The Journal of FINANCE
The Journal of THE AMERICAN FINANCE ASSOCIATION

THE JOURNAL OF FINANCE • VOL. LXXI, NO. 6 • DECEMBER 2016

Credit Rationing, Income Exaggeration, and Adverse Selection in the Mortgage Market

BRENT W. AMBROSE, JAMES CONKLIN, and JIRO YOSHIDA*

ABSTRACT

We examine the role of borrower concerns about future credit availability in mitigating the effects of adverse selection and income misrepresentation in the mortgage market. We show that the majority of additional risk associated with "low-doc" mortgages originated prior to the Great Recession was due to adverse selection on the part of borrowers who could verify income but chose not to. We provide novel evidence that these borrowers were more likely to inflate or exaggerate their income. Our analysis suggests that recent regulatory changes that have essentially eliminated the low-doc loan product would result in credit rationing against self-employed borrowers.

DURING THE GREAT RECESSION OF 2007 to 2008, the United States experienced a massive increase in residential mortgage defaults and foreclosures not seen since the Great Depression. For example, the Financial Crisis Inquiry Commission reports that 9.7% of all mortgages were in default by the end of 2009, compared to approximately 1% at the start of the decade.[1] While the decline in house prices between 2007 and 2009 is obviously one of the primary causes for the significant number of mortgage defaults registered during the crisis, financial economists have only recently begun to examine the role of mortgage fraud and adverse selection in exacerbating the consequences of the 2007 to 2009 housing bust. Evidence is mounting that the great mortgage expansion that accompanied the rise in home prices coincided with increases in mortgage fraud related to misrepresentations of borrower income (Jiang, Nelson, and

*Brent W. Ambrose and Jiro Yoshida are with the Pennsylvania State University Smeal College of Business. James Conklin is with the University of Georgia Terry College of Business. We thank Itzhak Ben-David, Sumit Agarwal, Manuel Adelino, Christopher Palmer, Sam Kruger, Ken Singleton (Editor), the anonymous Associate Editor and referees, and seminar participants at the National University of Singapore, University of California–Berkeley, Research Institute of Capital Formation, Housing Research and Advancement Foundation of Japan, Board of Governors of the Federal Reserve System, University of Cambridge, Federal Reserve Bank of Atlanta Real Estate Finance Conference, and the 2015 ASSA meeting in Boston for their helpful comments and suggestions. We also thank the Penn State Institute for Real Estate Studies for providing access to the New Century Mortgage database. We gratefully acknowledge research assistance from Sergio Garate. The authors have no material financial interests related to this research, as identified in the journal's Disclosure Policy; furthermore, no party had the right to review this paper prior to circulation.

[1] See U.S. Financial Crisis Inquiry Commission (2011, p. 215). Default is defined as "90-days or more past due or in foreclosure."

DOI: 10.1111/jofi.12426

Vytlacil (2014), Mian and Sufi (2015)), borrower assets (Garmaise (2015)), inflated appraisals (Ben-David (2011), Agarwal, Ambrose, and Yao (2014), Agarwal, Ben-David, and Yao (2015), Griffin and Maturana (2015)), and second liens and owner-occupancy status (Piskorski, Seru, and Witkin (2015)).[2] As a result, regulators and policy makers have implemented new rules to combat perceived abuses in mortgage lending.[3] The purpose of this paper is to shed light on how borrower heterogeneity with respect to employment status contributed to income misrepresentation and adverse selection, and how lender actions and borrower concerns about preserving future access to credit mitigated these risks. From a policy perspective, our results echo the concerns raised by Keys et al. (2009), Rajan, Seru, and Vig (2010), and Piskorski, Seru, and Vig (2010), among others, concerning the need to carefully weigh the costs and benefits of new financial regulations.

With respect to income misrepresentation, we present several novel insights. First, by comparing individual incomes within job titles, we provide new evidence that strongly suggests that income misrepresentation was concentrated primarily among borrowers who originated low-documentation loans but could have easily originated full-documentation mortgages instead. Second, unlike previous studies that show that misrepresentation in mortgage originations resulted from lender actions at origination,[4] we find that income falsification was essentially a borrower-level phenomenon.[5] Our results therefore suggest that excesses in the mortgage market in the last decade resulted from both borrower and lender actions. Third, we provide new evidence on lender actions in response to potential borrower income falsification. Finally, in additional analysis examining the role of borrower income falsification in facilitating the expansion in mortgage credit, we provide new insights into one of the possible causes of the Great Recession.

The role of borrower income misrepresentation leading up to the financial crisis is a source of considerable debate. For example, Mian and Sufi (2009) and Mian and Sufi (2015) argue that borrower income falsification was a leading culprit in facilitating the expansion of mortgage credit during the 2002 to 2006 housing boom. Supporting this argument, Jiang, Nelson, and Vytlacil (2014) show that income falsification occurred on low-documentation loans,

[2] In addition to mispresentation at the loan origination level, Piskorski, Seru, and Witkin (2015) find that misrepresentation was endemic in the secondary market (between originators and investors) as well. Furthermore, Agarwal and Evanoff (2013) and Agarwal et al. (2014) provide evidence of systematic predatory lending practices by loan originators. These practices may have exacerbated the consequences of mortgage fraud.

[3] For example, the Consumer Financial Protections Bureau adopted the "Ability to Repay Rule" that requires lenders to provide greater documentation of borrower income, and the Federal Housing Finance Agency in conjunction with the New York Attorney General's office issued the Home Valuation Code of Conduct (HVCC) that was designed to reduce the incidences of inflated appraisals.

[4] Piskorski, Seru, and Witkin (2015) is a notable exception. The authors provide evidence that borrowers misrepresented occupancy status on mortgage applications.

[5] Note that we are not arguing that the originator was unaware of the misrepresentation, but rather that borrowers are complicit in falsifying income.

resulting in elevated defaults, particularly for loans originated through the wholesale channel. By focusing on differences in employment status, we show that the majority of adverse selection and income falsification is confined to a specific borrower group that was never intended to use the low-documentation product. Our results therefore suggest that broad policies designed to eliminate activities associated with excesses in mortgage originations during the housing boom may have unintended consequences.

Since the potential for mortgage fraud and adverse selection has always been present, lenders have long relied on underwriting guidelines to limit this risk. However, Burke, Taylor, and Wagman (2012) illustrate how lender screening to reject higher risk applicants results in greater adverse selection.[6] One such underwriting metric is the debt-to-income (DTI) ratio, which limits the loan amount based on the borrower's income. This metric, in combination with the loan-to-value (LTV) ratio, serves to limit the borrower's housing consumption. As a result, borrowers seeking to maximize their housing consumption or investment have an incentive to exaggerate their income in order to reduce their DTI ratio and thereby qualify for a higher loan amount.

Recognizing the borrower's incentive to circumvent these metrics, mortgage lenders require proof of reported assets and incomes to verify that the borrower is capable of repaying the debt. Of course, verification of borrower income and assets comes at a cost. Not only do lenders bear the costs associated with verification activities, but borrowers also bear costs of collecting and reporting income and assets to the lender. For some borrowers, these costs are relatively minor and involve simply submitting the prior two years' W-2 tax documents from their employer along with their past two months' pay stubs. However, for many other potential borrowers, the costs of verifying income and assets are not so trivial. For example, self-employed individuals would need to provide full tax returns for the previous two years. Yet, self-employed individuals often file for tax return extensions due to the complexity of their tax situation, and as a result, the returns are not available to the lender. Self-employed individuals are also required to provide current profit and loss statements along with bank statements for several months to prove sufficient cash flow to service the debt. Furthermore, to comply with underwriting DTI guidelines, lenders may require additional documentation from self-employed borrowers to determine the nature of deposits and withdrawals and ascertain those expenses that are personal versus those associated with their business.[7]

[6] However, the presence of adverse selection at mortgage origination is not universally accepted. For example, Agarwal, Chang, and Yavas (2012) rely on differences in loan performance between prime and subprime markets to claim that adverse selection was less severe in the subprime market.

[7] Anecdotal discussions with mortgage brokers and other industry participants provide examples of the verification costs self-employed borrowers face. For example, lenders may require that self-employed borrowers provide written explanations for every deposit over the previous year. For a business with just two transactions per week, that would necessitate over 100 separate written explanations. Furthermore, many self-employed borrowers face serious confidentiality issues in revealing client names as the source of deposits.

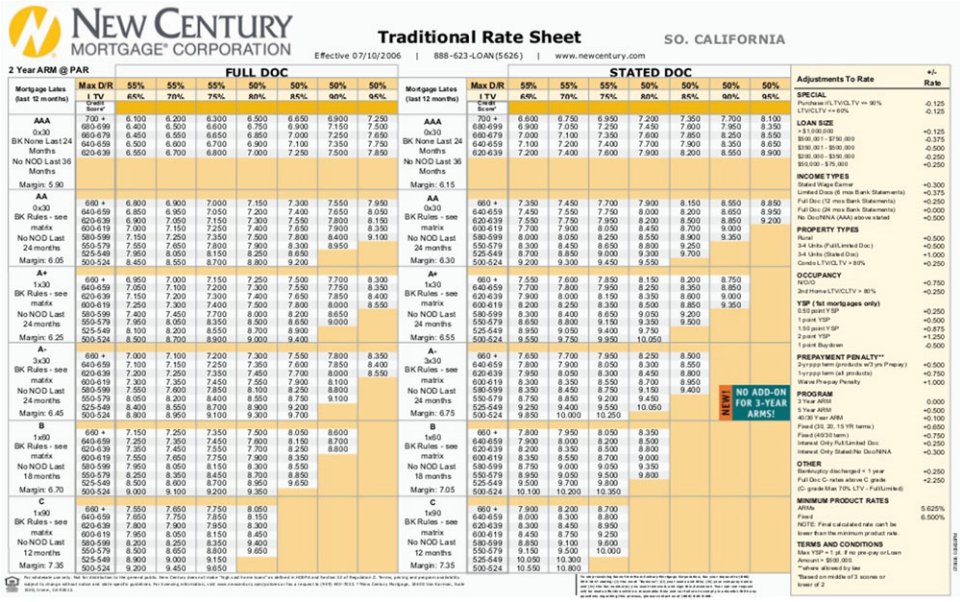
Over time the mortgage industry has developed different products designed to cater to borrowers with varying degrees of information verification costs. The traditional mortgage referred to as a full documentation (or full-doc) loan is designed for borrowers who can easily and with low cost document their financial situation. Recognizing that many self-employed borrowers would be effectively credit-rationed in the traditional loan market due to the costs associated with documenting income and assets, the mortgage industry developed an alternative low-documentation (low-doc), or stated-income stated-asset loan.[8] However, the low-doc product also provides an avenue for some borrowers to inflate or exaggerate their income in order to qualify for larger mortgages. While borrowers are still subject to civil or criminal legal actions for providing inaccurate information, the costs associated with pursuing borrowers who fraudulently overstate income or assets often exceed the possible claims, particularly if the loan is still performing. Herein lies the tension in the low-doc product: as long as the borrower is making payments, the lender does not have an incentive to take actions against the borrower for falsely representing her income or assets.

To clarify the constraints facing borrowers, we present the mortgage rate sheet for New Century Financial Corporation (Figure 1). The rate sheet lists the interest rates charged on mortgages (as of July 10, 2006) originated by New Century based on whether the borrower was willing to verify income and assets ("Full Doc") or did not provide tax returns and bank accounts to verify income and assets ("Stated Doc"). Each block in the rate sheet represents a borrower risk class ("AAA through C") that is based on the number of late payments, prior default records, or bankruptcy filings. Shaded areas without interest rates indicate that the loan product is not offered to borrowers who have credit scores in those risk categories.

To illustrate how borrower information verification costs and loan performance could interact to result in credit rationing, consider a high information-cost borrower rated "A+" with a credit score of 660 who seeks an 85% LTV ratio mortgage.[9] Since this borrower finds it costly to verify income, he applies under the "Stated Doc" product type and is quoted a contract interest rate of 8.200%. The impact of potential credit rationing becomes apparent if the borrower is downgraded to the "B" category (e.g., by a 60-day-late experience) before seeking to refinance into a new mortgage. Under the "B" category, New Century does not offer a stated-doc loan at an 85% LTV, that is, the borrower is effectively credit-constrained unless he is willing to move to a lower LTV mortgage at a higher contract rate. In contrast, a comparable low-information-cost borrower who experienced a similar downgrade could easily switch to a full-doc product with the same LTV. Since both borrowers are aware of this

[8] See Paley and Tzioumis (2011) and LaCour-Little and Yang (2013). We use the terms low-doc, no-doc, and stated-income interchangeably.

[9] The "A+" category indicates that a borrower was 30 days late on a previous mortgage only once in the last 12 months.



NEW CENTURY MORTGAGE® CORPORATION

Traditional Rate Sheet SO. CALIFORNIA

Effective 07/10/2006 | 888-623-LOAN(5626) | www.newcentury.com

2 Year ARM @ PAR

FULL DOC

Mortgage Lates (last 12 months)	Max D/R / TV / Credit Score	55% 65%	55% 70%	55% 75%	50% 80%	50% 85%	50% 90%	50% 95%
AAA 0x30 BK None Last 24 Months No NOD Last 36 Months Margin: 5.90	700 +	6.100	6.200	6.300	6.500	6.650	6.900	7.250
	680-699	6.400	6.500	6.600	6.750	6.900	7.150	7.500
	660-679	6.450	6.550	6.650	6.850	7.000	7.250	7.650
	640-659	6.500	6.600	6.700	6.900	7.100	7.350	7.750
	620-639	6.550	6.700	6.800	7.000	7.250	7.500	7.850
AA 0x30 BK Rules - see matrix No NOD Last 24 months Margin: 6.05	660 +	6.800	6.900	7.000	7.150	7.300	7.550	7.950
	640-659	6.850	6.950	7.050	7.200	7.400	7.650	8.050
	620-639	6.900	7.050	7.150	7.300	7.550	7.800	8.150
	600-619	7.000	7.150	7.250	7.400	7.650	7.900	8.350
	580-599	7.150	7.250	7.350	7.500	7.800	8.400	9.100
	550-579	7.550	7.650	7.800	7.900	8.300	8.950	
	525-549	7.950	8.050	8.150	8.250	8.650		
	500-524	8.450	8.550	8.700	8.800	9.200		
A+ 1x30 BK Rules - see matrix No NOD Last 24 months Margin: 6.25	660 +	6.950	7.000	7.150	7.250	7.500	7.700	8.300
	640-659	7.050	7.100	7.200	7.300	7.550	7.750	8.350
	620-639	7.150	7.200	7.300	7.400	7.650	7.850	8.400
	600-619	7.250	7.300	7.400	7.500	7.800	8.000	8.550
	580-599	7.400	7.450	7.700	8.000	8.200	8.650	
	550-579	7.950	8.050	8.350	8.500	8.650	9.000	
	525-549	8.100	8.200	8.550	8.700	8.900		
	500-524	8.500	8.700	8.900	9.000	9.400		
A- 3x30 BK Rules - see matrix No NOD Last 24 months Margin: 6.45	660 +	7.000	7.100	7.200	7.300	7.550	7.800	8.350
	640-659	7.100	7.150	7.250	7.350	7.600	7.850	8.400
	620-639	7.200	7.250	7.350	7.450	7.700	8.000	8.550
	600-619	7.300	7.350	7.450	7.550	7.900	8.100	
	580-599	7.550	7.600	7.850	8.100	8.250	8.800	
	550-579	8.050	8.200	8.400	8.550	8.750	9.100	
	525-549	8.400	8.550	8.700	8.900	9.200		
	500-524	8.800	8.950	9.100	9.300	9.700		
B 1x60 BK Rules - see matrix No NOD Last 18 months Margin: 6.70	660 +	7.150	7.250	7.350	7.500	8.050	8.600	
	640-659	7.250	7.350	7.450	7.600	8.150	8.700	
	620-639	7.350	7.450	7.550	7.700	8.250	8.800	
	600-619	7.550	7.650	7.750	7.900	8.350		
	580-599	7.950	8.050	8.150	8.300	8.550		
	550-579	8.250	8.350	8.450	8.700	8.850		
	525-549	8.500	8.600	8.700	8.950	9.650		
	500-524	9.000	9.100	9.200	9.350			
C 1x90 BK Rules - see matrix No NOD Last 12 months Margin: 7.35	660 +	7.550	7.650	7.750	8.050			
	640-659	7.650	7.750	7.850	8.150			
	620-639	7.800	7.900	7.950	8.300			
	600-619	7.950	8.050	8.150	8.450			
	580-599	8.200	8.250	8.350	9.400			
	550-579	8.500	8.650	8.800	9.650			
	525-549	8.900	9.000	9.150				
	500-524	9.200	9.450	9.650				

