	−8.402***	0.657	2.308**	1.066*	−1.071	0.540
Prepayment option value			1.111***			
Ln (inferred-MSA median-income ratio)						0.195***
Origination year dummies	Yes	Yes	Yes	Yes	Yes	Yes
Default Probability Residual Regression						
Intercept	−0.182***	−0.004***	−0.006***	−0.006***	−0.005***	
Ln (inferred-MSA median-income ratio)	0.009***	0.008***	0.010***	0.008***	0.005***	
Number of observations	14,401	13,587	7,294	10,664	3,737	14,401

Exhibit 7 | Selected Results from Robustness Checks for the Loan Default Rate Regressions: Non-subprime Stated-Doc Loans in the Restricted Sample

Notes: The first stage is a default rate logit regression on explanatory variables excluding the natural logarithm of inferred-MSA median-income ratio. In the second stage, the residual from this regression is converted into default probability residual, which is regressed on the natural logarithm of inferred- MSA median-income ratio. In Regression 1, the dependent variable is changed into the early payment default dummy, which is 1 if the loan is defaulted within one year. In regression 2, the inferred-MSA median-income ratio is replaced by the inferred-correct income ratio, where the correct income for each stated-doc loan borrower is estimated using the full-doc loan borrower income regression, in which the explanatory variables include the square of credit score, wealth, appraisal value, MSA-level median household income, origination year dummy and state dummy. Sample size is shrunken due to the missing data on appraisal value, and the fact that the inferred-correct income is limited to range [20,000, 500,000]. Regression 3 is for loans with prepayment option value information available. By definition, "prepayment option value" equals one minus the matching loan rate on May 2009 / initial loan rate. In Regression 4, the loans were generated during 2006–2007. In Regression 5, the loans were generated during 2000–2005. Regression 6 is a one-stage default dummy logit regression, where the explanatory variables include Ln(inferred-MSA median-income ratio) and other variables.

[a] The dependent variable is the default dummy in Regressions 2 to 6, while the Early Payment Default (EPD), is defined as default within one year, in Regression 1.

*Significant at the 10% level.

**Significant at the 5% level.

***Significant at the 1% level.

Finally, we also estimate a single-stage regression, where loan documentation dummies or income-overstatement proxies are included together with other determinants of default risk. We obtain results consistent with the previous multi-stage regression results.

Hypothesis 3: Low documentation risk and income overstatement risk are priced with low-risk premiums: Thus far we have shown that these loans are risky and performed poorly after controlling for other risk factors. But were they appropriately priced for that risk?

To address this question, we examine risk premiums for stated documentation loans and high income ratio cases by employing multiple-stage loan pricing regressions (9) to (13). For loan pricing, we need to further restrict the sample to otherwise comparable products. We select one of the most frequently occurring loan types in our sample, the one-year ARM. After excluding subprime loans, we construct a subsample of the full sample that consists of 12,083 one-year ARM loans with full-doc, stated-doc, or no-doc characteristics; and a subsample from the restricted sample that consists of 6,256 one-year ARM loans with either full-doc or stated-doc.

Using the full sample, we regress in the first stage the gross margin of the loan on a set of pricing factors such as credit score, varied LTV measures, prepayment penalty dummy, and loan contract type. In the second stage, we regress the residual from the first-stage regression on low-documentation dummies including stated-doc, no-doc, and low-doc dummies. The risk premium for low documentation should be reflected in the coefficients for these dummy variables. As shown in Exhibit 8, results with varying specifications produce a coefficient of around 7% to 10% for the stated-doc dummy, a coefficient around 5% to 8% for the no-doc dummy, and a coefficient around 8% to 9% for the low-doc dummy, implying a rate differential of 5 to 10 basis points for low documentation on a risk-adjusted basis. Surprisingly, the loan margins are lower for no-doc loans than for stated-doc loans. Such a pattern could arise if no-doc loans were restricted to only the highest credit score borrowers.

Using the restricted subsample, the first-stage rate regression can include additional factors such as debt-to-income ratio, local housing market conditions, and capital market factors; in the second stage, the residual from the first-stage regression is again regressed on the income ratio, the coefficient of which captures the risk premium for income overstatement. In the third stage, the residual from the second-stage regression is regressed on the stated-doc dummy, the coefficient of which should capture the low documentation risk premium. The results are reported in Panel A of Exhibit 9. The stated-doc risk premium is around 6–7 basis points, and the natural logarithm of inferred-MSA median income ratio generates a coefficient of around 3% to 4%. The results support Hypothesis 3 that the loan price is increasing in stated documentation loans and the income ratio. So there is evidence that lenders priced the additional risk of low documentation and income exaggeration; however, the magnitude of these risk premiums is quite