STATED DOC

Mortgage Lates (last 12 months)	Max D/R / TV / Credit Score	55% 65%	55% 70%	55% 75%	50% 80%	50% 85%	50% 90%	50% 95%
AAA 0x30 BK None Last 24 Months No NOD Last 36 Months Margin: 6.15	700 +	6.600	6.750	6.950	7.200	7.350	7.700	8.100
	680-699	6.900	7.050	7.250	7.450	7.600	7.950	8.350
	660-679	7.000	7.100	7.350	7.600	7.850	8.250	8.550
	640-659	7.100	7.200	7.400	7.700	7.900	8.350	8.650
	620-639	7.200	7.400	7.600	7.900	8.200	8.550	8.900
AA 0x30 BK Rules - see matrix No NOD Last 24 months Margin: 6.30	660 +	7.350	7.450	7.700	7.900	8.150	8.550	8.850
	640-659	7.450	7.550	7.750	8.000	8.200	8.650	8.950
	620-639	7.550	7.750	7.950	8.200	8.500	8.850	9.200
	600-619	7.700	7.900	8.050	8.450	8.700	9.000	
	580-599	8.000	8.050	8.250	8.550	8.900	9.350	
	550-579	8.300	8.450	8.650	8.800	9.250		
	525-549	8.700	8.850	9.000	9.300	9.700		
	500-524	9.200	9.300	9.450	9.550			
A+ 1x30 BK Rules - see matrix No NOD Last 24 months Margin: 6.55	660 +	7.550	7.600	7.850	8.150	8.200	8.750	
	640-659	7.700	7.800	7.950	8.250	8.350	8.850	
	620-639	7.900	8.000	8.150	8.350	8.600	9.000	
	600-619	8.200	8.250	8.350	8.500	8.850	9.350	
	580-599	8.300	8.400	8.650	9.050	9.200		
	550-579	8.650	8.800	9.150	9.350	9.500		
	525-549	8.950	9.050	9.400	9.750			
	500-524	9.550	9.750	9.950	10.050			
A- 3x30 BK Rules - see matrix No NOD Last 24 months Margin: 6.75	660 +	7.650	7.700	7.950	8.250	8.500		
	640-659	7.800	7.900	8.050	8.300	8.550		
	620-639	7.950	8.050	8.300	8.450	8.800		
	600-619	8.300	8.350	8.550	8.700	8.950		
	580-599	8.350	8.450	8.750	9.150	9.400		
	550-579	8.750	8.850	9.200	9.450			
	525-549	9.250	9.400	9.550	10.050			
	500-524	9.850	10.000	10.250				
B 1x60 BK Rules - see matrix No NOD Last 18 months Margin: 7.05	660 +	7.800	7.950	8.050	8.350			
	640-659	7.900	8.000	8.200	8.500			
	620-639	8.200	8.350	8.500	8.800			
	600-619	8.350	8.550	8.700	9.000			
	580-599	8.750	9.000	9.050	9.350			
	550-579	8.950	9.050	9.500	9.800			
	525-549	9.500	9.700	9.800				
	500-524	10.100	10.200	10.350				
C 1x90 BK Rules - see matrix No NOD Last 12 months Margin: 7.35	660 +	7.900	8.200	8.700				
	640-659	8.000	8.300	8.800				
	620-639	8.300	8.450	8.950				
	600-619	8.450	8.750	9.250				
	580-599	8.850	9.100	9.600				
	550-579	9.150	9.500	10.000				
	525-549	10.050	10.300					
	500-524	10.550	10.800					

NEW! NO ADD-ON FOR 3-YEAR ARMS!

Adjustments To Rate	+/- Rate
SPECIAL	
Purchase if LTV/CLTV <= 90% LTV/CLTV <= 80%	-0.125 -0.125
LOAN SIZE	
> $1,000,000	+0.125
$698,001 - $750,000	-0.375
$350,001 - $500,000	-0.500
$200,000 - $350,000	-0.250
$50,000 - $75,000	+0.250
INCOME TYPES	
Stated Wage Earner	+0.300
Limited Docs (6 mos Bank Statements)	+0.375
Full Doc (12 mos Bank Statements)	+0.250
Full Doc (24 mos Bank Statements)	+0.000
No Doc/NINA (AAA) above stated	+0.500
PROPERTY TYPES	
Rural	+0.500
3-4 Units (Full/Limited Doc)	+0.500
3-4 Units (Stated Doc)	+1.000
Condo LTV/CLTV > 80%	+0.250
OCCUPANCY	
NOO	+0.750
2nd Home LTV/CLTV > 80%	+0.250
YSP (flat mortgages only)	
0.50 point YSP	+0.250
1 point YSP	+0.500
1.50 point YSP	+0.875
2 point YSP	+1.250
1 point Buydown	-0.500
PREPAYMENT PENALTY**	
3yr ppp term (products w/3 yrs Prepay)	+0.500
1yr ppp term (all products)	+0.750
Waive Prepay Penalty	+1.000
PROGRAM	
3 Year ARM	0.000
5 Year ARM	+0.500
40/30 Year ARM	+0.100
Fixed (30, 20, 15 YR terms)	+0.650
Fixed (40/30 term)	+0.750
Interest Only Full/Limited Doc	+0.250
Interest Only Stated/No Doc/NINA	+0.300
OTHER	
Bankruptcy discharged < 1 year	+0.250
Full Doc C- rates above C grade (C- grade Max 70% LTV - Full/Limited)	+2.250
MINIMUM PRODUCT RATES	
ARMs	5.625%
Fixed	6.500%
NOTE: Final calculated rate can't be lower than the minimum product rate.	
TERMS AND CONDITIONS	
Max YSP = 1 pt. If no pre-pay or Loan Amount > $636,000	
**where allowed by law	
*Based on middle of 3 scores or lower of 2	

Figure 1. New Century rate sheet. This figure includes a pricing matrix from New Century Mortgage Corporation for the Southern California Region as of 07/10/2006. The original image is available at http://www.slideshare.net/devonauerswald/new-century-subprime-matrix-80-5.

difference in borrowing constraints, the high-verification-cost borrower would have greater concern about future credit rationing.[10]

To confirm that the insights obtained from the New Century rate sheet were common across the mortgage industry during the period prior to the Great Recession, we collected wholesale rate sheets for several other mortgage lenders originating loans during that period. Figures IA.1 to IA.7 in the Internet Appendix show the wholesale rate sheets for First Franklin on July 10, 2006, Countrywide on August 16, 2006, and Argent Mortgage on July 21, 2006.[11] Although it seems implausible in the context of current day mortgage underwriting practices, the similar pricing patterns across these three rate sheets and the New Century rate sheet reveal that full-doc loans were available to borrowers who had declared bankruptcy or had a mortgage default within two years of the origination date, while low-doc loans were not available to borrowers with these characteristics at any price. These rate sheets thus reinforce the

[10] Although our example assumes that borrowers accurately report their income, we recognize that the low-information-cost borrower may have falsified her income and thus not have sufficient "true" or verifiable income to qualify for a full-doc product with the same LTV after a downgrade. However, this does not alter our intuition above because the LTV available to this borrower is still higher than the LTV available on a low-doc mortgage to a comparable high-verification-cost borrower.

[11] The Internet Appendix is available in the online version of the article on the *Journal of Finance* website.

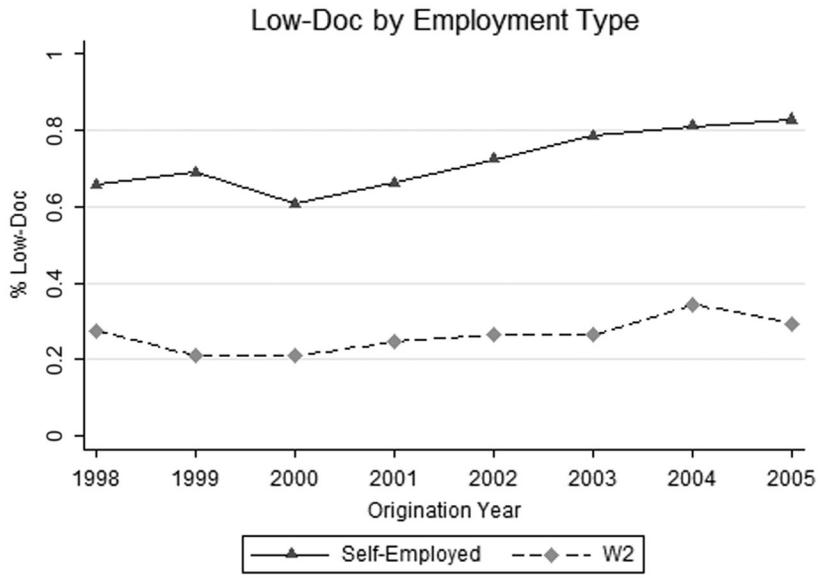


Figure 2. Share of originations that are low-doc by employment type. This figure shows the proportion of originated loans that are low-doc by employment type in each origination year. The sample includes funded loans from the New Century database as described in Section II.A.

expectation that borrowers who could not easily verify income via a low-cost W-2 would face credit rationing as a result of a prior bankruptcy or mortgage default.

Figure 2 demonstrates why understanding the role of future credit concerns in limited-information contracts is particularly important for self-employed borrowers. Using data from one of the largest subprime lenders in the run-up to the crisis, the figure shows the proportion of low-doc loans to self-employed and W-2 borrowers by origination year. Roughly 80% of self-employed borrowers obtain low-doc loans, compared to only 30% for W-2 borrowers. Clearly, low-doc loans are favored by the type of borrowers that they were originally intended for—the self-employed. Stated differently, limited-information debt contracts are an important source of credit for borrowers who are likely to be credit-rationed under full-information (full-doc) mortgage contracts.

To better understand the link between mortgage type and borrower employment status, we theoretically and empirically demonstrate that low-doc loans experience higher ex post default rates than full-doc loans, and that the relationship is strongest for low-doc W-2 loans—borrowers with the ability to access the full-documentation origination channel. In other words, we find that the majority of the additional risk associated with low-doc loans is due to adverse selection on the part of borrowers with verifiable income. We conjecture that these borrowers likely selected into low-doc loans in order to inflate income and, in turn, increase housing consumption. Thus, our analysis is connected

to the theoretical insights developed in Diamond (1989) and Diamond (1991) regarding the role of borrower reputation and concerns about future credit rationing in ameliorating adverse selection and income falsification.

Our results are related to an important recent attempt by Jiang, Nelson, and Vytlacil (2014) to quantify the amount of income inflation on low-doc loans to W-2 borrowers. Their results suggest that W-2 borrowers with low-doc loans exaggerated income by 20% to 25%. Using a similar methodology on loans originated by a different lender, we estimate that income inflation ranged between 7% and 13% on low-doc loans to W-2 borrowers. Our study thus provides an additional point estimate for the level of income overstatement on so-called "liars' loans." Additionally, to our knowledge, we are the first to provide evidence that, relative to W-2 employees, self-employed borrowers refrain from overstating income when applying for mortgage loans. In fact, our regression results provide no evidence that self-employed borrowers selecting low-doc loans reported incomes that were above predicted levels from an income estimation model. Furthermore, while we show that income inflation is directly related to ex post mortgage default for W-2 borrowers, the connection is less clear for the self-employed, which suggests that income falsification is most problematic on low-doc loans originated by W-2 borrowers.

One of the unique features of our data is that we have information on loan applications. We thus also investigate lender actions to mitigate borrower adverse selection by documenting that the probability of lender loan application rejection was positively associated with borrowers most likely to engage in income falsification. Additionally, we find that premiums were set at a level that allowed adverse selection and untruthful reporting to persist in equilibrium. We also show that the low-doc effect on mortgage performance is less pronounced for borrowers with high FICO scores or histories of mortgage repayment. Taken together, these results suggest that borrower concerns about possible credit rationing can mitigate adverse selection and private information in debt contracts. Finally, supporting the findings of Mian and Sufi (2015), we document that mortgages to borrowers who were most likely to overstate income were concentrated in lower income neighborhoods.

Our findings are particularly important in light of the Consumer Financial Protections Bureau's (CFPB) "Ability to Repay Rule," which went into effect in January of 2014. This rule implements sections 1411 and 1412 of the Dodd-Frank Wall Street Reform and Consumer Protection Act (Dodd-Frank Act), requiring that lenders verify and document a potential borrower's ability to repay the loan.[12] Loans that do not meet the rule leave the lender exposed to significant litigation risks, which effectively eliminates the low-doc loan market.

[12] The Dodd-Frank Act is available online at https://www.sec.gov/about/laws/wallstreetreformcpa.pdf. Information on the "Ability to Repay Rule" is available at http://www.consumerfinance.gov/regulations/ability-to-repay-and-qualified-mortgage-standards-under-the-truth-in-lending-act-regulation-z/#rule.

However, eliminating the low-doc market likely results in regulator-imposed credit rationing against self-employed borrowers. Consistent with this view, Green (2014, p. 19) provides a telling description of the current mortgage market: "[W]hile people who draw regular salaries and receive W-2 forms from the Internal Revenue Service at the end of each year have fairly ready access to mortgage credit, self-employed people find it very difficult to obtain a mortgage. This is even true for people who can document a long history of self-employment income." This credit rationing against self-employed borrowers can have significant negative consequences for the economy. For example, Adelino, Schoar, and Severino (2015b) provide direct evidence that employment in small businesses is related to home price appreciation. Their analysis suggests that rising home prices allowed mortgage credit to expand via the collateral channel, which, in turn, created equity that could be used as working capital in small businesses. As a result, eliminating the low-doc loan market may have adverse consequences on future employment growth. However, this credit rationing against self-employed borrowers is likely unnecessary. We argue that the low-doc loan channel provides access to credit for self-employed borrowers without a large increase in default risk, since self-employed borrowers' concerns for future credit significantly reduce the problems of adverse selection and income exaggeration endemic in low-doc loans originated by W-2 borrowers. Our analyses are thus consistent with the intuition embedded in models of reputation in financial contracting (e.g., Diamond (1989)).

Our paper proceeds as follows. In Section I, we discuss the interaction of borrower type based on income verification costs and mortgage product selection to develop a stylized model that motivates our empirical analysis. Section II discusses the data and summary statistics. Section III presents the empirical results linking mortgage performance to borrower concerns over future credit. Section IV provides evidence documenting the extent of borrower income misrepresentation and its impact on mortgage performance, and presents robustness checks to control for income differences across job types (Section IV.A) and within job types (Section IV.B). Next, Section V presents an analysis of lender responses to potential borrower income falsification. Specifically, we focus on lender screening at the time of application (Section V.A), links between observable credit quality measures to mortgage performance (Section V.B), and loan pricing (Section V.C). In Section VI, we highlight several important policy implications by examining the role of borrower income misrepresentation in facilitating the expansion of mortgage credit. Finally, Section VII concludes.

I. A Simple Model

To formulate testable hypotheses concerning the presence of adverse selection and a borrower's future access to credit, we first categorize mortgage contracts into high and low information loans based on the amount and extent of borrower information collected by the lender during the underwriting process. High information contracts represent full-doc mortgages where the loan originator collects and verifies the borrower's financial information (income

and assets) as reported on the loan application. In contrast, low information contracts represent low-doc mortgages where the originator does not independently verify the borrower's claims concerning assets or income.

Next, we categorize borrowers with respect to information verification costs. For example, borrowers who are self-employed often face high information verification costs since they are unable to provide lenders with a W-2 tax document from an employer. In contrast, borrowers who are employed by a third party have low information verification costs since they can easily produce an employer-generated W-2 statement that documents their income.

The lender understands that low information contracts are ex ante riskier and prices them accordingly. Furthermore, since the level of borrower income is often a critical component in determining the maximum loan amount, the lender is aware of the possibility that some borrowers may inflate their reported income using the low information contract in order to secure a higher loan amount than would otherwise be available. Thus, in the spirit of the Diamond (1991) model, we introduce three aspects of borrower heterogeneity into borrowers' contract selection decision: information verification costs, future credit concerns, and loan demand relative to income.[13]

We specify the borrower's reduced-form objective function on the basis of the amount of debt originated today and at some future date.[14] Specifically, the borrower's utility is expressed by the equation

$$U = u(L_1; \mu) - C_1 + \rho E\left[u(L_2; \mu) - C_2\right], \tag{1}$$

where L_t and C_t $(t = \{1, 2\})$ denote the debt amount and costs associated with the loan at period t. Parameter $\rho \in [0, 1]$ represents the borrower's probability of originating a future loan; a borrower has no concern about future credit access if $\rho = 0$ and has maximal concern if $\rho = 1$. We assume that $u(L; \mu)$ is a felicity function with $\partial u/\partial L > 0$, $\partial^2 u/\partial L^2 < 0$, and $\partial u/\partial \mu > 0$.[15] Parameter μ represents the borrower's loan demand. Loan demand is large if a borrower expects larger income growth, puts a higher utility weight on housing consumption, or is more tolerant of higher amounts of leverage.

Two types of loans are available to a borrower: full-doc loans and low-doc loans. For a full-doc loan, a borrower must prepare an income document. A borrower's true income y is private information, but the lender can verify this income by obtaining an appropriate document. For borrowers who have W-2 tax documents, the cost of producing an income verification document is low (c^L). In contrast, self-employed borrowers incur high income verification costs (c^H).

[13] Key differences between the two models are that we consider (1) the borrower's optimal choice, (2) debt instruments with complete and incomplete monitoring, (3) liquidity default, and (4) two borrowing opportunities for an individual borrower.

[14] This two-loan objective function can be derived from a standard consumption choice model, in which a borrower gains utility by intertemporally smoothing consumption or by owning a house that better matches her unique personal taste.

[15] We use $u(L; \mu) = \mu\sqrt{L}$ for analytical convenience, but another concave function such as a log utility function would give essentially the same result.

These documentation costs are measured in the unit of utility. We normalize c^L to be zero.

The lender uses the borrower's reported income to determine the loan amount. For a full-doc loan, the loan amount (L^F) is a linear function of the borrower's true income: $L^F = \alpha y$, where α is a constant DTI ratio. For a low-doc loan, a borrower reports stated income y^S. The stated income can deviate from the true income by an unobserved positive factor $x : y^S = xy$. The variable x represents the degree of income exaggeration. For a low-doc loan, the lender uses an alternative DTI ratio β to determine the loan amount: $L^N(x) = \beta y^S = \beta xy$.[16]

For simplicity, we model the mortgage as similar to a discount bond: the borrower receives the loan amount and pays the entire interest cost at origination, and then pays back the total loan amount at maturity. Between origination and maturity, the borrower regularly sets aside part of her income in a sinking fund to pay off the loan at maturity. The borrower defaults at maturity if she cannot build a sufficient fund due to negative income shocks during the term of loan. We abstract from stochastic income and collateral processes to keep the model simple. Instead, we assume that the probability of default $D \in (0, 1)$ is an increasing function of the relative DTI ratio: $D'(z) > 0, z \equiv \beta x/\alpha$.[17] When $z = 1$, the default probability is the same for the low-doc loan and full-doc loans because the ratio of the sinking fund payment to the initial true income is identical. As z increases, the borrower is less likely to accumulate a sufficient repayment fund because the annuity payment is large relative to the initial true income.

The lender cannot infer the borrower's loan demand from the loan amount because a large loan amount can arise from large loan demand or large income. Without verification, the lender has no information about the borrower's true income. The lender cannot infer the borrower's loan demand from a default event because a nonexaggerating borrower may also default on a loan. However, based on the inference about the average loan demand of a borrower group, the lender determines the loan interest rate. The interest rate for a full-doc loan is normalized to zero, and the interest spread for a low-doc loan is rL^N.

A W-2 or self-employed low-doc borrower may face higher future credit costs or be credit-rationed after originating the first loan with probability p, due to the lender's random audit.[18] However, the W-2 borrower can still arrange a standard full-doc loan in the second period (possibly from another lender).

[16] Technically, x could be less than one if the borrower wanted to underreport income. However, we view this as a relatively uninteresting and rare case since loan amounts are jointly determined by the borrower's DTI ratio and LTV ratio. If a borrower were to underreport, then the DTI ratio would be higher for a given loan amount. All else being equal, lenders view loans with higher DTIs as having higher default risk and subject them to increased underwriting scrutiny. As a result, these loans would face either an elevated probability of lender rejection or higher interest rates due to risk-based pricing reducing the incentive to underreport.

[17] For example, if $D(z) = (1 + \delta/z)^{-1}$, where δ is a positive constant, then $D(z)$ has the following properties: $\lim_{z\to 0} D(z) = 0$, $\lim_{z\to\infty} D(z) = 1$, and $D(1) = 1/(1 + \delta)$.

[18] If p represents the default probability, our result will be stronger because a borrower who exaggerates income and subsequently defaults will be more likely to face higher future credit costs or credit rationing.

In contrast, a self-employed borrower can only arrange a smaller low-doc loan: $L_D^N = \beta y$. Furthermore, the borrower pays a penalty that depends on the degree of income exaggeration in the first period: $\gamma x L_D^N$.

The utility gains from full-doc loans for W-2 (U_W^F) and self-employed borrowers (U_S^F) are, respectively,

$$U_W^F = u(L^F;\mu) + \rho u(L^F;\mu) \tag{2}$$

and

$$U_S^F = u(L^F;\mu) - c^H + \rho(u(L^F;\mu) - c^H). \tag{3}$$

The utility gains from low-doc loans for W-2 (U_W^N) and self-employed borrowers (U_S^N) are, respectively,

$$\begin{aligned} U_W^N(x) = {} & u(L^N(x);\mu) - rL^N(x) \\ & + \rho[pu(L^F;\mu) + (1-p)(u(L^N(x);\mu) - rL^N(x))] \end{aligned} \tag{4}$$

and

$$\begin{aligned} U_S^N(x) = {} & u(L^N(x);\mu) - rL^N(x) \\ & + \rho\left[p\left(u\left(L_D^N;\mu\right) - rL_D^N - \gamma x L_D^N\right) + (1-p)(u(L^N(x);\mu) - rL^N(x))\right]. \end{aligned} \tag{5}$$

We first analyze a borrower's utility-maximizing choice of income exaggeration for a low-doc loan, given a loan cost r. We then analyze the borrower's choice between a low-doc loan and a full-doc loan. Details of the solution are outlined in the Internet Appendix. We obtain the following three propositions.

PROPOSITION 1: *The level of income exaggeration is equal to*

$$x_W = \frac{\mu^2}{4r^2\beta y}$$

for W-2 borrowers and self-employed borrowers without concerns about future credit access, and is equal to

$$x_S = Ax_W, \text{ where } A \equiv \left[1 + \frac{\rho p \gamma}{(1+\rho(1-p))r}\right]^{-2} \in (0,1],$$

for self-employed borrowers with concerns about future credit rationing. Thus, $x_S \leq x_W$.

The degree of income exaggeration (x) is small if loan demand (μ) is small, the interest cost (r) is large, or the borrower can arrange a large loan amount on the basis of true income (βy). The difference in income exaggeration between a self-employed borrower and other borrowers is greater if the penalty for untruthful reporting is more severe (γ is larger), the probability of detection is greater (p is larger), or the self-employed borrower has greater concerns about

future access to credit (ρ is large). Note that the amount of a low-doc loan does not depend on β because the borrower can adjust her stated income in response to the lender's DTI criterion. It is straightforward to link the degree of income exaggeration to the probability of default.

PROPOSITION 2: *The probability of default is lower for a self-employed borrower who has greater concerns about future access to credit than for an otherwise identical W-2 borrower or a self-employed borrower without concerns about future credit availability. Specifically, the default probability is*

$$D\left(\frac{\beta x_W}{\alpha}\right) \geq D\left(\frac{\beta x_S}{\alpha}\right).$$

The above equation holds with equality if $\rho p\gamma = 0$.

A borrower chooses between a full-doc loan and a low-doc loan on the basis of the relative utility benefit. The utility benefit of a low-doc loan over a full-doc loan for a W-2 borrower is

$$B_W^N(\mu) \equiv U_W^N(x_W) - U_W^F = (1 + \rho(1-p))\left(\frac{\mu^2}{4r} - \sqrt{\alpha y}\mu\right). \tag{6}$$

For a self-employed borrower, the utility benefit is

$$B_S^N(\mu) \equiv U_S^N(x_S) - U_S^F = \theta_1\mu^2 + \theta_2\mu + \theta_3, \tag{7}$$

where $\theta_1 > 0, \theta_2 < 0$, and $\theta_3 \equiv (1+\rho)c^H - \rho pr\beta y$ are specified in the Internet Appendix. Both (6) and (7) are convex quadratic functions of μ. The former takes a value of zero when $\mu = \mu^* \equiv 4r\sqrt{\alpha y}$. The latter exhibits the following properties: $B_S^N(0) = \theta_3$ and $\min B_S^N(\mu) = \theta_3 - \theta_2^2/4\theta_1$. Depending on the value of θ_3, the solution to $B_S^N(\mu) = 0$ has zero, one, or two roots. Using these properties, we obtain the following proposition.

PROPOSITION 3: *W-2 borrowers, irrespective of their future credit availability concerns, choose low-doc loans if and only if loan demand* μ *is greater than* $\mu_W^* \equiv 4r\sqrt{\alpha y}$. *Self-employed borrower choice depends on the cost of income verification. For* $c^H > (\rho pr\beta y + \theta_2^2/4\theta_1)/(1+\rho)$, *all self-employed borrowers choose low-doc loans. For* $c^H < \rho pr\beta y/(1+\rho)$, *a self-employed borrower chooses a low-doc loan if and only if* $\mu > \mu_S^* = -\frac{\theta_2}{2\theta_1} + \sqrt{\frac{\theta_2^2}{4\theta_1^2} + \frac{\theta_3}{\theta_1}}$; *otherwise, a self-employed borrower chooses a low-doc loan if and only if* $\mu > \mu_S^*$ *or* $\mu < \mu_S^{**} = -\frac{\theta_2}{2\theta_1} - \sqrt{\frac{\theta_2^2}{4\theta_1^2} + \frac{\theta_3}{\theta_1}}$.

On the basis of the comparative statics of μ_W^*, μ_S^*, and μ_S^{**}, more borrowers will choose a low-doc loan if the low-doc loan is less costly (r is smaller) or the full-doc loan amount is small (αy is small). In addition, more self-employed borrowers will choose a low-doc loan if the income verification cost is larger (c^H is large). As a consequence, when the income verification cost is sufficiently large, use of a low-doc loan should be more prevalent in a self-employed sample than in a W-2 sample.

In equilibrium, the lender will charge a positive interest rate premium on low-doc loans, recognizing that borrowers who have stronger incentives to exaggerate income will select low-doc loans. Moreover, the rate premium will be greater for W-2 borrowers because the average default risk of W-2 low-doc borrowers is higher than that of self-employed low-doc borrowers. Furthermore, if the lender can estimate the level of income falsification of an individual borrower, the lender may charge a larger rate premium for a high estimated value of income falsification. Although the rate premium will mitigate adverse selection and untruthful income reporting, it will not completely eliminate these problems. By increasing the premium, the lender faces a trade-off between the benefit of mitigating these problems and the cost of decreasing total loan volume. By charging a sufficiently high premium to completely eliminate these problems, the lender will forgo opportunities to extend low-doc loans to borrowers who only moderately exaggerate income. Thus, the problems of adverse selection and untruthful reporting will persist in equilibrium.

To summarize, the insights derived from our theoretical model lead to the following empirical predictions. First, low-doc loans will be preferred by borrowers with high information verification costs, namely, self-employed borrowers (Section II). Second, the ex post probability of default will be lower for self-employed low-doc borrowers than for W-2 low-doc borrowers (Section III). Third, on average, borrowers will exaggerate income for low-doc mortgages, and the level of income falsification will be higher among W-2 borrowers than self-employed borrowers (Section IV). Fourth, there will be a positive mortgage rate premium for low-doc loans, and the premium will be larger for W-2 low-doc borrowers than self-employed low-doc borrowers (Section V). Finally, the rate premium will be positively related to income falsification (Section V).

II. Data and Summary Statistics

A. Data

The main data set used in the analysis contains loans originated by New Century Financial Corporation (New Century). New Century was one of the largest subprime lenders in the run-up to the recent mortgage crisis, with a large portion of its business originated through independent mortgage brokers. Along with originations, New Century serviced mortgage loans and held a portfolio of loans as investments. New Century collected detailed borrower and collateral information at the time of origination, as well as contractual features of the loans. Also, for the loans that New Century serviced, monthly mortgage performance data are available.

From the loan origination records, we identify a borrower's employment type (W-2 versus self-employed) and the level of income documentation (full-doc versus stated income.)[19] We focus on first-lien loans with complete servicing data that were originated through the mortgage broker channel between 1998

[19] Although proof of income is not required on low-doc loans, verification of employment is required. For W-2 borrowers, this usually entails a verbal verification of employment from the

and 2005. Following Conklin (2016), to limit the effect of outliers and data entry errors, we exclude loans for which: (1) total fees are negative or greater than 15% of the loan amount; (2) the yield spread premium paid from the bank to the broker is negative or greater than 5% of the loan amount; (3) the combined LTV at origination is negative or greater than 125%; (4) the borrower's FICO score is less than 450 or greater than 850; (5) the borrower's DTI ratio is negative or greater than 60%; (6) the borrower's monthly income is negative or greater than $26,900; and (7) borrower age is less than 18 or greater than 99. The final sample includes 458,872 funded mortgage loans.

We obtain additional data from several other sources. Market interest rate data come from the Federal Reserve Bank of St. Louis's Federal Reserve Economic Data and Freddie Mac's Primary Mortgage Market Survey. Monthly metropolitan statistical area (MSA) unemployment rates come from the Bureau of Labor and Statistics. Time-varying MSA-level house price indices come from the Federal Housing Finance Agency, and ZIP code-level income information is obtained from the 2000 Census and IRS individual income tax statistics. Finally, the Pahl Index for mortgage broker regulations at the state level comes from Pahl (2007); higher values of the Pahl index indicate stricter regulation of brokers at the state level.

B. Summary Statistics

Table I provides definitions for the variables used throughout the analysis. Summary statistics for key variables of interest separated by employment status and loan type are presented in Table II.[20] We find that 21% of the borrowers are self-employed, with the remainder classified as W-2. Consistent with New Century's concentration in the subprime market niche, nearly 40% of the mortgages are low-doc loans. In comparison, Paley and Tzioumis (2011) document that roughly one-third of all loans originated between 2000 to 2007 were low/no-doc loans. We also find that 5% of the loans fall at least 60 days behind on their mortgage within the first 24 months after origination. Since New Century sold the majority of its loans within six months of origination, the observed default is a lower bound on the actual default rate.[21] Furthermore, our loan sample covers loans originated from 1998 to 2005, with performance data ending in early 2007. Since this covers a period prior to the financial crisis during which house prices were generally rising, most of the loans in the

borrower's employer. For self-employed borrowers, lenders typically require a signed letter from a CPA or copies of business licenses. The New Century data set contains a field indicating whether the borrower is self-employed. Throughout the paper, we will refer to all borrowers who are not self-employed as W-2 borrowers.

[20] Table IA.III in the Internet Appendix reports the complete set of summary statistics.

[21] Some of the loans that exit the sample due to the transfer of servicing rights likely defaulted in a later period. Unfortunately, we cannot distinguish between loans that prepaid and loans for which the servicing rights were transferred, and thus, standard techniques for handling competing risks with censored data cannot be employed.

Table I
Variable Definitions

This table presents the variable names and descriptions used in subsequent tables.

Variable	Definition
Loan Characteristics	
RATE_SPREAD	The difference between the rate on the mortgage and the two-year Treasury rate
CLTV	Combined loan-to-value ratio at origination
LOAN_AMOUNT	Loan amount at origination
FEES	Fees as a percentage of loan amount
ARM	An indicator set to one if the loan was an adjustable-rate mortgage
PREPAY	An indicator set to one if there was a prepayment penalty on the loan
PURCH	An indicator variable set to one if the loan was for a home purchase
CASH	An indicator variable set to one if the loan was a cash-out refinance
FICO	The FICO score of the primary borrower at origination
IO	An indicator set to one if the loan had interest-only payments
MONTHS	Months since origination
Property Characteristics	
INVESTMENT	An indicator set to one if the property was an investment property
TWO_UNIT	An indicator set to one if the property was a two-unit property
CONDO	An indicator set to one if the property was a condo
Borrower Characteristics	
AGE	The age of the primary borrower
MINORITY	An indicator set to one if the borrower was a minority
INCOME	The total monthly income of the borrowers
FACE	Face-to-face interview between broker and borrower
Interest Rate Environment	
RATE_30	The average monthly prime 30-year fixed rate at the time of origination
Area Characteristics	
UNEMP	Monthly unemployment rate at the MSA level
HHI	MSA-level Herfindahl-Hirschman index for broker competition
REG	Pahl index for state-level broker regulations
HPI_2YR	MSA house price growth over previous two years
HPI_GROWTH	MSA house price growth since origination

sample had not yet experienced significant declines in house prices that would trigger negative-equity-induced default.[22]

[22] We also confirmed that the reported default rate in the New Century data is roughly comparable to default rates on subprime mortgages as reported in the BlackBox (BBX) data. For example, for subprime loans originated in 2004, BBX reports an average 24-month default rate of 7.3%, compared to the average default rate of 5% in the New Century data. In addition, to mitigate concerns about the degree to which the New Century loans are representative of the overall subprime market, we compared our sample to the loans in Demyanyk and Van Hemert (2011), a highly cited paper on the subprime mortgage crisis. Their sample spans many subprime lenders and covers roughly half of the subprime mortgage market. Table IA.IV in the Internet Appendix compares descriptive statistics for loans originated in 2004 and 2005 (the years with the most originations in our data) in Table 1 of Demyanyk and Van Hemert (2011) with the New Century loans. The

Table II

Summary Statistics by Employment and Documentation Type

This table presents the mean of key variables for the funded loans from the New Century database. For continuous variables, the standard deviation is reported in parentheses. Summary statistics and definitions for all variables are presented in Table IA.III of the Internet Appendix.