Exhibit 8 | Two-stage Loan Margin Regression Results

	Regression 1	Regression 2	Regression 3	Regression 4
Variable	Coeff.	Coeff.	Coeff.	Coeff.
First-Stage Loan Margin Regression				
Intercept	3.263***	3.263***	3.268***	3.205***
Ln (loan balance)		0.011*	0.011*	
Original LTV (%)	0.012***		0.012***	0.012***
Original LTV>80% dummy		0.435***		
Credit score	−0.001***	−0.0009***	−0.001***	−0.001***
Prepayment penalty dummy		0.074***		0.073***
Ln (wealth)	−0.080***	−0.082***	−0.080***	−0.077***
IO dummy		0.049***		
Neg. amort. ARM dummy			−0.023	
Origination year and month dummies	Yes	Yes	Yes	Yes
Second-Stage Loan Margin Residual Regression				
Intercept	−0.053***	−0.051***	−0.054***	−0.054***
Stated document dummy	0.077***	0.074***	0.078***	0.078***
Intercept	−0.075***	−0.083***	−0.075***	−0.074***
Stated document dummy	0.099***	0.106***	0.099***	0.098***
No document dummy	0.053***	0.078***	0.052***	0.048***
Intercept	−0.068***	−0.074***	−0.068***	−0.067***
Low document dummy	0.082***	0.090***	0.082***	0.081***

Notes: The first stage is a loan margin regression on explanatory variables excluding the documentation type dummies. In the second stage, the residual from this regression is regressed on the stated documentation dummy and the no documentation dummy, or on the low documentation dummy (which is 1 if either stated-doc or no-doc). The data are based on 12,083 loans in the full sample that are one-year ARMs, excluding subprime loans.
*Significant at the 10% level.
**Significant at the 5% level.
***Significant at the 1% level.

small. These results are broadly consistent with LaCour-Little and Yang (2010), who report similarly small risk premiums for IO and other deferred amortization loan types. Linking this to the high default risk of low-doc loans, we infer that more accurate risk-based pricing would have reduced default rates.[23]

Finally, we estimate a three-stage pricing regression for each of 2005, 2006, and 2007. As shown in Panel B of Exhibit 9, the effect of underpricing for stated documentation is most pronounced in 2007 (with the coefficient of the stated-doc dummy equal to an insignificant 0.012, versus a significant 0.058 in 2005 and a significant 0.105 in 2006). The result suggests that low pricing might be an important contributor to the collapse of Bears Stearns starting in 2007.

Exhibit 9 | Three-Stage Loan Margin Regression Results

	Regression 1	Regression 2	Regression 3	Regression 4	Regression 5	Regression 6
Variable	Coeff.	Coeff.	Coeff.	Coeff.	Coeff	Coeff
Panel A: Low documentation risk premium						
First-Stage Loan Margin Regression						
Intercept	3.295***	3.626***	3.199***	3.395***	3.568***	4.742***
Original LTV (%)	0.014***	0.015***	0.015***	0.015**	0.014***	0.015***
Original LTV>80% dummy						
Credit score	−0.001***	−0.001***	−0.001***	−0.001***	−0.001***	−0.001***
Prepayment penalty dummy						0.059***
Debt to income ratio (front end)						−0.004***
Ln (wealth)						−0.083***
MSA ratio of household income to HPI		−0.001***				
MSA HPI			−0.001***			
5-year historical growth rate in MSA HPI				0.102***		
IO dummy	0.037***					
Neg. amort. ARM dummy					−0.095***	
Level of 30 Year FRM					0.003	
Yield curve slope	0.214***					
SP500 1-year return	0.130	0.160	0.142	0.174	0.163	
Origination year and month dummies	Yes	Yes	Yes	Yes	Yes	Yes
Second-Stage Loan Margin Residual Regression						
Intercept	−0.009	−0.006	−0.006	−0.007	−0.009	−0.007
Ln (inferred-MSA median-income ratio)	0.041***	0.029***	0.028***	0.034***	0.041***	0.032***
Third-Stage Loan Margin Residual Regression						
Intercept	3.468***	3.465***	3.465***	3.466***	3.468***	3.466***
Stated document dummy	0.063***	0.066***	0.067***	0.065***	0.063***	0.065***

Exhibit 9 | (continued)

Three-Stage Loan Margin Regression Results

Variable	2005 Coeff.	2006 Coeff	2007 Coeff
Panel B: By year			
First-Stage Loan Margin Regression			
Intercept	4.705***	−0.745	2.897***
Original LTV (%)	0.003***	0.002**	0.028***
Credit score	0.0002***	0.0007***	−0.004***
IO dummy	0.223***	0.032***	0.023***
Yield curve slope	−1.012***	3.942***	2.771***
SP500 1-year return	−5.718***	4.106***	−14.126***
Origination year & month dummies	Yes	Yes	Yes
Second-Stage Loan Margin Residual Regression			
Intercept	−0.006	−0.002	−0.007
Ln (inferred-MSA median-income ratio)	0.017*	0.028**	0.032**
Third-Stage Loan Margin Residual Regression			
Intercept			
Intercept	3.181***	3.266***	3.873***
Stated document dummy	0.058***	0.105***	0.012

Exhibit 9 | (continued)

Three-Stage Loan Margin Regression Results

Notes: The first stage is a loan margin regression on explanatory variables excluding the documentation type dummies and the natural logarithm of inferred-MSA median-income ratio. In the second stage, the residual from this regression is regressed on the natural logarithm of inferred-MSA median-income ratio. In the third stage, the residual from the second regression is regressed on the stated documentation dummy. The data are based on 6,256 loans in the restricted sample that are one-year ARMs, excluding subprime loans and no-doc loans. In 2005, the number of observations is 1,871; in 2006, the number of observations is 1,905; in 2007, the number of observations is 2,428.

*Significant at the 10% level.

**Significant at the 5% level.

***Significant at the 1% level.

In summary, regression the results provide strong support for our three major hypotheses, confirming the importance of loan documentation on default risk.

Conclusion

In this paper we have examined the effects of documentation type on default risk. Although loan documentation requirements have changed significantly in recent years, their contribution to increasing rates of residential mortgage default has not been rigorously analyzed. We believe our study is among the first to comprehensively examine these issues. We do so using a large database of home purchase loans securitized by Bear Stearns over the period 2000–2007 and observed up through May 2009.

We find that reduced documentation does increase the likelihood of loan default after controlling for other risk factors. The problem is particularly acute for stated documentation loans, which are offered to lower quality borrowers (as measured by credit score and LTV) compared to no documentation loans, for which higher credit scores and lower LTV ratios mitigate some of the incremental risk. Simulation based on our default regression models suggests roughly a 10% increase in default risk when a loan with average characteristics switches from full documentation to stated documentation, after controlling other observable risk factors. This is almost surely a lower bound estimate, as we have eliminated loans with piggyback seconds and overall default rates are high in our data even for full documentation loans.

It appears that the reason that these loans to mid-quality borrowers perform worse than no documentation loans is that lenders allowed borrowers to simply state, as opposed to verify, income or assets, while not allowing lower quality borrowers to do so. We also find evidence of income exaggeration in the stated documentation category and show that the degree of income exaggeration is also related to default risk. Finally, in terms of pricing, we estimate that the roughly 10 basis point premiums associated with low documentation lending were not sufficient given the increased risk.

Further research efforts on this topic should involve measuring local area income at a finer level of geography, relating stated income to stated occupation, replicating the analysis using multi-lender data, and estimating competing risks using a more complete panel-data format.

Appendix

A. Loans Underwritten by Bear Stearns

	Bear Stearns' Market Share and MBS Industry Ranking								Dollar % of Prime/Non-subprime Loans in All Loans Underwritten		
	All Non-agency		Non-agency Prime		Alt A		Subprime		Non-agency Prime	Non-agency Prime	Non-subprime
Year	Share	Rank	Market Share	Rank	Market Share	Rank	Market Share	Rank	National	Bear Stearns	Our Sample
Panel A: Market share and loan quality											
2000	13.40%	2			16.90%	3	12.40%	3			68.95%
2001	19.90%	1			24.70%	1	10.90%	4			78.60%
2002	12.30%	3	17.30%	1	17.20%	3	<5.7%	>10	55.30%	78.02%	81.66%
2003	10.70%	2	15.70%	1	21.10%	1	<4.5%	>10	53.16%	78.13%	61.40%
2004	11.90%	1	19.10%	1	21.30%	1	5.50%	6	45.36%	72.62%	51.76%
2005	11.00%	1	14.00%	1	15.10%	1	7.50%	6	51.46%	65.75%	74.95%
2006	9.00%	3	12%*	1	15.10%	1	4.70%	10	51.04%	68.06%	72.63%
2007	9.00%	2	10.4%**	1	10.90%	3	6.50%	6	60.83%	70.04%	79.79%

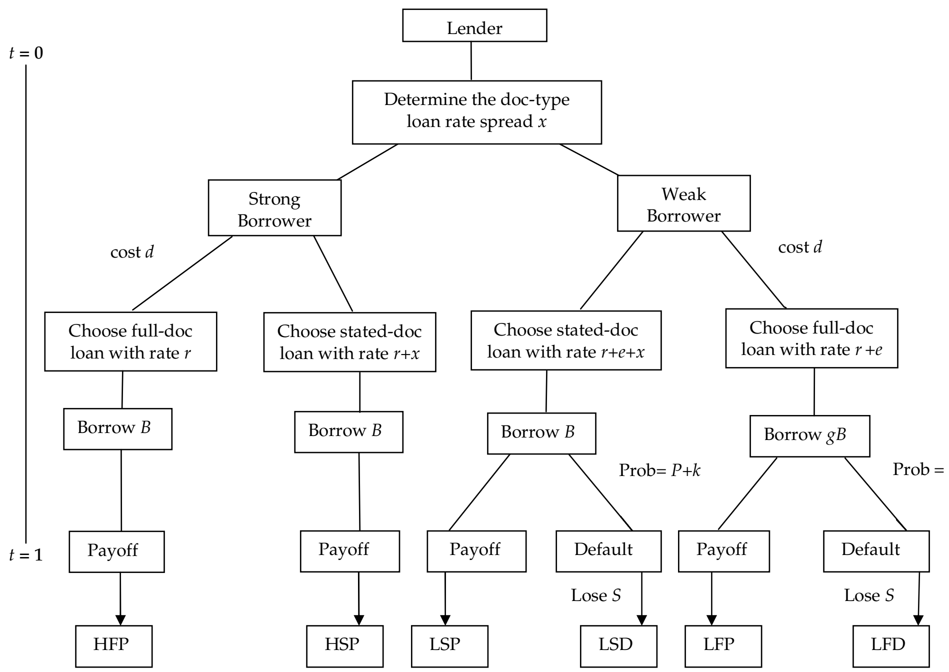
Appendix (continued)