	W-2		Self-Employed	
	Full-Doc (1)	Low-Doc (2)	Full-Doc (3)	Low-Doc (4)
DEFAULT	0.0478	0.0525	0.0512	0.0505
RATE SPREAD	4.6351	4.9516	4.5239	4.7337
	(1.5288)	(1.4065)	(1.4989)	(1.4566)
CLTV	82.9510	83.5768	83.6182	83.6585
	(14.2817)	(14.0507)	(13.4387)	(14.9095)
LOAN AMOUNT	174,019	207,296	224,821	231,555
	(107,689)	(115,962)	(138,936)	(134,122)
ARM	0.7234	0.8311	0.7316	0.8285
PURCHASE	0.2802	0.4186	0.3203	0.4327
CASH	0.5974	0.5063	0.5669	0.4986
FICO	600.5364	628.0193	613.4986	632.4919
	(59.7866)	(61.7492)	(59.2660)	(63.2627)
AGE	43.1441	40.6433	43.9410	42.9128
	(11.6602)	(10.8312)	(11.1064)	(11.0918)
MINORITY	0.4126	0.4187	0.3248	0.3405
INCOME	5,396	6,398	8,465	8,251
	(3,014)	(3,199)	(5,245)	(4,403)
MIDWEST	0.2088	0.1588	0.1406	0.1276
SOUTH	0.2543	0.2198	0.2280	0.2330
NORTHEAST	0.1219	0.1674	0.0941	0.1197
PACIFIC	0.0079	0.0073	0.0122	0.0099
1999	0.0216	0.0137	0.0363	0.0216
2000	0.0273	0.0173	0.0468	0.0193
2001	0.0464	0.0366	0.0663	0.0344
2002	0.0972	0.0836	0.1004	0.0702
2003	0.2238	0.1924	0.1973	0.1923
2004	0.2803	0.3537	0.2353	0.2695
2005	0.2921	0.2923	0.2956	0.3815
N	256,747	107,621	19,823	74,681

Turning to loan characteristics, the average interest rate spread is 4.72%, and an overwhelming majority are adjustable rate mortgages (ARMs).[23] The mean loan amount is $193,000 with a combined LTV (CLTV) ratio at origination of 83%. Furthermore, 34% of the loans are originated to purchase a home,

samples appear to be quite similar, although the New Century data include a larger proportion of low-doc loans.

[23] The rate spread is the initial contract rate minus the two-year constant maturity Treasury rate at the time of origination. The average note rate on the mortgages is 7.68%, and the ARMs are actually "hybrid ARMs," with an initial fixed-rate period (typically two years) after which the interest rate adjusts every six months.

while 56% are refinance loans with the borrower extracting equity (CASH). The remaining 10% of loans are rate/term refinances. These are generally cases in which the borrower is refinancing to obtain a lower interest rate than the rate on the current mortgage. The average FICO score is 613. Taken together, the summary statistics clearly reflect the fact that New Century was primarily a subprime lender with mortgages originated to higher risk borrowers.

In terms of observable borrower characteristics, Table II shows that the average borrower is 43 years old with an income of $6,200 per month. In addition, 40% of the borrowers are minorities, and a large share of mortgage loans (44%) were originated in the West region of the United States as classified by the U.S. Census Bureau. Since New Century began its operations in California, the strong focus in the West is not surprising. Furthermore, consistent with the entire subprime market, New Century experienced significant growth from 2000 through 2005 (Chomsisengphet and Pennington-Cross (2006)).

Table II shows that there are several key differences across the borrower groups considered. First, consistent with first and second predictions in Section I, loans to self-employed borrowers are much more likely to be low-doc (79% of the self-employed subsample are low-doc loans, compared to 30% for the W-2 subsample.) This is not surprising since the low-doc product was designed specifically for borrowers with difficult-to-verify financial situations. Also, the average loan amount in the W-2 subsample is $46,000 lower than that for the self-employed group. Consistent with the difference in average loan sizes, self-employed borrowers report a higher average income. Finally, the average FICO score is higher in the self-employed subsample.

Since the summary statistics suggest that differences exist across the four borrower $\times$ loan product groups (low-doc self-employed, low-doc W-2, full-doc self-employed, and full-doc W-2), in Figure IA.8 in the Internet Appendix, we plot kernel density distributions for borrower and mortgage characteristics. First, we see that the credit risk distributions for full-doc loans (W-2 and self-employed) are wider and skewed more to the left than those for low-doc loans. This is consistent with the lender imposing a higher underwriting screen on low-doc mortgages, where borrowers have greater opportunity to embellish their debt payment capacity. Second, the borrower income distribution for full-doc W-2 loans is skewed more to the left than the other groups. In terms of borrower age, we see little difference across the four groups' kernel density distributions. Turning to loan characteristics, Figure IA.8 in the Internet Appendix reveals a sizable difference in the distribution of mortgage amounts between full-doc W-2 borrowers and the other three groups. Figure IA.8 also reveals an interesting difference in loan pricing across the four groups. In particular, it appears that full-doc W-2 borrowers have a higher proportion of high-fee mortgages, and the interest rate spread on full-doc loans (regardless of whether to a W-2 borrower or self-employed borrower) is essentially the same. However, the interest rate spread distribution for low-doc W-2 borrowers is skewed more to the right. It therefore appears that, from a pricing perspective, the lender did anticipate that borrowers with W-2s who selected low-doc loans were potentially a higher risk and priced those loans accordingly. Interestingly,

full-doc W-2 borrowers tended to pay higher origination fees (as a percentage of their loan amount) than low-doc borrowers.

To summarize, Table II and Figure IA.8 indicate that several important differences exist between full-doc and low-doc loans according to borrower information verification cost type. First, the data support our theoretical prediction that borrowers with high information verification costs (self-employed) will prefer low-doc loans. Second, borrowers with low information verification costs (W-2 borrowers) that select the low-doc loan product have higher average reported incomes and loan amounts than similar borrowers who select full-documentation loans. Third, low-doc loans for W-2 borrowers experience higher levels of ex post default. Fourth, we do not observe a similar pattern for borrowers with high information verification costs. For self-employed borrowers, the average income and loan amount are similar regardless of loan type. Furthermore, low-doc loans to self-employed borrowers do not have higher average default rates. Taken together the summary statistics provide preliminary evidence consistent with the popular view that low-doc loans were "liar's loans," but the role of borrower concerns, to preserve access to credit, may have ameliorated this tendency as low-doc loans to self-employed borrowers do not appear to have the same level of income overstatement, loan amount distortion, or increased mortgage default risk.

III. Borrower Type and Mortgage Performance

The univariate analysis in the previous section confirms our first prediction that low-doc loans are preferred by self-employed borrowers. The unconditional analysis in the previous section also supports our second prediction: that W-2 low-doc borrowers are riskier than comparable self-employed borrowers. In this section, we conduct multivariate analysis in which we compare ex post default rates conditional on borrower characteristics observable at loan origination as well as macroeconomic factors and changes in house prices and interest rates after origination. In particular, we estimate the following loan-level regression of mortgage default:

$$\Pr(DEFAULT_i) = \Phi(\alpha + \beta_1 W2_i + \beta_2 Lowdoc_i + \beta_3 W2_i \times Lowdoc_i + \delta X_i + \theta R + \vartheta W + \gamma T), \tag{8}$$

where $DEFAULT_i$ is an indicator for mortgage default for loan i and Φ is the standard normal cumulative distribution function.[24] The vector X_i contains information collected and recorded on the loan application. This information includes loan characteristics (fees charged on the loan, loan amount, CLTV, whether the loan has a prepayment penalty, purchase or refinance, cash-out or

[24] The default variable takes a value of one if the loan becomes 60 or more days delinquent within 24 months of origination. In robustness checks reported in Table IA.VIII, we used alternate windows for delinquency (12 and 36 months). The results were qualitatively unchanged. Unfortunately, data limitations prevent us from observing loan performance during the mortgage crisis since payment history is only available through the beginning of 2007 in the New Century database.

rate/term refinance, whether the payments are interest-only), property characteristics (two-unit, condominium, owner-occupied, or investment property), and borrower characteristics (FICO score, borrower age, borrower income, DTI ratio, whether the borrower met in person with the loan officer, minority status). The variable R captures market interest rates at the time of origination. Area characteristics, W, include the monthly MSA unemployment rate, the level of broker competition, the Pahl index capturing the level of broker regulation at the state level, and the census region (West, Midwest, South, Northeast, or Pacific).[25] Since mortgage defaults are closely related to house prices, W also includes MSA-level house price changes in the two years leading up to origination as well as MSA-level house price changes between origination and the last month the loan is observed in the performance data.[26] Finally, T contains variables denoting mortgage origination year to control for loan cohort effects. Throughout the analysis, unless otherwise stated, the reported standard errors are robust to heteroskedasticity and within-cluster correlation of errors at the MSA level.

The parameters β_1, β_2, and β_3 are the primary coefficients of interest and capture the differential effect of borrower concerns about future credit access on the probability of default. Specifically, β_1 represents the difference in outcome for borrowers with low information verification costs (i.e., the employment type is W-2), β_2 captures the difference in outcome when the loan type is low-doc, and $\beta_1 + \beta_2 + \beta_3$ reflects the difference in outcome for borrowers with the least concerns about future credit access, that is, borrowers with low information verification costs ($W2 = 1$) who originate a low-information-content mortgage ($Lowdoc = 1$).

Table III presents estimated marginal effects from maximum likelihood estimation of equation (8). Since Ai and Norton (2003), Williams (2012), and Buis (2010) note that interpreting a single marginal effect of an interaction term in a nonlinear model can be problematic, we follow Williams (2012) and report the marginal effects of low-doc at representative values for borrower employment type (i.e., at values of zero and one for $W2$).[27] In column (1), the marginal effects indicate that low-doc loans are associated with higher ex post default rates, regardless of employment type. This is consistent with the increased risk associated with low-doc loans and supports the pricing effect observed in Table II. However, the difference in magnitude between the effects for self-employed and W-2 borrowers shows a more complex relationship and is consistent with

[25] Broker competition is computed as the quarterly Herfindahl-Hirchman index in each MSA as in Ambrose and Conklin (2014).

[26] Although we report results using preorigination MSA house price changes over a two-year period, results are insensitive to other window lengths (e.g., one-, three-, and five-year house price changes).

[27] We also calculate a single estimate for the marginal effect of the interaction term using marginal effects at the sample means. The results are qualitatively unchanged. Williams (2012) provides a detailed discussion of the differences between average marginal effects and marginal effects at the mean at http://www3.nd.edu/~rwilliam/stats/Margins01.pdf. As a robustness check, we employed a linear probability model of default. Consistent with the findings reported in Table III, we find that the relationship between low-doc and default is driven by W-2 borrowers. Table IA.V in the Internet Appendix reports the estimated marginal effects for this specification.

Table III
Relationship between Low-Doc, Employment Type, and Mortgage Performance

This table presents the marginal effects of low-doc loan status on default by employment type. Columns (1) and (2) are the baseline model without and with MSA fixed effects, respectively. Column (3) presents the marginal effects of low-doc loan status on default by employment type after matching low-doc loans with full-doc loans based on the propensity score. The propensity score is derived using loan, property, borrower, and area characteristics as well as MSA and origination year. Column (4) presents marginal effects of low-doc loan status on default by employment type for the subsample of loans where the borrower's liquid assets field is not missing in the New Century data. The marginal effects are derived from a probit model of mortgage performance on income documentation, employment type, an interaction term between income documentation and employment type, loan characteristics, borrower characteristics, property characteristics, and area characteristics for the funded loans from the New Century database. Heteroskedasticity-robust standard errors adjusted for clustering at the MSA level are reported in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% level, respectively.

	Baseline Model		Propensity Score Matching	Controlling Borrower's Liquid Assets
Dependent Variable: Default	M.E. (1)	M.E. (2)	M.E. (3)	M.E. (4)
Self-Employed (Low-Doc)	0.0053***	0.0054***	0.0043**	0.0062
	(0.0019)	(0.0019)	(0.0020)	(0.0060)
W-2 (Low-Doc)	0.0124***	0.0119***	0.0111***	0.0090***
	(0.0013)	(0.0012)	(0.0012)	(0.0022)
Loan Characteristics	Yes	Yes	Yes	Yes
Property Characteristics	Yes	Yes	Yes	Yes
Borrower Characteristics	Yes	Yes	Yes	Yes
Interest Rate Environment	Yes	Yes	Yes	Yes
Area Characteristics	Yes	Yes	Yes	Yes
Origination Year Fixed Effects	Yes	Yes	Yes	Yes
MSA Fixed Effects	No	Yes	Yes	Yes
N	455,546	455,058	269,662	50,231
Mean Default Rate (Full-Doc Self)	0.0512	0.0512	0.0505	0.0426
Mean Default Rate (Full-Doc W-2)	0.0478	0.0478	0.0412	0.0422
Log Likelihood	−77,364	−76,716	−45,197	−7,853

borrower concerns about future credit access mitigating default risk. For borrowers with the greatest such concerns (self-employed borrowers), the marginal effect of *Lowdoc* is modest (0.53%). To place this result in perspective, dividing the marginal effect by the mean default rate (0.0053/0.0512) indicates that self-employed borrowers originating low-doc loans have a 10.4% higher probability of default than the reference group (self-employed borrowers originating full-doc loans.) In contrast, for borrowers with the least concerns about future credit access (W-2 borrowers), moving from a full-doc to a low-doc mortgage is associated with a 25.9% increase about the mean in mortgage default, ceteris paribus.[28] In other words, low-doc loans to self-employed borrowers pose

[28] This is calculated by taking the ratio of the marginal effect to the average W-2 borrower default rate (0.0124/0.0478).

modest additional default risk, consistent with the prediction that borrowers with high information verification costs value the ability to obtain credit. However, low-doc loans to W-2 borrowers have substantially higher default rates, in line with the view that they are less concerned about being credit-rationed in the future since they can easily switch to full-doc mortgages in the future.

Although we include time-varying controls at the MSA level to account for local economic conditions (e.g., pre- and postorigination house price changes and unemployment), the possibility remains that unobserved time-constant geographic effects are driving the observed effects. In a robustness check, we address this concern by including MSA fixed effects (column (2)).[29] The results are virtually identical and confirm that low-doc mortgages have a higher likelihood of default, but that the marginal effect is much larger for W-2 borrowers.

Another concern with our analysis is that, while our regression framework controls for all observable information available at loan origination, there might exist omitted variables bias. Thus, in an additional robustness check, we present two alternative specifications in columns (3) and (4). In column (3), we use a propensity score matching approach. We match low-doc W-2 observations with full-doc W-2 observations using a nearest-neighbor propensity score based on observable loan, borrower, and geographic characteristics. We use the same matching procedure for self-employed low-doc observations. After creating our matched sample, we repeat the estimation of equation (8). The results remain qualitatively unchanged. Next, because we are able to observe the borrower's liquid assets (e.g., checking, savings, stocks, etc.) for a subsample of the borrowers, in column (4), we repeat our main default regression controlling for borrower liquid assets. We again find that our primary results remain unchanged.

To further investigate the low-doc effect, we divide the sample according to the purpose of the debt. First, we examine differences across purchase and refinance loans. Second, we split the sample according to residence type (e.g., investment versus noninvestment property). In the first analysis, we split our sample into four categories according to employment type and loan purpose (W2/purchase, W2/refinance, self-employed/purchase, and self-employed/refinance). Panel A of Table IV presents the average default rate for each subsample. We see that the average default rate for refinancing borrowers is lower than that for purchase borrowers. We next estimate equation (8) for each subsample and report the marginal effect of *Lowdoc* in Panel B of Table IV.[30] The marginal effects show that borrowers selecting low-doc loans (*Lowdoc*) have a higher probability of default in all subsamples except for the self-employed/refinance category. To put the marginal effects into perspective, Panel C divides the marginal effects by the mean probability of default for each

[29] Since several MSAs had no defaults, the number of observations included in the regression in column (2) is lower than in column (1).

[30] Full tables are available from the authors upon request. Since the probit regressions are run separately for each subsample, $W2$, $Lowdoc \times W2$, and the purchase indicator variable are not included in the regressions. We include all other control variables from equation (8) in the estimation.

Table IV
Employment Type, Loan Purpose, and Default

Panel A presents the mean (and standard deviation) of default for each of the subsamples. Panel B reports the marginal effects of low-doc from four separate probit regressions of default on income documentation, loan characteristics, borrower characteristics, property characteristics, and area characteristics for the funded loans from the New Century database. Heteroskedasticity-robust standard errors adjusted for clustering at the MSA level are reported in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% level, respectively. Panel C reports the increase in sample mean probability of default due to low-doc.

Loan Purpose	W-2 (1)	Self-Employed (2)
Panel A: Sample Mean Probability of Default (Standard Deviation)		
	Mean (Std. Dev.)	Mean (Std. Dev.)
Purchase	0.0616 (0.2404)	0.0507 (0.2194)
Refinance	0.0433 (0.2036)	0.0505 (0.2191)
Panel B: Marginal Effect (M.E.) of Low-Doc on Default (Standard Error)		
	M.E.	M.E.
Purchase	0.0192*** (0.0020)	0.0154*** (0.0035)
Refinance	0.0085*** (0.0010)	0.0035 (0.0022)
Panel C: Percentage Increase in Sample Mean Default Probability Due to Low-Doc (Panel B/Panel A)		
Purchase	31%	30%
Refinance	20%	7%

subsample. The top row in Panel C (purchase mortgages) shows that low-doc loans have a similar effect on default risk for purchase mortgages, regardless of whether they are originated by self-employed or W-2 borrowers. In contrast, the bottom row in Panel C indicates that low-doc refinancing loans have higher default risk than full-doc refinancing mortgages, but the difference between W-2 and self-employed borrowers is striking.