Panel B: Loan distribution across year and across the loan original state, based on loan volume

Year	All Non-agency MBS	Non-agency MBS Underwritten by Bear Stearns	All Loans in Our Sample	Non-subprime Loans in Our Sample	State	HMDA	All Loans in Our Sample	Non-subprime Loans in Our Sample
2000	1.37%	1.70%	0.31%	0.36%	California	24.24%	22.00%	27.00%
2001	3.10%	5.70%	0.56%	0.83%	Florida	6.69%	14.11%	15.22%
2002	8.06%	9.15%	1.07%	1.71%	New York	4.36%	3.26%	2.52%
2003	11.41%	11.28%	5.06%	5.19%	Texas	3.91%	5.17%	3.84%
2004	16.82%	18.53%	7.07%	4.65%	Illinois	4.64%	4.31%	2.11%
2005	23.19%	23.54%	24.57%	25.09%	New Jersey	3.56%	2.58%	2.82%
2006	22.30%	18.60%	40.03%	35.10%	Other states	52.61%	48.57%	46.49%
2007	13.76%	11.50%	21.34%	27.07%				
Sum	100.00%	100.00%	100.00%	100.00%	Sum	100.00%	100.00%	100.00%

Notes: The data are from the 2011 Mortgage Market Statistical Annual.
*The data is the Bear Stearns' market share in the 2006 non-agency prime/Alt A MBS underwriting business.
**The data is the Bear Stearns' market share in the 2007 non-agency prime/Alt A MBS underwriting business.

B. Decision Flows



In this one-period game, at time $t = 0$, the lender decides on and announces the loan rate spread associated with the documentation type, x. Observing the doc-type loan rate spread x, the base loan rate r and the credit risk premium e, each borrower chooses the loan documentation type. The strong borrower can either pay a documentation fee d to get a full-doc loan at rate r, or choose not to pay documentation fee, hence get a stated-doc loan at rate $r + x$, both with a loan amount B. Similarly, the low-credit borrower can either pay d to get a full doc loan with amount gB at rate $r + e$, or pay nothing and get a stated-doc loan with amount B at rate $r + e + x$, where $g\epsilon[0,1]$. At time $t = 1$, the high-credit borrower will pay off the loan principal and interest no matter what documentation type she has chosen earlier; the low-credit borrower will default the loan obligations with probabilities P_f and P_s, respectively, under the full-doc loan and the stated-doc loan, and a default will bring the borrower a lump-sum loss S. The outcome nodes for the high-credit borrower are HFP (payoff under full-doc) and HSP (payoff under stated-doc), and the outcome nodes for the low-credit borrower are LSP (payoff under stated-doc), LSD (default under stated-doc), LFP (payoff under full-doc), and LFD (default under full-doc).

C. Proof for Proposition 1

The game involves two sequential decisions both made at $t = 0$: (1) the lender's choice of documentation risk premium x, and (2) the two borrowers' loan documentation type choices. We solve the game with backward induction, starting from the borrowers' documentation type decisions. The strong borrower compares her total cash outflow under full-doc, C_f^S, and that under stated-doc, C_s^S,

$$C_f^S = (1 + r)B + d, \tag{14}$$
$$C_s^S = (1 + r + x)B, \tag{15}$$

and will choose stated-doc if:

$$C_s^S < C_f^S \rightarrow x < x^S = \frac{d}{B}. \tag{16}$$

Similarly, the weak borrower compares her total cash outflow under the full-doc, C_f^W, and that under the stated-doc, C_s^W,

$$C_f^W = (1 - P)(1 + r + e)gB + PS + d, \tag{17}$$
$$C_s^W = (1 - P - k)(1 + r + e + x)B + (P + k)S, \tag{18}$$

and will choose stated-doc if:

$$C_s^W < C_f^W \rightarrow x < x^W = \frac{d - kS + [k - (1 - g)(1 - P)](1 + r + e)B}{(1 - P - k)B}, \tag{19}$$

Comparing the doc risk premiums that make the two types of borrowers indifferent to the documentation type choices, x^S and x^W, we find:

$$x^W - x^S = \frac{(P + k)d - kS + [k - (1 - g)(1 - P)](1 + r + e)B}{(1 - P - k)B} \tag{20}$$
$$\rightarrow x^W > x^S > 0 \text{ when } k > (1 - g)(1 - P) \tag{21}$$

(given loan amount B much larger than other parameters).