In our second analysis, we investigate whether the increase in risk on low-doc loans is associated primarily with investment properties. To do so, we split our sample into loans used to finance the borrower's primary or secondary residence versus loans used to finance investment properties.[31] We then estimate equation (8) for each sample.[32] For both property types, the marginal effect

[31] As noted in Piskorski, Seru, and Witkin (2015), we recognize that our measure of borrower owner-occupancy status may be biased. However, it is unclear how such bias may affect the results.

[32] Table IA.VII in the Internet Appendix reports the marginal effects.

of low-doc is significant and positive for W-2 borrowers. Moreover, the size of the effect is nearly identical across property types. Thus, consistent with our previous results, regardless of the property type, low-doc appears to have a smaller impact on default risk for self-employed borrowers.

To summarize, Table III provides several key insights that are consistent with our second prediction. First, full-doc loans to self-employed borrowers are ex post marginally riskier than full-doc loans to W-2 borrowers. This makes sense as income for self-employed borrowers is likely more volatile. Second, low-doc loans are riskier than full-doc loans. Third, and most importantly, a distinction exists between low-doc loans originated to self-employed borrowers and low-doc mortgages originated to W-2 borrowers. In particular, consistent with our prediction that preserving access to future capital is valuable, the magnitude of the change in default risk is considerably larger for W-2 borrowers originating low-doc loans.

IV. Income Exaggeration and Mortgage Performance

In the previous section, we establish that the majority of the increased risk associated with low-doc mortgages resulted from the set of borrowers who were capable of verifying income at relatively low cost by providing a W-2 statement. Having established that the problems documented with the low-doc product arose from a particular set of borrowers, we now test our third prediction by exploring the interaction of adverse selection and expectations regarding future access to credit with respect to borrower income falsification as a possible causal link for this increased risk.

To measure income exaggeration, we follow the method outlined in Jiang, Nelson, and Vytlacil (2014) and estimate a semi-log model of borrower income as a function of borrower characteristics (credit rating, race, sex, and age), area characteristics (income per capita in the borrower's ZIP code and house price growth over the previous two years in the borrower's MSA), loan amount, an indicator for whether the property is an investment property, origination year dummies, and state dummies.[33] Table IA.V in the Internet Appendix reports results for the borrower income model estimation.[34]

[33] Two potential issues may result from including the loan amount as a control variable. First, by including the loan amount as an explanatory variable in the income regression, we introduce a conservative bias into our estimation of income falsification. This bias may arise since borrower income is one of the metrics used in mortgage underwriting to determine the loan amount. Thus, in estimating income falsification, our method will tend to have higher predicted incomes for low-doc loans (and thus underestimate income falsification) if these borrowers used inflated incomes to qualify for higher loan amounts. A second and closely related concern is that the loan amount is endogenous. Since we are primarily interested in predictive accuracy, we do not view this as a major concern. Results in later sections that rely on our income estimates are not materially affected if we exclude the loan amount from the income regressions.

[34] We use the coefficients reported in Table IA.V in the Internet Appendix to compute estimates of income for the full-doc (in-sample) and low-doc (out-of-sample) loan borrowers. To calculate an estimate of income exaggeration (INC_EXAG), we subtract the estimated income from the reported income. Since estimated and reported incomes are both in logs, INC_EXAG represents the

Table V
Descriptive Statistics for Income Exaggeration Measure by Employment and Documentation Type

This table presents descriptive statistics for the income exaggeration measure. Column (1) presents the mean of income exaggeration for W-2 borrowers split according to income documentation. Column (2) presents the mean of income exaggeration for self-employed borrowers split according to income documentation. In columns (1) and (2), ***, **, and * indicate rejection of the null hypothesis that income exaggeration is equal to zero at the 1%, 5%, and 10% level of significance, respectively. Column (3) includes the mean difference, with ***, **, and * indicating significance at the 1%, 5%, and 10% level, respectively, from a mean difference test assuming unequal variance across the two groups.

Document Type	W-2 [1]	Self-Employed [2]	Mean Difference [3]
Full-Doc	0.0004	0.0009	−0.0005
	(0.0007)	(0.0031)	(0.0026)
Low-Doc	0.0799***	0.0106***	0.0693***
	(0.0010)	(0.0013)	(0.0016)

Table V presents descriptive statistics for *INC_EXAG* across employment and documentation type. For low information verification cost borrowers (W-2 borrowers) originating low-doc loans, the average estimated income overstatement is approximately 8%. In comparison, the average income overstatement associated with full-doc self-employed mortgages is 1%. For both W-2 and self-employed borrowers, *INC_EXAG* is significantly different from zero.[35]

To formally identify the extent of income falsification, we estimate the following regression:

$$INC\_EXAG_i = \alpha + \beta_1 W2_i + \beta_2 Lowdoc_i + \beta_3 W2_i \times Lowdoc_i + \delta X_i + \theta R + \vartheta W + \gamma T + \varepsilon_i, \quad (9)$$

where INC_EXAG_i is our measure of income exaggeration, and X_i, R, W, and T are defined as in equation (8).[36] Equation (9) tests whether borrowers selecting low-doc loans are correlated with our measure of income exaggeration and whether this effect depends on employment status. Table VI reports the results when we estimate equation (9) using OLS.[37] We first find that the

percentage difference between the borrower's reported income and estimated income. We winsorize *INC_EXAG* at the 1% level, but the main results are unchanged without winsorization.

[35] By construction, *INC_EXAG* is not different from zero for the full-doc loans.

[36] Equation (9) is analogous to explaining the residuals from equation (8), so we do not include the control variables from equation (8) in equation (9). The reason we say "analogous" is because *INC_EXAG* includes both in-sample (full-doc) and out-of-sample (low-doc) estimates. Results are qualitatively unchanged if we include controls from equation (8) in equation (9).

[37] Since some of the observations in our full sample are missing information used in the income prediction model, regressions using our income exaggeration measure have a somewhat smaller sample size. For example, the sex of the primary borrower, which is used as an explanatory variable in our income prediction model, is missing for 5,144 observations.

Table VI
Explaining Income Exaggeration

This table presents coefficient estimates from an OLS regression of *INC_EXAG* on employment type, type of income documentation, loan characteristics, borrower characteristics, property characteristics, and area characteristics for the funded loans from the New Century Database. Heteroskedasticity-robust standard errors adjusted for clustering at the MSA level are reported in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% level, respectively.

Dependent Variable: Income Exaggeration	
Employment Type	
W2	0.0002
	(0.0012)
Income Documentation	
Lowdoc	0.0046
	(0.0108)
Lowdoc × W2	0.0662***
	(0.0094)
Loan Characteristics	Yes
Property Characteristics	Yes
Borrower Characteristics	Yes
Interest Rate Environment	Yes
Area Characteristics	Yes
Origination Year Fixed Effects	Yes
MSA Fixed Effects	Yes
Constant	−0.2293***
	(0.0195)
N	449,916
Adj. R^2	0.03

parameter estimate for *W2* is small and not statistically significant, suggesting no material difference on average in income exaggeration between W-2 and self-employed borrowers. We next find that *Lowdoc* is positively related to our measure of income exaggeration, but again is not statistically significant. Further, the positive and statistically significant coefficient on the interaction term (*Low-Doc* × *W2*) indicates that income exaggeration is more pronounced in low-doc loans when the borrower is likely to have less concern about future credit access (i.e., W-2 borrowers).[38] These results are consistent with the

[38] As a robustness check, we estimate the following probit model of income exaggeration:

$$\Pr(INC\_EXTREME_i) = \Phi(\alpha + \beta_1 W2_i + \beta_2 Lowdoc_i + \beta_3 W2_i \times Lowdoc_i + \delta X_i + \theta R + \vartheta W + \gamma T),$$

where $INC\_EXTREME_i$ is a dummy variable equal to one if INC_EXAG_i is in the top decile for the borrower's employment type, and X_i, R, W, and T are as defined in equation (8). Table IA.IX in the Internet Appendix reports the estimated marginal effects for this regression. The results confirm that, for self-employed borrowers, selection of a low-doc loan is not significantly related to the probability of extreme income exaggeration. However, W-2 borrowers originating low-doc loans are significantly more likely to have extreme income overstatement. We also confirm that the results remain unchanged if we use the top quartile of *INC_EXAG* as our cutoff for *INC_EXTREME*. Finally, Table IA.X in the Internet Appendix reports the estimated coefficients assuming a linear probability model of *INC_EXTREME*. Again, the results confirm that income exaggeration increases among low-doc loans originated by W-2 borrowers.

hypothesis that borrowers with the least ex ante concerns about future credit availability (W-2 borrowers originating low-doc loans) are likely to inflate income. Focusing on the low-doc loan type and comparing income exaggeration across self-employed and W-2 borrowers, the interaction term shows that W-2 borrowers are associated with a significantly higher level of income exaggeration than self-employed borrowers. To put our income exaggeration measure into perspective, using a sample of loans from a different lender, Jiang, Nelson, and Vytlacil (2014) estimate income overstatement of 20% to 25% on low-doc loans to W-2 borrowers. Although the magnitudes differ somewhat across the two studies, both estimates suggest that low-doc loans to W-2 borrowers are, in fact, "liars' loans."

To estimate the impact of income exaggeration on ex post mortgage default, we estimate the following regression:

$$\begin{aligned}\Pr(DEFAULT_i) = \Phi(&\alpha + \beta_1 W2_i + \beta_2 Lowdoc_i + \beta_3 W2_i \times Lowdoc_i \\ &+ \lambda_1 INC\_EXAG_i + \lambda_2 W2_i \times INC\_EXAG_i + \lambda_3 Lowdoc_i \\ &\times INC\_EXAG_i + \lambda_4 W2_i \times Lowdoc_i \times INC\_EXAG_i \\ &+ \delta X_i + \theta R + \vartheta W + \gamma T), \end{aligned} \tag{10}$$

where $DEFAULT_i$ measures whether the loan is 60 days delinquent over the 24 months following origination, and X_i, R, W, and T are as defined in equation (8). To provide more comprehensive insight into the observed effects, we compute the average marginal effects of *Lowdoc* at different levels of income exaggeration across employment types and display the results in Panel A of Figure 3.[39] The horizontal axis in this panel runs from the 5th to the 95th percentile of *INC_EXAG*. Displaying the marginal effects across a range of income exaggeration levels reveals several interesting insights. Specifically, we see that higher income exaggeration among W-2 borrowers has a larger impact on the probability of default, while the slope of the marginal effect for self-employed borrowers is negative but not significantly different from zero. We therefore find that income falsification is positively related to default for low-doc loans with low information verification costs (W-2 borrowers), whereas the same relation does not hold for low-doc self-employed borrowers. The results thus support the hypothesis that borrowers with relatively low ex ante concerns about future credit access that self-select into low information

[39] Since *INC_EXAG* is a generated regressor, we use a bootstrapping procedure to estimate the marginal effects and standard errors. The first step of the bootstrapping procedure is to take a random sample of size N (with replacement) from the overall sample size of N. Because the subsample is drawn with replacement, the effective probability that the subsample of size N will be the same as the sample of size N is zero. Next, we follow the methodology outlined above to create estimates of *INC_EXAG*. Third, we estimate equation (10) on the sample and record the marginal effect and heteroskedasticity-robust standard error estimates. We then repeat the sampling and estimating procedure 400 times and use the average marginal effects and standard errors across the 400 replications. Table IA.XI in the Internet Appendix presents the results.

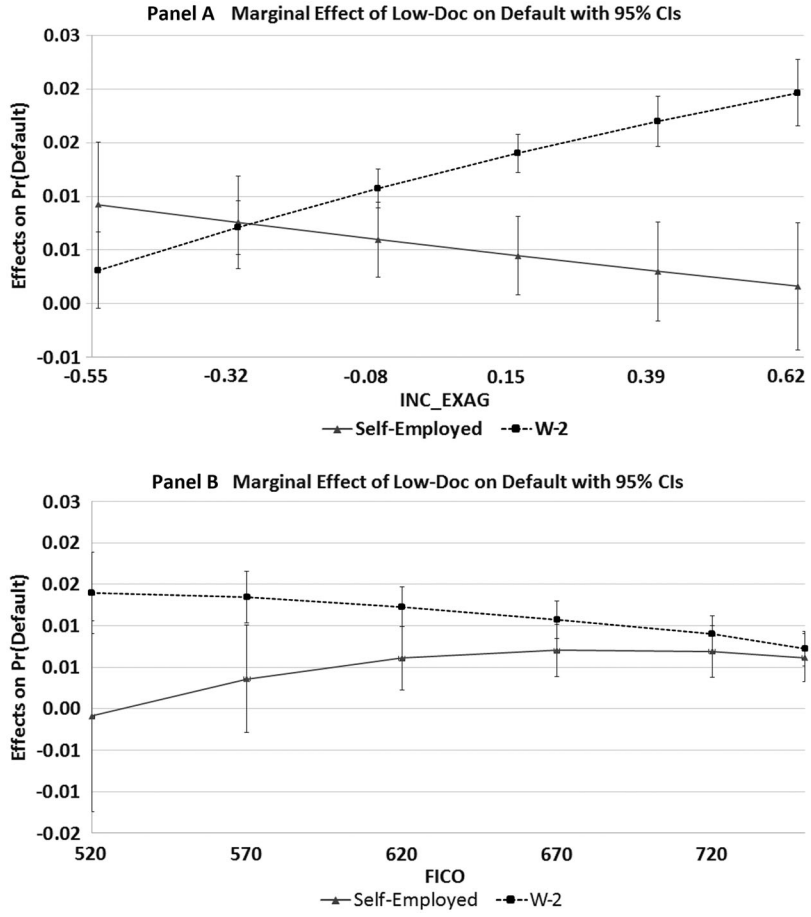


Figure 3. Marginal effect of low-doc on default by income exaggeration and FICO score. Panel A shows the average marginal effects of low-doc at different levels of estimated income exaggeration by employment type. Note that −0.53 and 0.64 are the 5^{th} and 95^{th} percentiles of income exaggeration, respectively. The marginal effects are derived from the probit model of mortgage default described in equation (10) for the funded loans from the New Century database. Panel B shows the average marginal effects of low-doc across different FICO scores by employment type. The marginal effects are derived from the probit model of mortgage default described in equation (12) for the funded loans from the New Century database.

mortgages are most likely to inflate income during loan origination and that this risk manifests itself in higher ex post default rates.[40]

[40] Even though income is not verified on low-doc loans, verification of employment and employment type (W-2, self-employed) is required, as described in Section II.A. Thus, employment-type misrepresentation is not a major concern for our study. A related but separate concern is that low-doc borrowers may receive both W-2 and self-employment income (dual-employment). While we do not observe this in our data, such imperfect measurement of W-2 status, which would be more

A. Robustness Check: Job-Specific Overstatement

To better understand the extent of income exaggeration, we create a second measure of income overstatement. For a subset of the applications in the New Century database, the lender recorded the borrower's line of business or job title (e.g, "TEACHER" and "PRESIDENT"). Using these classifications, we can compute the average income for low-doc and full-doc borrowers within each job title classification. Comparing average incomes across low-doc and full-doc loans within the same job title and employment type (W-2, self-employed) provides another measure of whether low-doc borrowers systematically inflate income, and if so, whether this varies according to employment type.[41]

Table VII presents the average incomes across documentation types for the 25 most frequently used job titles by W-2 borrowers.[42] In the first column, we see that the 1,855 W-2 borrowers with loc-doc loans whose job title is "MANAGER" have an average income of $6,720 per month. In contrast, column (3) reports that the 1,794 full-doc W-2 borrowers with a job title of "MANAGER" have an average income of $5,563 per month. Column (5) presents the mean difference test across the documentation types. In column (6), the mean difference is divided by the average income for the full-doc group to create *Job-Specific Overstatement (%)*, which can be interpreted as the percentage increase in reported income for a job type when no income documentation is provided.[43]

For every job title in Table VII, the average low-doc/W-2 income is significantly higher than the average full-doc/W-2 income. The differences are

pronounced in the low-doc sample, may potentially introduce two sources of bias. First, there could be an attenuation bias in the estimated coefficients on W2, especially the interaction term with *Lowdoc*. This is not a major concern, however, as we find statistically significant effects of these variables despite the potential bias. Second, our income exaggeration measure for W-2 low-doc borrowers may be biased. This effect is likely minor. For example, if a dual-employment borrower is less productive in the non-W-2 occupation, then the reported total income would be evaluated as a negative exaggeration because we predict income on the basis of full-time W-2 income. However, since W-2 and self-employment productivity should be highly correlated for an individual borrower, this bias is likely to be small.