The result $x^W > x^S$ means that if x is low enough to attract the strong borrower, that is, $x < x^S$, it must also attract the weak borrower, as $x < x^S < x^W$. However, the opposite is not necessary true. We hence obtain the first three columns of the table in Proposition 1.

After solving for the borrowers' doc-type decisions, we move backward to solve for the lender's decision on the size of the documentation risk premium x. The lender will choose x to maximize his expected rate of return from offering the two loans, $E(R)$, which may take three forms under the condition $k > (1 - g)(1 - P)$,

$$E(R) = \max_{(x)} \begin{cases} E(R)_1 = \dfrac{1}{2B}[(1 + r + x)B \\ \quad + (1 - P - k)(1 + r + x + e)B - 2B], \quad \text{if } x \in (-\infty, x^S], \\ E(R)_2 = \dfrac{1}{2B}[(1 + r)B + (1 - P - k)(1 + r + x + e)B \\ \quad - B - (1 + u)B], \quad \text{if } x \in (x^S, x^W], \\ E(R)_3 = \dfrac{1}{B(1 + g)}[(1 + r)B + (1 - P)(1 + r + e)gB \\ \quad - (1 + u)(B + gB)], \quad \text{if } x \in (x^W, +\infty] \end{cases} \tag{22}$$

Specifically, when $x \in (-\infty, x^S]$, each borrower will choose a stated-doc loan at the amount B. The strong borrower will pay back $(1 + r + x)B$ for sure, while the weak borrower will pay back $(1 + r + x + e)B$ with a probability $(1 - P - k)$. Therefore the lender's expected rate of return is $E(R)_1 = 1/2B\,[(1 + r + x)B + (1 - P - k)(1 + r + x + e)B - 2B]$. When $x \in (x^S, x^W]$, the strong borrower chooses a full-doc loan at the amount B (incurring a documentation cost uB to the lender) and will pay back $(1 + r)B$ for sure, while the weak borrower chooses a stated-doc loan at amount B and will pay back $(1 + r + x + e)B$ with a probability $(1 - P - k)$. Therefore the lender's expected rate of return is $E(R)_2 = 1/2B\,[(1 + r)B + (1 - P - k)(1 + r + x + e)B - B - (1 + u)B]$. When $x \in (x^W, +\infty]$, the strong borrower chooses a full-doc loan at the amount B (incurring a documentation cost uB to the lender) and will pay back $(1 + r)B$ for sure, while the weak borrower chooses a full-doc loan at the amount gB (incurring a documentation cost ugB to the lender), and will pay back $(1 + r + e)gB$ with a probability $(1 - P)$. Therefore the lender's expected rate of return is $E(R)_3 = 1/B(1 + g)\,[(1 + r)B + (1 - P)(1 + r + e)gB - (1 + u)(B + gB)]$.

From Equation (22), we derive the local optimum under each condition as:

$$\max_{\{x \in (-\infty, x^S]\}} E(R)_1 \rightarrow \textit{corner solution } x^*$$
$$= x^S \left(as \; \frac{dE(R)_1}{dx} = \frac{B^2}{2}(2 - k - P) > 0 \right); \quad (23)$$
$$\max_{\{x \in (x^S, x^W]\}} E(R)_2 \rightarrow \textit{corner solution } x^*$$
$$= x^W \left(as \; \frac{dE(R)_2}{dx} = \frac{B^2}{2}(1 - k - P) > 0 \right); \quad (24)$$
$$\max_{\{x \in (x^W, +\infty]\}} E(R)_3 \rightarrow x^* = \textit{any value that is}$$
$$> x^W \left(as \; \frac{dE(R)_3}{dx} = 0 \right). \quad (25)$$

This proves the fourth column of the table in Proposition 1.

Comparing these three situations,

$$E(R)_2|_{x^*=x^W} - E(R)_1|_{x^*=x^S}$$
$$= \frac{B}{2}\{B[(1 + r + e)(k - (1 - g)(1 - P)) - u]$$
$$- kS - d(1 - k - P)\}$$
$$> 0 \text{ (given } B \text{ much larger than other parameters, and}$$
$$k > (1 - g)(1 - P) \text{ suggesting that in general}$$
$$(1 + r + e)(k - (1 - g)(1 - P)) > u); \quad (26)$$
$$E(R)_3|_{x^*>x^W} - E(R)_1|_{x^*=x^S}$$
$$= \frac{B}{2}\{B[(1 + r + e)(k + P) - e - 2r] - d(2 - k - P)\}$$
$$+ \frac{g(r + e)(1 - P) + r - u - g(P + u)}{1 + g}$$
$$> 0 \text{ (given } B \text{ much larger than other parameters, and}$$
$$\text{in general: } (k + P)(1 + r + e) > e + 2r); \quad (27)$$

$$E(R)_2|_{x^*=x^W} - E(R)_3|_{x^*>x^W}$$
$$= \frac{B}{2}\{B[r + g(1 - P)(1 + r + e) - (1 + u)] + d - kS\}$$
$$+ \frac{g[P - (r + e)(1 - P)] - r}{1 + g} + u$$
$$\geq 0 \text{ if } r + g(1 - P)(1 + r + e) \geq 1 + u, \text{ while } < 0$$
$$\text{if otherwise (given } B \text{ is much larger than other parameters).} \quad (28)$$