[41] We restrict our analysis in this section to observations for which there is no coborrower or the coborrower's income is listed as zero. On low-doc loans with multiple borrowers and multiple job types, detecting income overstatement becomes much more difficult as exaggeration could occur within either (or both) jobs.

[42] Borrower business type is not a standardized field in the New Century data. For example, Table VII shows borrower business types of "NURSE," "REGISTERED NURSE," and "RN." Although these classifications refer to similar (or the same) positions, we did not attempt to standardize the field for several reasons. First, there are over 39,000 unique borrower business types in the data, so manually reviewing and standardizing them is cost prohibitive. Second, any attempt to standardize the field, including fuzzy matching techniques, would require significant judgment calls. Rather than letting our own biases enter into the standardization algorithm, we chose to use the field in its raw form. This is a conservative treatment as it reduces the number of observations in each category and thereby reduces the overall statistical power of our subsequent tests. As a result, our analysis is biased toward not finding evidence of income exaggeration within job titles.

[43] We find that 0.18% of the borrowers in this subsample have a job title of "OWNER" but are not coded as self-employed in the New Century data. Since we cannot determine whether the job title or the self-employment flag was coded incorrectly, we exclude these 329 observations from the analysis.

Table VII

Average Income by Documentation Status for the Most Frequently Used Borrower Business Types (W-2)

This table presents summary statistics by income documentation status for the top 25 W2 borrower business types in the subsample of loans where business type was not missing. Columns (1) and (2) ((3) and (4)) report the average income and its standard deviation for each business type for low-doc (full-doc) loans. Column (5) includes the mean difference, with ***, **, and * indicating significance at the 1%, 5%, and 10% level, respectively, from a mean difference test assuming unequal variance across the two groups. Column (6) lists *Job-Specific Overstatement* (%) that is the difference in column (5) divided by the mean income for full-doc in column (3).

	Low-Doc Income		Full-Doc Income			Job-Specific Overstatement %
Borrower Business Type	Mean (1)	Std. Dev. (2)	Mean (3)	Std. Dev. (4)	Difference (5)=(1)−(3)	(6)
MANAGER	$6,720	($2,831)	$5,563	($2,946)	$1,157***	17%
SUPERVISOR	$6,378	($2,280)	$4,879	($2,200)	$1,499***	24%
TEACHER	$6,162	($2,552)	$4,704	($1,655)	$1,458***	24%
DRIVER	$5,582	($2,165)	$4,399	($1,919)	$1,183***	21%
SALES	$6,491	($3,448)	$5,590	($3,196)	$901***	14%
TRUCK DRIVER	$5,996	($2,141)	$4,582	($2,085)	$1,414***	24%
OFFICE MANAGER	$5,830	($2,270)	$4,704	($2,707)	$1,126***	19%
CONSTRUCTION	$5,743	($2,584)	$4,653	($2,404)	$1,090***	19%
GENERAL MANAGER	$8,478	($3,793)	$7,788	($3,871)	$690**	8%
SALES MANAGER	$8,326	($3,825)	$7,787	($4,191)	$539*	6%
NURSE	$6,685	($2,724)	$5,236	($2,475)	$1,449***	22%
FOREMAN	$6,777	($2,509)	$5,226	($2,612)	$1,551***	23%
MECHANIC	$5,219	($1,918)	$4,375	($1,814)	$844***	16%
REGISTERED NURSE	$8,993	($3,387)	$6,812	($2,722)	$2,181***	24%
MACHINE OPERATOR	$5,200	($1,651)	$3,685	($1,513)	$1,515***	29%
EDUCATION	$5,540	($2,483)	$4,579	($1,915)	$961***	17%
LABORER	$4,261	($1,704)	$3,334	($1,430)	$927***	22%
MEDICAL	$5,894	($3,505)	$5,224	($3,527)	$670*	11%
ELECTRICIAN	$6,303	($2,347)	$4,949	($2,173)	$1,354***	21%
CUSTOMER SERVICE	$4,345	($1,788)	$3,377	($1,293)	$968***	22%
RN	$7,872	($2,830)	$5,908	($2,287)	$1,964***	25%
MAINTENANCE	$4,693	($1,637)	$3,425	($1,501)	$1,268***	27%
CARPENTER	$5,837	($2,359)	$4,329	($1,921)	$1,508***	26%
POLICE OFFICER	$7,581	($2,946)	$6,157	($2,210)	$1,424***	19%
ENGINEER	$9,426	($3,735)	$6,511	($2,667)	$2,915***	31%

significant in economic terms as well. For example, the average low-doc W-2 school teacher's income is $1,458 greater per month ($17,496 annually) than the average full-doc W-2 teacher's income. If we take the average full-doc income as an unbiased estimate of the average teacher's "true" income,[44] this suggests that low-doc teachers inflated their income by 24%. Within these 25

[44] We believe that this is a reasonable assumption since full-doc borrowers provided proof of income in the underwriting process.

Table VIII

Average Income by Documentation Status for the Most Frequently Used Borrower Business Types (Self-Employed)

This table presents summary statistics by income documentation status for the top 25 self-employed borrower business types in the subsample of loans where business type was not missing. Columns (1) and (2) ((3) and (4)) report the average income and its standard deviation for each business type for low-doc (full-doc) loans. Column (5) includes the mean difference, with ***, **, and * indicating significance at the 1%, 5%, and 10% level, respectively, from a mean difference test assuming unequal variance across the two groups. Column (6) lists *Job-Specific Overstatement* (%) that is the difference in column (5) divided by the mean income for full-doc in column (3).

	Low-Doc Income		Full-Doc Income			Job-Specific Overstatement %
Borrower Business Type	Mean (1)	Std. Dev. (2)	Mean (3)	Std. Dev. (4)	Difference (5)=(1)−(3)	(6)
OWNER	$8,540	($4,400)	$9,431	($5,389)	−$891***	−9%
PRESIDENT	$10,989	($5,355)	$11,133	($6,438)	−$144	−1%
REALTOR	$10,149	($5,087)	$11,408	($6,090)	−$1,259**	−11%
CONSTRUCTION	$7,829	($3,955)	$8,253	($5,278)	−$424	−5%
OWNER/OPERATOR	$7,965	($4,104)	$8,137	($4,401)	−$172	−2%
TRUCK DRIVER	$6,870	($2,844)	$6,547	($3,308)	$323	5%
REAL ESTATE AGENT	$10,809	($5,431)	$10,504	($5,255)	$305	3%
MANAGER	$7,412	($3,508)	$7,863	($4,227)	−$451	−6%
SALES	$7,387	($3,584)	$9,170	($6,812)	−$1,783**	−19%
REAL ESTATE	$9,600	($5,094)	$8,770	($5,326)	$830	9%
CONTRACTOR	$7,126	($3,109)	$7,997	($4,334)	−$871	−11%
CONSULTANT	$8,944	($4,408)	$8,059	($3,597)	$885	11%
LANDSCAPING	$6,514	($2,593)	$7,362	($4,794)	−$848	−12%
OWNER/MANAGER	$9,640	($4,144)	$10,587	($5,748)	−$947	−9%
OWNER OPERATOR	$8,272	($3,976)	$8,573	($6,232)	−$301	−4%
DRIVER	$6,369	($2,784)	$4,547	($2,143)	$1,822**	40%
LOAN OFFICER	$10,038	($4,155)	$12,521	($5,944)	−$2,483***	−20%
ATTORNEY	$12,635	($5,176)	$11,039	($5,411)	$1,596	14%
HAIR STYLIST	$5,843	($2,560)	$4,544	($1,973)	$1,299**	29%
HANDYMAN	$4,696	($1,794)	$7,916	($3,387)	−$3,220***	−41%
VICE PRESIDENT	$10,620	($4,713)	$9,356	($4,395)	$1,264	14%
CARPENTER	$6,259	($2,605)	$7,551	($5,077)	−$1,292	−17%
SALES MANAGER	$9,077	($3,562)	$10,273	($6,138)	−$1,196	−12%
PAINTER	$5,909	($2,611)	$7,300	($1,961)	−$1,391*	−19%
OWNER/PRESIDENT	$11,363	($5,386)	$13,078	($7,818)	−$1,715	−13%

most frequently used W-2 job titles, the average *Job-Specific Overstatement (%)* is 20%.

Next, we turn our attention to the 25 most frequently used job titles by self-employed borrowers. In Table VIII, the same pattern of overstatement does not emerge for self-employed borrowers. First, for many of the job titles, no significant difference exists across the low-doc and full-doc groups. In addition, whereas in Table VII, all of the mean differences are positive, for self-employed borrowers, there are both positive and negative differences, and the average

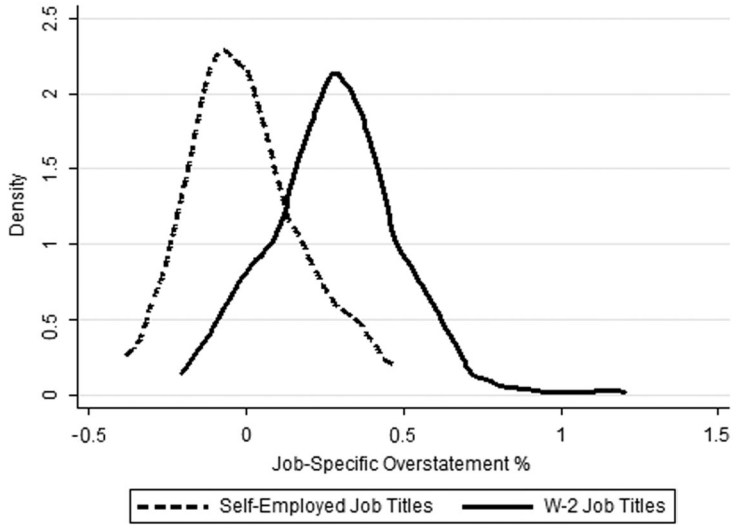


Figure 4. Distribution of job-specific overstatement. This table presents the distribution of *Job-Specific Overstatement (%)* (defined in Tables VII and VIII) for the 313 W-2 job titles that had at least 10 low-doc and 10 full-doc observations as well as the 55 self-employed job titles that had at least 10 full-doc and 10 low-doc observations.

overstatement is −3%. Consistent with our previous findings, this suggests that income exaggeration is systematic for low-doc W-2 borrowers, but not for the self-employed borrowers.

To ensure that our results are not driven by including only the 25 most frequently reported borrower business types, we broadened our sample to include any job titles that meet at least one of the two following requirements: (1) there are 10 low-doc W-2 observations and 10 full-doc W-2 observations with the job title, or (2) there are 10 low-doc self-employed observations and 10 full-doc self-employed observations with the job title. We find that 313 job titles meet the first criterion and 55 job titles meet the second. Requiring 10 loans of each documentation type limits the ability of outliers to drive our results.[45] For each of these job titles, we calculate *Job-Specific Overstatement (%)* as above. Figure 4 in the Internet Appendix presents the distribution of *Job-Specific Overstatement (%)* by employment type (W-2, self-employed). The distribution of job title overstatement for the 313 W-2 job titles is clearly shifted to the right of that for the 55 self-employed job titles, providing further evidence that income inflation is a problem on low-doc W-2 loans.

To formalize the results in Figure 4, Table IX presents the mean of each distribution. The average overstatement for W-2 jobs is 28%, versus −0.42% for self-employed job titles. In the second row of Table IX, we report the proportion

[45] Although our choice of 10 loans per documentation type within a job title is somewhat arbitrary, our results are robust to other limits (7, 12) and the use of median income rather than mean income.

Table IX

Job-Specific Overstatement Summary Statistics by Employment Type

This table presents summary statistics by employment type for *Job-Specific Overstatement* (%) as defined in Tables VII and VIII. Column (1) includes borrower business types that had at least 10 full-doc/W-2 and 10 low-doc/W-2 observations. Column (2) includes borrower business types that had at least 10 full-doc/self-employed and 10 low-doc/self-employed observations. *% of Borrower Business Types with Job-Specific Overstatement > 0* is significantly different from 50% at the 1% level of confidence for W-2 borrower business types, but is not significantly different from 50% for self-employed business types. We perform a one-tail mean difference test with the null hypothesis that the low-doc average income is less than or equal to the full-doc income for each of the borrower business types. *% of Borrower Business Types with Job-Specific Overstatement Significantly > 0* reports the percentage of borrower business types for which we were able to reject the null hypothesis.

Borrower Business Types That Have at Least 10 Full- and 10 Low-Doc Observations	W-2 (1)	Self-Employed (2)
Average Job-Specific Overstatement	28.10%	−0.42%
% of Borrower Business Types with Job-Specific Overstatement > 0	90.10%	41.82%
% of Borrower Business Types with Job-Specific Overstatement Significantly > 0	72.52%	16.36%
N	313	55

of job titles with overstatement above zero. If borrowers report true income on low-doc loans, then we would expect this number to be 50%. For W-2 job titles, the number is 90%, and thus, using a two-tailed t-test, the null hypothesis is rejected at the 1% level of confidence. Turning to self-employed job titles, where 39% have overstatement above zero, we fail to reject the null of 50%.

Next, we calculate a mean difference test of average incomes across documentation types (as in column (5) of Tables VII and VIII) for each borrower business type. For each job title, we test the null hypothesis of $H_0 : \theta \leq 0$ against $H_a : \theta > 0$. The third row of Table IX reports the fraction of the mean differences for which the null hypothesis of $H_0 : \theta \leq 0$ is rejected at the 10% level of confidence or better. For the 313 W-2 job titles (of which 90% had higher average income on low-doc loans), the null hypothesis is rejected 73% of the time. In comparison, the null hypothesis is rejected for only 16% of the job titles for self-employed borrowers. The results in Table IX provide strong evidence of income inflation on low-doc loans within the W-2 employment type, but again, we find no evidence of income exaggeration by the self-employed.[46]

We also investigate which jobs tend to have the largest income inflation. Table X reports the top 25 job titles by employment type in terms of *Job-Specific*

[46] One concern is that our analysis of job title incomes does not control for differences across location. To alleviate concerns about geographic differences in incomes, Table IA.XII in the Internet Appendix repeats the analysis from Table IX for borrowers located in California and Florida. Although the sample sizes are much smaller, the results are consistent with those in Table IX and lead us to conclude that geographic differences in incomes are not biasing the analysis.

Table X

Borrower Business Types with the Largest Job-Specific Overstatement (%)

This table includes the 25 borrower business types with the largest values for *Job-Specific Overstatement* (%), as defined in Tables VII and VIII, by employment type. The first set of columns include borrower business types that had at least 10 full-doc/W-2 and 10 low-doc/W-2 observations. The second set of columns includes borrower business types that had at least 10 full-doc/self-employed and 10 low-doc/self-employed observations.

	W-2		Self-Employed	
Ranking	Business Type (1)	Job-Specific Overstatement (2)	Business Type (3)	Job-Specific Overstatement (4)
1	PERSONAL BANKER	120%	CLEANING	47%
2	PHLEBOTOMIST	110%	DRIVER	40%
3	FORKLIFT OPERATOR	97%	HAIR SALON	36%
4	KITCHEN MANAGER	85%	DAY CARE	30%
5	ASST MGR	82%	HAIR STYLIST	29%
6	DENTAL ASSISTANT	77%	INSTALLER	28%
7	LETTER CARRIER	71%	DENTIST	24%
8	NURSING	70%	LAW	23%
9	BUS OPERATOR	69%	CHILD CARE	18%
10	MEDICAL BILLER	69%	ATTORNEY	14%
11	PROGRAM MANAGER	66%	CONSULTING	14%
12	SENIOR ACCOUNTANT	64%	VICE PRESIDENT	14%
13	DEALER	62%	CHILD CARE PROVIDER	12%
14	FINISHER	62%	CONSULTANT	11%
15	MEDICAL ASSISTANT	62%	REAL ESTATE	9%
16	CAN	61%	CEO	8%
17	CSR	61%	SUB CONTRACTOR	7%
18	CARRIER	61%	OPERATOR	6%
19	OFFICER	60%	TRUCK DRIVER	5%
20	PLUMBING	60%	REAL ESTATE SALES	5%
21	FORMAN	59%	REAL ESTATE AGENT	3%
22	COURIER	59%	INSUARNCE	1%
23	CONSTRUCTION SUPERVISOR	58%	TRUCKING	0%
24	WELDER	58%	OWNER	−1%
25	BUYER	58%	PRESIDENT	−1%

Overstatement (%). For the top-ranking W-2 job title (PERSONAL BANKER), the average annual income for low-doc borrowers ($84,672) is more than double that for full-doc borrowers with the same job title ($38,412). The average low-doc W-2 letter carrier reports an annual income of $103,128, compared with his full-doc W-2 counterpart of $60,252. The table also shows that the largest job title overstatement for self-employed borrowers (CLEANING) has overstatement below the 25th highest job title (WELDER) for W-2 borrowers

(58% versus 47%), again providing evidence that overstatement is particularly problematic among the low-doc W-2 job titles.