Since $E(R)_1|_{x^*=x^S}$ is lower than both $E(R)_2|_{x^*=x^W}$ and $E(R)_3|_{x^*>x^W}$, x^S can not be the optimal doc risk premium. The lender will then choose x^W if $r + g(1 - P)(1 + r + e) \geq 1 + u$, while any doc risk premium that is greater than x^W if otherwise. We hence prove the last column of the table in Proposition 1. ■

Endnotes

[1] As of December 2009, a record 10.04% of home mortgages was in either default or the foreclosure process (Inside Mortgage Finance, 2010). Rogers and Winter (2009), Ding, Quercia, and Ratcliffe (2010), and Daneshvary, Clauretie, and Kader (2011) discuss the negative spillover effects of foreclosures.

[2] Bear Stearns collapsed and was sold to JPMorgan Chase in March 2008. An obvious question is whether data from Bear Stearns mortgage-backed securities is representative of the broader mortgage market, or at least of that segment of the market on which we focus. To address this issue, we compare our data to aggregate measures reported from the 2011 Mortgage Market Statistical Annual. Bear Stearns was one of the top three originators of residential mortgage-backed securities during each year of our study period. Moreover, its market share tended to be higher in the Alt-A segment, as opposed to the subprime segment, where specialty firms such as New Century tended to dominate (at least prior to the bankruptcies that began in 2007). The phenomenon we focus on, stated documentation loans, may be characterized as an underwriting method that originated in the prime segment, but migrated down the credit spectrum over time. Indeed, the prevalence of reduced documentation in the 2006 subprime cohort is probably one of the reasons that vintage has performed so poorly. In any event, we are relatively more comfortable with our inferences as they apply to Alt-A, as compared to subprime, where Bear Stearns did not hold a commanding market share. We note, too, that Bear Stearns volume was more highly geographically concentrated in risky areas such as California and Florida, which undoubtedly further contributed to the high default rates we observe in our data. As to whether Bear Stearns was simply a poorly-managed firm or not, we would argue that it was emblematic of an over-leveraged investment bank that was over-invested in the mortgage sector at the time of a market panic. See Cohen (2009) for a more comprehensive history of the rise and fall of the firm.

[3] See Inside Mortgage Finance (2007).

[4] Besanko and Thakor (1992) suggest that increasing competition in the banking industry due to deregulation will induce a bank to adopt or change policies to be more favorable to its borrowers and savers, albeit at its own cost. The reduced loan documentation requirement studied in our paper is one of the outcomes. Another analogous example is the increase in service costs associated with the pressure to attract customers (Maudos and Guevara, 2007).

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

[6] Uhde and Michalak (2010) provide empirical evidence for a positive relation between the credit risk securitization practice and a bank’s risk exposure. Peni, Smith, and Vähämaa (2013) explore the relations between banks’ corporate governance strengths and their risk-taking in real estate loan lending before and during the recent financial crisis.

[7] In our model, the strong borrower does not need to speculate as she can get the same size loan regardless of documentation type. In the real world, some good borrowers may also chase bigger houses upon reduced documentation in a rising market which is not the focus of this study.

[8] Assuming that documentation cost is proportional to loan size can greatly simplify the math. If alternatively we assume that the documentation cost is constant, the essence of our findings discussed later will not change.

[9] Since this study focuses on the impact of documentation type on default risk, we ignore other issues including the possibility that credit type will affect the loan size, as well as the impact of loan pricing.

[10] Due to missing data and apparent errors in records for the variable ‘‘original combined LTV’’, we exclude loans for properties with multiple loans (that is, loans with ‘‘original combined LTV’’ exceeding ‘‘original LTV’’). In other words, we exclude all loans with piggyback seconds from our data to focus only on the documentation issue. For research specifically focused on that segment, see LaCour-Little, Calhoun, and Yu (2011).

[11] For example, if PITI is $2,500 per month and front-end ratio is 0.25, then inferred monthly income is $10,000.

[12] Intuitively, given a national average homeownership rate at around 70%, an ordinary home purchaser’s income should be within a reasonable variation range of the local average income. Of course, some real estate investors who take low documentation loans may also have higher than average income. But loans identified as for investment property purchase are excluded from our data. In this process, we effectively ignore occupancy fraud issues.

[13] Note that the ‘‘no-doc’’ category here covers not only no documentation loans, but also no-ratio loans, an intermediate category between stated-income and no-doc. We merge these two types to simplify our categories and focus on the stated-doc loans, the only category where borrowers (or brokers) can actively misrepresent borrower income. Due to data limitations, we also simply exclude another category ‘‘lite doc.’’

[14] Inclusion of the yield curve and the level mortgage rate helps control for the influence of changes in credit supply that have been found significant during our sample period (Mian and Sufi, 2009).

[15] 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 the 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 the auction date. Capturing those outcomes in detail is important for measurement of loss severity, as opposed to default rate, which is our focus here. In practical applications, lenders need to develop both default probability and loss severity models. Medema, Koning, and Lensink (2009) provide a useful analysis of methods for validating the former in accordance with regulatory capital rules.

[16] Without sufficient indication of a clear direction of causality between loan documentation type and other variables, we view this more as a multicollinearity problem than an endogeneity problem, and hence do not employ the popular method to address the latter, namely, Heckman's two-stage model.

[17] Our data does show that our proxy for income exaggeration, the ratio between the inferred borrower income and the MSA-median household income, exhibits a positive correlation with credit score, LTV, local housing price appreciation rate, and the use of low documentation, which turns out to also affect default risk.

[18] Low-doc loans typically did not require disclosure of downpayment source, potentially allowing borrowers to use ''silent seconds'' to fund apparent equity contributions, effectively understating true LTV, another example of risk-layering. We cannot, of course, identify such cases.

[19] The difference in results between fast and slow appreciation cities may be associated with the problem of qualifying in markets with higher prices. Households in such markets already have incentives to increase loan size (to capture capital gains if the fast appreciation trend continues), and now have greater opportunity to do so due to the availability of the low documentation option.

[20] A recent study An and Qi (2012) discusses the problems with the hazard method for mortgage duration data.

[21] Our result that the prepayment option value (which reflects the prepayment risk) increases default risk, is consistent with the literature (e.g., Pennington and Ho, 2010). One rationale is that if a borrower could benefit from refinancing but does not, that suggests that he or she may have some credit or financial constraints the presence of which are, in turn, related to default risk.

[22] Note that the signs of the effects of these risk factors in the default regressions with the full sample and the restricted sample are in general consistent with what found in literature, such as Pennington-Cross and Ho (2010).

[23] Note that the signs of the effects of other pricing factors in the loan pricing regressions with the full sample and the restricted sample are in general consistent with what found in literature, such as Courchane (2007) and Pennington and Ho (2008). As a robustness test, we also try one-stage pricing regressions, where loan documentation dummies or income-overstatement proxies are included in the first-pass pricing regressions, together with other determinants of loan pricing. We find that the low-documentation risk premium and the income-overstatement risk premium are with similar magnitudes as in the multi-stage regressions.

References

An, M. and Z. Qi. Competing Risks Models Using Mortgage Duration Data under the Proportional Hazards Assumption. *Journal of Real Estate Research*, 2012, 34:1, 1–26.

Berkovec, J., Y. Chang, and D. McManus. Alternative Lending Channels and the Crisis in U.S. Housing Markets. Freddie Mac Working Paper, 2011.

Besanko, D. and A.V. Thakor. Banking Deregulation: Allocational Consequences of Relaxing Entry Barriers. *Journal of Banking and Finance*, 1992, 16, 909–32.

Brueckner, J.K. Mortgage Default with Asymmetric Information. *Journal of Real Estate Finance and Economics*, 2000, 20:3, 251–74.

Campbell, J.Y. and J. Cocco. Household Risk Management and Optimal Mortgage Choice. *Quarterly Journal of Economics*, 2003, 118:4, 1449–94.

Cohen, N. Banks Still Lending at Many Times Earnings. T*he Financial Times*, January 31, 2009.

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

Daneshvary, N., T. Clauretie, and A. Kader. Short-Term Own-Price and Spillover Effects of Distressed Residential Properties: The Case of a Housing Crash. *Journal of Real Estate Research*, 2011, 33:2, 179–207.

Deng, Y., J. Quigley, and R. Van Order. Mortgage Terminations, Heterogeneity and the Exercise of Mortgage Options. *Econometrica*, 2000, 68:2, 275–308.

Ding, L., R. Quercia, and J. Ratcliffe. Neighborhood Subprime Lending and the Performance of Community Reinvestment Mortgages. *Journal of Real Estate Research*, 2010, 32:3, 341–76.

Ding, L., R. Quercia, W. Li, and J. Ratcliffe. Risky Borrowers or Risky Mortgages Disaggregating Effects Using Propensity Score Models. *Journal of Real Estate Research*, 2011, 33:2, 245–78.

Firestone, S., R. Van Order, and P. Zorn. The Performance of Low-Income and Minority Mortgages. *Real Estate Economics*, 2007, 35:4, 479–504.

Gerardi, K., A. Lehnert, S. Sherlund, and P. Willen. Making Sense of the Subprime Crisis. Federal Reserve Bank of Atlanta Working paper, 2009-2.