To summarize, in this section, we create a second measure of income overstatement (*Job-Specific Overstatement (%)*) based on borrower business type. This variable is unique to the New Century data set, and allows us to test for differences in average income within a specific job title. Our results show that income overstatement is systematic on low-doc loans within W-2 job titles, whereas we find no evidence of the same phenomenon among self-employed job titles, consistent with our earlier results on income exaggeration. To our knowledge, our study is the first to exploit differences in income across documentation types within job titles. Similar to the estimates of 20% to 25% in Jiang, Nelson, and Vytlacil (2014), our results suggest that, on average, low-doc W-2 borrowers inflate income by 28%. For low-doc self-employed borrowers, in contrast, the average inflation is 0%.

B. Robustness Check: Income within Jobs

In Section IV.A, we show that, within the same job title, average income for low-doc W-2 borrowers is significantly higher than for full-doc borrowers, but the same relationship does not hold for self-employed borrowers. However, there are several potential concerns with that analysis. First, we limit our sample to jobs that have at least 10 full-doc and 10 low-doc observations within an employment type (W-2, self-employed). Second, the averages reported may simply be picking up systematic differences in salaries across geographic locations. For example, if most low-doc loans to W-2 teachers occur in areas with relatively high teacher salaries, while the majority of full-doc loans to W-2 teachers occur in regions where teachers' salaries are low, then we would be incorrectly attributing differences to income falsification when the causal mechanism is actually benign. Third, we do not control for individual borrower characteristics that may be correlated with income. To address these concerns, we estimate the following loan-level regression on the subsample of loans for which borrower business type is not missing:

$$\ln(INCOME_{ikj}) = \alpha + \beta_1 W2_i + \beta_2 Lowdoc_i + \beta_3 W2_i \times Lowdoc_i + \delta X_i + \vartheta W_k + \gamma T + \left(\sum_{j=1}^{J} \delta_j JOB\_TITLE_j + \sum_{j=1}^{J} W2_i \times \delta_j JOB\_TITLE_j \right) + \varepsilon_{ikj}, \quad (11)$$

where JOB_TITLE_j is the borrower's business type as listed in the New Century database,[47] and the other variables are as defined above. The first term

[47] Due to the large number of fixed effects, we only include observations for which the job title has three or more observations. These observations can come from any of the employment-type/documentation-type combinations (e.g., low-doc/W-2, full-doc/W-2, low-doc/self-employed, and full-doc/self-employed). The subsample includes 2,934 unique job titles, of which 448 job titles are

Table XI
Explaining Income *within* Job Title

This table presents coefficient estimates from an OLS regression of log borrower income controlling for the borrower's job title for the subsample of New Century loans for which borrower business type was not missing. Standard errors are reported in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% level, respectively.

Dependent Variable: Log Borrower Monthly Income	(1)	(2)	(3)	(4)
Employment Type				
W2	−0.5349***	−0.2796**	−0.2311**	−0.0978
	(0.0080)	(0.1193)	(0.1103)	(0.0878)
Income Documentation				
Lowdoc	−0.0428***	−0.0162**	−0.0343***	−0.0170***
	(0.0082)	(0.0075)	(0.0070)	(0.0056)
W2 × Lowdoc	0.2815***	0.2721***	0.2275***	0.1494***
	(0.0093)	(0.0086)	(0.0079)	(0.0063)
Job Title Fixed Effects	No	Yes	Yes	Yes
W2 × Job Title Fixed Effects	No	Yes	Yes	Yes
MSA Fixed Effects	No	No	Yes	Yes
Origination Year Fixed Effects	No	No	Yes	Yes
Borrower Characteristics	No	No	No	Yes
Area Characteristics	No	No	No	Yes
Constant	8.9432***	8.6742***	8.6440***	−0.5838***
	(0.0075)	(0.1187)	(0.1136)	(0.1247)
N	86,717	86,717	86,717	86,717
Adj. R^2	0.15	0.34	0.44	0.65

in parentheses allows us to compare within-job-specific income differences between low-doc and full-doc loans, while the second term controls for the possibility that W-2 and self-employed borrowers in the same position earn different incomes.

Column (1) of Table XI serves as a baseline regression of equation (11) where we include no additional control variables. The estimated coefficients on *Lowdoc* and *W2* × *Lowdoc* are consistent with the average overstatement in Table IX.[48] In column (2), we introduce job title fixed effects and the interaction of job title fixed effects with the W-2 indicator (the omitted job title category is "TEACHER"). The coefficients on the employment type and income documentation variables represent income differences within a specific job title. The income difference becomes somewhat smaller in magnitude, indicating that income documentation and job title are correlated. In column (3), we additionally include MSA fixed effects and origination year fixed effects to control for

held by both W-2 and self-employed borrowers, 468 are held by only self-employed borrowers, and 2,018 job titles are held only by W-2 borrowers.

[48] The average monthly incomes for full-doc self-employed and W-2 observations in this subsample are $8,363 and $5,122, respectively. Note that these averages are higher than in the full sample used in Table II. As noted above, the subsample in this section includes only observations for which borrower business type is not missing and there is no coborrower income.

geographic income variation and nationwide changes in economic conditions, respectively. The results are qualitatively similar to those in column (2).

In column (4), we further control for borrower and area characteristics. More specifically, we include the natural logarithm of FICO score, an indicator for female, the natural logarithm of age, an indicator for minority status, the natural logarithm of the ZIP code per capita income reported annually, an indicator for investment property, and MSA-level house price growth over the previous two years.[49] Although the signs and significance of the coefficients are similar to those in column (3), the magnitude of the income difference is significantly smaller. For example, the coefficient on *W2* is not statistically significant, indicating that income is now comparable between W-2 and self-employed full-doc borrowers. The coefficient on *W2* × *Lowdoc* is 0.149, which is smaller than that in column (3), but is statistically significant at the 1% level. The low-doc W-2 borrowers appear to overreport income by approximately 13.24% relative to the full-doc W-2 mean income.[50] In contrast, the income reported by self-employed borrowers for low-doc loans is slightly lower (by 1.70%) than the full-doc self-employed income. Thus, consistent with all of our previous findings, income falsification is only problematic on low-doc loans to W-2 borrowers after controlling for job title and other relevant factors.

To summarize, several important facts emerge from the results in Table XI. First, job titles are important in explaining income. Although this may not be surprising, to our knowledge, this is the first study to control for the borrower's job type when examining income on low-doc loans. Second, even after controlling for job titles, area characteristics, and borrower characteristics, the results are consistent with our previous findings: income overstatement appears to be problematic only on low-doc loans to W-2 borrowers. However, when we control for borrower and area characteristics as well as the borrower's job title, the average amount of income exaggeration by low-doc W-2 borrowers is reduced to 13%, a smaller number than the 20% to 25% reported in Jiang, Nelson, and Vytlacil (2014) or than the 28% that we find in Section IV.A.

V. Lender Attempts to Control Falsification

The fourth and fifth predictions from our theoretical model imply that lenders should react to potential borrower income falsification by charging higher interest rate premiums on low-doc loans and to borrowers with low information verification costs that seek out low-doc loans. In this section, we test these predictions using a unique feature of the New Century data that allows us to examine the loan applications as well as loans that were actually originated. By using loan applications, we make a novel contribution to the literature in that

[49] The IRS income data are not available for 1999, 2000, and 2003. Loans originated in those years are matched to IRS data from the most recent prior year for which data are available. The results remain qualitatively unchanged if we exclude loans originated in 1999, 2000, and 2003.

[50] We report $\beta_2 + \beta_3 + 1/2 \times$ standard errors of $(\beta_2 + \beta_3)$.

we are able to examine the impact of potential borrower income falsification on the underwriting decision.

A. Loan Application Rejection

Lenders make decisions on loan applications along two important dimensions: pricing and application acceptance. Because most mortgage databases contain information only on funded loans, previous studies on low-doc loans focus on the former. Since the New Century data include information on funded and nonfunded mortgage applications, we are able to help fill this gap in the literature. We ask several questions regarding the lender's accept/reject decision. First, since agents (borrowers or brokers) likely inflate income with the goal of increasing the probability of application acceptance, are low-doc loans less likely to be declined by the lender? Second, does the lender reject low-doc loans differently across employment types? If the risk of default on low-doc loans varies with employment type, the lender may base its rejection decision on this information. Finally, is income exaggeration taken into account in the lender's rejection decision? To examine these questions, we expand our sample to include 698,019 funded and nonfunded applications. The percentage of loans that are funded, approved but not funded, and rejected are 67%, 19%, and 14%, respectively.[51]

To investigate whether the lender's rejection decision varies with documentation type, we first estimate a probit regression similar to equation (8) with the dependent variable taking a value of one if the loan application is rejected. Whereas, in the default regressions, we included postorigination variables to control for changing market conditions (e.g., house price changes), the explanatory variables in this regression only include information available to the lender at the time of the accept/reject decision. As in Section III, we follow Williams (2012) and report the marginal effects of low-doc at representative values for borrower employment type (i.e., at values of 0 and 1 for *W2*). Table XII presents the results. For self-employed borrowers, low-doc borrowers are associated with a 1% reduction in the probability of being rejected, or a 4.5% reduction relative to the mean for full-doc self-employed borrowers. However, the relationship reverses for W-2 borrowers. The probability of application rejection is 1% higher on low-doc loans to W-2 borrowers, or a 6.6% increase relative to the mean rejection rate for full-doc W-2 borrowers. Thus, documentation type clearly affects the application rejection decision. Moreover, the results suggest that the lender recognizes that the propensity for income falsification is greater on W-2 low-doc loans.

To test whether the lender incorporates income falsification into the rejection decision, we estimate a linear probability regression similar to equation

[51] Due to missing variables, the sample size for the regression in this section is 697,020 observations. Also, our sample of funded loans is larger in this section than the sample used in Section III, since in the default regressions, we drop observations that are missing postorigination information, while we impose no such requirement in this section.

Table XII
Relationship between Low-Doc, Employment Type, and Application Rejection

This table presents marginal effects of no-doc on application rejection by employment type. The marginal effects are derived from a probit model of application rejection on income documentation, employment type, the interaction between income documentation and employment type, loan characteristics, borrower characteristics, property characteristics, and area characteristics at the time of the accept/reject decision for the loan applications from the New Century database. Heteroskedasticity-robust standard errors adjusted for clustering at the MSA level are reported in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively.

Dependent Variable: Application Rejected	
Self-Employed (Low-Doc)	−0.0083***
	(0.0028)
W-2 (Low-Doc)	0.0093***
	(0.0018)
Loan Characteristics	Yes
Property Characteristics	Yes
Borrower Characteristics	Yes
Interest Rate Environment	Yes
Area Characteristics	Yes
Origination Year Fixed Effects	Yes
MSA Fixed Effects	Yes
N	697,009
Mean Rejection Rate (Full-doc Self-Employed)	0.1840
Mean Rejection Rate (Full-doc W-2)	0.1407
Log Likelihood	−254,462

(10) where we include our measure of income exaggeration. We now use the rejection indicator as the dependent variable and include all of the independent variables from equation (10) that are observable to the lender at the time of the rejection decision. Table XIII reports the marginal coefficient estimates for the OLS estimation.[52] The results indicate that W-2 borrowers are 2.4% less likely to be rejected than self-employed borrowers. Similarly, borrowers originating low-doc loans are 1.06% less likely to be rejected than borrowers seeking full-doc loans. However, interaction of *W2* and *Lowdoc* confirms that borrowers with low costs of verifying information faced significantly higher lender scrutiny, as the probability of rejection is 1.77% higher than for self-employed low-doc borrowers. Finally, the positive and significant coefficient for the triple interaction between *Lowdoc*, *W2*, and *INC_EXAG* indicates that the lender recognized the risk of income falsification on low-doc loans to W-2 borrowers and adjusted the probability of rejection accordingly.

[52] Since *INC_EXAG* is a generated regressor, we use a bootstrapping procedure similar to that outlined for equation (10). However, the bootstrapping procedure causes problems when using a probit model on all loan applications (versus funded loans in earlier sections), so we report bootstrap estimates from an OLS model in Table XIII. Column (1) reports coefficient estimates when we do not include the generated regressor in the model.

Table XIII

Relationship between Low-doc, Employment Type, and Application Rejection (OLS)

This table presents marginal coefficient estimates from a linear probability model of application rejection on income documentation, employment type, the interaction term between income documentation and employment type, loan characteristics, borrower characteristics, property characteristics, and area characteristics at the time of the accept/reject decision for the loan applications from the New Century database. Heteroskedasticity-robust standard errors adjusted for clustering at the MSA level are reported in parentheses in column (1). Bootstrapped standard errors are reported in parentheses in column (2). ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively.

Dependent Variable: Application Rejected	(1)	(2)
W2	−0.0209***	−0.0243***
	(0.0027)	(0.0024)
Lowdoc	−0.0096***	−0.0106***
	(0.0030)	(0.0022)
Lowdoc × W2	0.0176***	0.0177***
	(0.0027)	(0.0024)
INC_EXAG		0.0028
		(0.0046)
INC_EXAG × W2		−0.0089*
		(0.0049)
Lowdoc × INC_EXAG		0.0009
		(0.0054)
Lowdoc × W2 × INC_EXAG		0.0162***
		(0.0061)
Loan Characteristics	Yes	Yes
Property Characteristics	Yes	Yes
Borrower Characteristics	Yes	Yes
Interest Rate Environment	Yes	Yes
Area Characteristics	Yes	Yes
Origination Year	Yes	Yes
MSA Fixed Effects	Yes	Yes
N	697,020	687,199
Adj. R^2	0.11	0.10

Taken together, the results in Tables XII and XIII provide several new insights on low-doc loans. First, low-doc loans are treated differently from full-doc loans with regard to loan approval. Second, the relationship varies according to employment type: low-doc loans are associated with a lower likelihood of rejection for self-employed borrowers, but for W-2 borrowers, low-doc loans are more likely to be declined. Third, the lender appears to incorporate income exaggeration into the rejection decision for low-doc W-2 borrowers, but not for self-employed borrowers. This finding is consistent with our previous results: that income falsification is problematic only on low-doc loans to W-2 borrowers.

B. Low-Doc Loans and Credit Reputation

Next, we examine the interaction of low-doc, employment type, and credit history. We measure credit history observable at origination using the borrower's credit (FICO) score, a standard risk metric used in mortgage underwriting in the United States. An individual develops a reputation with creditors over time through credit usage and debt repayment patterns. The FICO score quantifies this reputation, with higher scores reflecting more creditworthy borrowers, ceteris paribus. Since credit scores are widely used for lending, insurance, and employment decisions, a strong credit reputation, as indicated by a high FICO score, is a valuable asset for a borrower.

In this section, we test whether observed credit reputation mitigates the default risk of borrowers that have otherwise signaled low concerns about access to future credit (W-2 borrowers selecting low-doc loans). Our regression now takes the form

$$\begin{aligned}\Pr(DEFAULT_i) = \Phi(&\alpha + \beta_1 W2_i + \beta_2 Lowdoc_i + \beta_3 W2_i \times Lowdoc_i \\ &+ \lambda_1 FICO_i + \lambda_2 W2_i \times FICO_i + \lambda_3 Lowdoc_i \times FICO_i \\ &+ \lambda_4 W2_i \times Lowdoc_i \times FICO_i + \delta X_i + \theta R + \vartheta W + \gamma T), \quad (12)\end{aligned}$$

where $FICO_i$ is the borrower's credit score at origination, and all other variables are as defined in equation (8). The three-way interaction of $W2$ with $Lowdoc$ and $FICO$ allows us to test whether an established credit reputation ameliorates the additional default risk of low-doc loans.

Panel B of Figure 3 graphs the average marginal effects of low-doc by employment type across FICO scores.[53] For low-verification-cost borrowers (W-2), the downward-sloping line provides some evidence that credit reputation counteracts the income exaggeration problem inherent in low-doc loans. That is, borrowers with higher FICO scores have lower default probabilities. However, the same result does not hold for self-employed borrowers. Interestingly, we note that the average marginal effect of *Lowdoc* increases over the lower range of FICO scores for self-employed borrowers. Given the wide confidence intervals, we are careful not to interpret the results in this section too strongly. However, Panel B of Figure 3 does suggest that the increased risk associated with low-doc loans is most severe for borrowers that are least likely to be concerned about future credit rationing, namely, W-2 borrowers with low FICO scores.

C. Income Falsification and Mortgage Pricing

The previous sections demonstrate that low-doc loans to borrowers with low concerns about future credit rationing are riskier due to income inflation. In this section, we examine whether the lender priced this risk. To test this hypothesis,

[53] Table IA.XIV in the Internet Appendix reports the marginal effects of low-doc at different levels of FICO score by employment type in tabular form.

Table XIV
Relationship between Low-Doc, Employment Type, and Mortgage Pricing

This table presents coefficient estimates from an OLS regression of mortgage pricing on type of income documentation, loan characteristics, borrower characteristics, property characteristics, and area characteristics for the funded loans from the New Century database. The average *RATE_SPREAD* for the reference group (full-doc/self-employed) is 4.5672 holding all other control variables at their average values. Heteroskedasticity-robust standard errors adjusted for clustering at the MSA level are reported in parentheses in column (1). Bootstrapped standard errors are reported in parentheses in column (2). ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively.