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

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

——. Mortgage Delinquency Rates Continue to Worsen in Late 2009 as Virus Spreads to Prime Loans. *Inside Mortgage Finance*, February 12, 2010.

Jiang, W., A. Nelson, and E. Vytlacil. Liar's Loans? Effects of Loan Origination Channel and Loan Sale on Delinquency. Paper presented at FHFA Symposium on Mortgage Default. Washington, D.C., September 16, 2009.

LaCour-Little, M. The Home Purchase Mortgage Preferences of Low-and-Moderate Income Households. *Real Estate Economics*, 2007, 35:4, 265–90.

LaCour-Little, M. and J. Yang. Pay Me Now or Pay Me Later: Alternative Mortgage Products and the Mortgage Crisis. *Real Estate Economics*, 2010, 38:4, 687–732.

LaCour-Little, M., C. Calhoun, and W. Yu. What Role Did Piggyback Lending Play in the Housing Bubble and Mortgage Collapse? *Journal of Housing Economics*, 2011, 20, 81–100.

MacDonald, D. and K. Winson-Geideman. Residential Mortgage Selection, Inflation Uncertainty, and Real Payment Tilt. *Journal of Real Estate Research*, 2012, 34:1, 51–71.

Maudos, J. and J.F. Guevara. The Cost of Market Power in Banking: Social Welfare Loss vs. Cost Inefficiency. *Journal of Banking and Finance*, 2007, 31, 2103–25.

Medema, L., R.H. Koning, and R. Lensink. A Practical Approach to Validating a PD Model. *Journal of Banking and Finance*, 2009, 33, 701–8.

Mian, A. and A. Sufi. The Consequences of Mortgage Credit Expansion: Evidence from the U.S. Mortgage Default Crisis. *Quarterly Journal of Economics*, 2009, 124:4, 1449–96.

Nichols, J., A. Pennington-Cross, and A. Yezer. Borrower Self-Selection, Underwriting Costs, and Subprime Mortgage Credit Supply. *Journal of Real Estate Finance and Economics*, 2005, 30:2, 197–219.

Pavlov, A. and S. Wachter. Subprime Lending and Real Estate Prices. *Real Estate Economics*, 2011, 39:1, 1–17.

Paley, I. and K. Tzioumis. Stated-Income Loans: Separating the Wheat from the Chaff. Office of the Comptroller of the Currency Working Paper, 2011.

Peni, E., S. Smith, and S. Vähämaa. Bank Corporate Governance and Real Estate Lending During the Financial Crisis. *Journal of Real Estate Research*, 2013, 35:3, 313–43.

Pennington-Cross, A. and G. Ho. Predatory Lending Laws and the Cost of Credit. *Real Estate Economics*, 2008, 36:2, 175–211.

——. The Termination of Subprime Hybrid and Fixed Rate Mortgages. *Real Estate Economics*, 2010, 38:3, 399–426.

Rajan, U., A. Seru, and V. Vig. The Failure of Models That Predict Failure: Distance, Incentives and Defaults. University of Chicago Working Paper, 2008.

Rogers, W. and W. Winter. The Impact of Foreclosures on Neighboring Housing Sales. *Journal of Real Estate Research*, 2009, 31:4, 455–79.

Rothschild, M. and J. Stiglitz. Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information. *Quarterly Journal of Economics*, 1976, 90:4, 629–49.

Sa-Aadu, J. and C.F. Sirmans. The Pricing of Adjustable Rate Mortgage Contracts. *Journal of Real Estate Finance and Economics*, 1989, 2, 253–66.

Uhde, A. and T.C. Michalak. Securitization and Systematic Risk in European Banking: Empirical Evidence. *Journal of Banking and Finance*, 2010, 3061–77.

United States Government Accountability Office. 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, 2006.

Wheaton, W. and G. Nechayev. The 1998–2005 Housing ‘‘Bubble’’ and the Current ‘‘Correction’’: What’s Different This Time? *Journal of Real Estate Research*, 2008, 30:1, 1–26.

We acknowledge helpful comments from the editor, four anonymous referees, Xudong An, Jan Brueckner, Jane Dokko, Stuart Gabriel, Richard Green, John Karkari, Andreas

Lehnert, Robert Van Order, Walter Torous, Kerry Vandell, Paul Willen, Ko Wang, Zhonghua Wu, Anthony Yezer, Peter Zorn, as well as other participants at the 2009 FDIC-FHFA Symposium on "Improving Assessment of the Default Risk or Single-Family Mortgages", the 2010 AREUEA Mid-Year Conference, the 2010 American Real Estate Society Annual Meeting, the 2010 Global Chinese Real Estate Congress Meeting, the 2011 AREUEA Annual Conference, and the 2011 Annual UCI-UCLA-USC Urban Research Day.

Michael LaCour-Little, California State University, Fullerton, CA 92834-6848 or mlacour-little@fullerton.edu.

Jing Yang, California State University, Fullerton, CA 92834-6848 or jyang@ fullerton.edu.