Dependent Variable: Rate Spread	(1)	(2)
W2	−0.0981***	−0.0717***
	(0.0083)	(0.0077)
Lowdoc	0.5305***	0.5308***
	(0.0173)	(0.0073)
Lowdoc × W2	0.1508***	0.1509***
	(0.0113)	(0.0079)
INC_EXAG		0.0852***
		(0.0172)
INC_EXAG × W2		0.0614***
		(0.0149)
Lowdoc × INC_EXAG		0.0235
		(0.0167)
Lowdoc × W2 × INC_EXAG		0.0291
		(0.0201)
Loan Characteristics	Yes	Yes
Property Characteristics	Yes	Yes
Borrower Characteristics	Yes	Yes
Interest Rate Environment	Yes	Yes
Area Characteristics	Yes	Yes
Origination Year	Yes	Yes
MSA Fixed Effects	Yes	Yes
N	455,687	449,917
Adj. R^2	0.69	0.69

we estimate the following OLS model of pricing:

$$\begin{aligned} RATE\_SPREAD_{ij} = {} & \alpha + \beta_1 W2_i + \beta_2 Lowdoc_i + \beta_3\{W2_i \times Lowdoc_i\} \\ & + \lambda_1 INC\_EXAG_i + \lambda_2\{W2_i \times INC\_EXAG_i\} \\ & + \lambda_3\{Lowdoc_i \times INC\_EXAG_i\} \\ & + \lambda_4\{W2_i \times Lowdoc_i \times INC\_EXAG_i\} \\ & + \delta X_i + \theta R + \vartheta W_j + \gamma T + \varepsilon_{ij}, \end{aligned} \tag{13}$$

where *RATE_SPREAD* is the note rate on the mortgage minus the two-year T-bill rate in the month of origination, with the control variables as defined in Section II. Column (1) of Table XIV reports the coefficient estimates from

the pricing regression using the full sample. Relative to full-doc self-employed borrowers, interest rate spreads on loans to W-2 borrowers are 9.8 basis points lower. The second and third rows of column (1) suggest that the lender recognized differences in low-doc loan quality according to borrower employment status: low-doc loans to borrowers with verifiable income (W2) carried an additional risk premium of 15 basis points relative to low-doc loans to self-employed borrowers. Interestingly, although most of the additional risk on low-doc loans is attributable to W-2 borrowers, the majority of the low-doc premium (53 basis points) applies to all borrower types.

The second column of Table XIV includes *INC_EXAG*.[54] The coefficients on *Lowdoc* and *W2* × *Lowdoc* are nearly identical to those in column (1). The coefficient on *INC_EXAG* suggests that full-doc self-employed borrowers with high income levels (relative to our model estimates) pay a rate premium. The same result holds for full-doc W-2 borrowers with high levels of income. Since *INC_EXAG* for a full-doc borrower does not contain income falsification, this rate premium corresponds to higher risk in a mortgage originated to a high-income individual, possibly due to higher risk in income or collateral value. This rate premium on *INC_EXAG* is not significantly different for low-doc loans to self-employed or W-2 borrowers.

The results in column (2) clearly show that the lender prices additional risk associated with low-doc loans to W-2 borrowers, but we find no evidence that the pricing is related to income exaggeration at the loan level. However, it is important to recognize that loan pricing is the result of two processes: (1) the lender's risk-based pricing and (2) negotiations between the borrower and the originator. Although we cannot fully disentangle each of these effects, Figure 1 provides some insight. Clearly, there is a risk-based premium moving from the "FULL DOC" to the "STATED DOC" side of the pricing sheet. In addition, the "Adjustments To Rate" section shows an additional rate premium of 30 basis points if the loan is for a "Stated Wage Earner." This indicates that the lender increased the low-doc risk premium for borrowers likely to have lower concerns about future credit rationing (W-2), consistent with our empirical results. Not surprisingly, the rate sheet does not contain any pricing adjustments for "income inflation," "income exaggeration," "unbelievable income," or any other variant of those phrases, since the lender would not have wanted to publicize that it had officially accepted falsified applications and that it had charged a higher rate on the basis of its imperfect assumption of income falsification. Our empirical results, combined with the New Century Rate Sheet, suggest that the lender did price low-doc loan risk explicitly, but we find no evidence that income exaggeration was priced at the loan level.[55] As we predict, the rate

[54] Since *INC_EXAG* is a generated regressor, we use a bootstrapping procedure similar to that outlined for equation (10).

[55] We are careful not to generalize from the rate sheet to our entire sample period, since rate sheets were region-specific and changed frequently. However, we note that the First Franklin rate sheet (Figure IA.1) contained a similar premium for "NIV Wage Earner," indicating that New Century was not alone in pricing low-doc loans to W-2 borrowers.

differentials did not completely eliminate the problems of adverse selection and income falsification in the mortgage market.

VI. Policy Implications: Income Falsification, Borrower Location, and Subsequent House Price Declines

As we note in the introduction, the role of borrower income misrepresentation in facilitating the expansion of mortgage credit is subject to debate. The extent to which borrowers (or lenders/brokers operating on behalf of borrowers) systematically inflated incomes in order to obtain larger loans is consistent with the view that the 2002 to 2006 housing boom resulted from an expansion in mortgage credit due to a decline in underwriting standards. Further, in support of this view, Mian and Sufi (2015) examine ZIP-code-level differences in income growth reported on mortgage applications and growth in IRS reported income and find that areas that experienced significant growth in subprime mortgage origination activity also saw higher levels of income overstatement.[56] Using microlevel mortgage data compiled by Piskorski, Seru, and Witkin (2015), Mian and Sufi (2015) also document that incidents of mortgage fraud were significantly more likely in areas identified as having higher levels of borrower income misrepresentation.

We contribute to our understanding of the linkage between income overstatement and mortgage fraud by analyzing income falsification by borrower employment status at the ZIP code level using a research design similar to that in Mian and Sufi (2015). Specifically, we regress the percentage of low-doc loans in each ZIP code on the natural logarithm of ZIP code median income from the 2000 Census.[57] Table XV reports the estimated coefficients, where columns (1) and (2) correspond to the W-2 borrower sample and columns (3) and (4) to the self-employed sample. Focusing first on the W-2 borrowers, when looking across MSAs (column (1) without MSA fixed effects), we see that higher income areas are correlated with a higher proportion of loans to low-doc W-2 borrowers. However, when looking within MSAs (column (2) with MSA fixed effects), the sign on the estimated coefficient becomes negative, suggesting that loans to low-doc W-2 borrowers concentrate in lower income ZIP codes. Taken together, the results in columns (1) and (2) suggest that low-doc loans to W-2 borrowers are more prevalent in wealthier (higher income) MSAs, but the origination activity is occurring in the lower income areas of those MSAs. In contrast, for the self-employed borrowers (columns (3) and (4)), the negative relation between low-doc loans and lower income ZIP codes holds regardless of whether we look across or within MSAs. Our results therefore support the

[56] In addition, Agarwal, Ambrose, and Yildirim (2015) provide evidence that the risks associated with subprime mortgage originations were transmitted to the prime mortgage market.

[57] We estimate the regression for the W-2 borrowers and self-employed borrowers separately. For the W-2 borrower group, we select ZIP codes that have at least nine loans to W-2 borrowers (the median number of W-2 borrowers across all ZIP codes). For the self-employed sample, we select ZIP codes that have at least four loans to self-employed borrowers (the median number of self-employed borrowers across all ZIP codes.)

Table XV
Low-Doc Penetration and ZIP Code Median Income

This table presents coefficient estimates from an OLS regression of the percentage of a ZIP code's loans that are low-doc on the natural logarithm of ZIP code median income from the 2000 Census. The dependent variable in columns (1) and (2) is the number of loans in a ZIP code that were low-doc to W-2 borrowers divided by the total number of W-2 loans originated in that ZIP code. The dependent variable in columns (3) and (4) is the number of loans in a ZIP code that were low-doc to self-employed borrowers divided by the total number of loans originated to self-employed borrowers in that ZIP code. ZIP codes included in columns (1) and (2) must have at least nine loans to W-2 borrowers (the median for all the ZIP codes), whereas columns (3) and (4) include ZIP codes that have at least four originations to self-employed borrowers (the median for all ZIP codes). The total number of ZIP codes in our sample is 14,090. Standard errors are reported in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively.

Dependent Variable: Low-Doc Share in ZIP	% Lowdoc W2 (1)	% Lowdoc W2 (2)	% Lowdoc Self-Employed (3)	% Lowdoc Self-Employed (4)
ZIP Median Income	0.0328***	−0.0202***	−0.0233***	−0.0505***
	(0.0046)	(0.0042)	(0.0060)	(0.0065)
MSA Fixed Effects	No	Yes	No	Yes
N	7,063	7,033	5,464	5,464
Adjusted R^2	0.01	0.36	0.00	0.13

finding of Mian and Sufi (2015) that mortgages to borrowers most likely to overstate income (W-2 borrowers originating low-doc loans) are concentrated in lower income neighborhoods.

Based on the evidence linking buyer income overstatement to specific areas, Mian and Sufi (2015) argue that this expansion in the supply of mortgage credit, including low-doc loans, put upward pressure on house prices.[58] However, this interpretation is controversial. Adelino, Schoar, and Severino (2015a) point out that the income distribution of mortgage purchase applicants may be different from the ZIP code income distribution reported in the IRS data. Thus, rather than reflecting income overstatement on low-doc mortgage applications, Mian and Sufi's (2015) measure may simply be reflecting the fact that home buyers have higher average incomes than the average income of all individuals within a ZIP code. The link between low-doc loans and house prices thus remains an empirical question.

We add to this debate by examining the relationship between low-doc market share by employment type and subsequent house price growth. We ask whether greater exposure to low-doc loans, especially low-doc loans to W-2 borrowers, is negatively related to house price growth rates after the housing boom. For each MSA i from 2004:Q1 to 2005:Q4, we measure the share of low-doc loan originations ($L_i \equiv \frac{Lowdoc_i}{All_i}$, where $Lowdoc_i$ and All_i denote the

[58] Pavlov and Wachter (2011), Agarwal et al. (2012), and Berkovec, Chang, and McManus (2012) also provide empirical and theoretical support for the credit supply effect of subprime lending on housing prices.

number of originated low-doc and all mortgages, respectively) and the proportion of low-doc W-2 borrowers in the low-doc loan originations ($W_i \equiv \frac{W2\&Lowdoc_i}{Lowdoc_i}$, where $W2\&Lowdoc_i$ denotes the number of W-2 low-doc loan originations). We compute the subsequent house price change starting from 2006:Q1, which corresponds to the time when a small number of MSAs started to exhibit price declines. We use three different periods of cumulative house price changes, 2006 to 2007, 2006 to 2008, and 2006 to 2009, and estimate the following MSA-level equation:

$$\Delta HPI_i = \alpha + \beta_1 (L_i) + \beta_2 (W_i) + \beta_3 \left(E_i^{-1}\right) + \beta_4 \left(L_i \times E_i^{-1}\right) + \beta_5 \left(W_i \times E_i^{-1}\right) + \beta_6 (L_i \times W_i) + \beta_7 \left(L_i \times W_i \times E_i^{-1}\right) + \gamma M_i + \delta \left(M_i \times E_i^{-1}\right) + \varepsilon, \quad (14)$$

where ΔHPI_i denotes the cumulative house price change since 2006 in MSA i measured by the Federal Housing Finance Agency (FHFA) MSA-level house price index, E_i^{-1} denotes the inverse elasticity of housing supply estimated by Saiz (2010), and M_i contains variables that control for changes in housing demand in MSA i, namely, house price growth between 2000 and 2005, and the change in population, per capita income, and unemployment rate since 2006. We include the interaction terms between the inverse of supply elasticity and other exogenous variables because the inverse of elasticities serves as a conditioning variable in the reduced-form equilibrium price equation. We only control for demand factors because the main cause of the housing bust was likely due to housing demand shocks. We require each MSA to have at least 23 loans to be included in the sample, with 95% of the MSAs meeting this requirement.[59]

Table XVI presents the marginal effect of the share of low-doc loans (L) and the proportion of low-doc W-2 borrowers (W), which are evaluated at the mean values of the interacted variables:

$$\left.\frac{\partial \Delta HPI_i}{\partial L_i}\right|_{\overline{W_i}, \overline{E_i^{-1}}} = \beta_1 + \beta_4 \overline{E_i^{-1}} + \beta_6 \overline{W_i} + \beta_7 \left(\overline{W_i} \times \overline{E_i^{-1}}\right), \quad (15)$$

$$\left.\frac{\partial \Delta HPI_i}{\partial W_i}\right|_{\overline{L_i}, \overline{E_i^{-1}}} = \beta_2 + \beta_5 \overline{E_i^{-1}} + \beta_6 \overline{L_i} + \beta_7 \left(\overline{L_i} \times \overline{E_i^{-1}}\right), \quad (16)$$

where the upper bar indicates the sample mean. We also evaluate how these effects vary with supply inelasticity by computing partial derivatives of equations (15) and (16) with respect to E_i^{-1}. Column (1) shows that both the share of low-doc loans and the proportion of low-doc W-2 borrowers are negatively associated with house price growth from 2006 to 2007. The estimated coefficients are statistically significant at the 1% level. The results indicate that a 10 percentage point increase in the share of low-doc loans in 2004 and 2005 is

[59] The results are qualitatively similar when we use other cutoff values for the minimum number of loans to be included in the sample.

Table XVI

MSA House Price Changes and Low-Doc Share of Originations

This table reports the marginal effects on the subsequent house price changes of the share of low-doc loans and the proportion of low-doc loans to W-2 borrowers as defined by equations (15) and (16). The marginal effect is evaluated at the mean values of the interacted variables. The regression results are presented in Table IA.XIII in the Internet Appendix. White heteroskedasticity-robust standard errors are reported in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively.

Dependent Variable	ΔHPI (2006 to 2007) (1)	ΔHPI (2006 to 2008) (2)	ΔHPI (2006 to 2009) (3)
Marginal Effect of the Share of Low-Doc Loans (equation (15))	−0.187***	−0.106*	−0.023
	(0.053)	(0.062)	(0.072)
Change in the Effect by Supply Inelasticity	−0.601***	−0.345**	0.074
	(0.144)	(0.159)	(0.206)
Marginal Effect of Proportion of W2 (equation (16))	−0.172***	−0.184***	−0.138**
	(0.056)	(0.068)	(0.070)
Change in the Effect by Supply Inelasticity	−0.355	−0.527**	−0.376
	(0.221)	(0.234)	(0.243)

associated with a 1.87% lower house price growth rate in 2006 and 2007. A 10 percentage point increase in the proportion of low-doc W-2 borrowers in 2004 and 2005 is associated with a 1.72% lower growth rate. These effects change with supply inelasticity. The change in the effect of low-doc share by supply inelasticity is −0.601 and statistically significant at the 1% level. Thus, for an MSA that exhibits a one-standard-deviation increase in the inverse elasticity of supply, a 10 percentage point increase in low-doc share is associated with a 3.75% lower house price growth rate. The change in the effect of low-doc W-2 borrowers' proportion by supply inelasticity (−0.355) is marginally significant at the 11% level. This coefficient suggests that a 10 percentage point increase in low-doc W-2 borrowers' proportion is associated with a 2.83% lower house growth rate in an MSA with a one-standard-deviation larger inelasticity value.[60] Columns (2) and (3) indicate that the relation between the low-doc share and the subsequent house price change becomes smaller and weaker as the recession grows in severity in 2008 and 2009. However, the relation between the low-doc W-2 borrowers' proportion and house price changes remains strong and statistically significant until 2009. Although we are careful not to claim a strong causal interpretation, this result suggests that exposure to low-doc loans, especially to W-2 low-doc loans, at the peak of the housing boom is closely related to the beginning of the housing bust.

[60] The estimated values of 3.75% and 2.83% are calculated as $10 \times (-0.187 - 0.601 \times 0.312)$ and $10 \times (-0.172 - 0.355 \times 0.312)$, respectively, where 0.312 is the standard deviation of inverse elasticity.

VII. Conclusion

Financial economists have only recently begun to examine the role of mortgage fraud and adverse selection as contributing factors to the Great Recession of 2007 to 2009. For example, growing evidence suggests that misrepresentations of borrower income, borrower assets, appraisals, second liens, and owner-occupancy status increased significantly during the period prior to the financial crisis. In contributing to this literature, we document how borrower heterogeneity with respect to employment status contributed to income misrepresentation and adverse selection.

Using a national data set of subprime mortgages originated by a major financial institution during the house price boom, we provide several novel insights into the role of borrower income misrepresentation. First, we document that income misrepresentation was concentrated among borrowers who originated low-doc loans but could have originated full-doc loans instead. Second, we show that the majority of additional risk associated with low-doc mortgages was due to adverse selection and income falsification on the part of borrowers with verifiable income. We also provide evidence that these borrowers were more likely to inflate or exaggerate their income on the mortgage application. As a result, we provide new evidence showing that income misrepresentation resulted from borrower actions, which extends prior results showing that excesses in the mortgage market arise from lender actions. Third, we document that lenders took actions designed to counter potential borrower income falsification. Finally, we discuss how borrower income falsification may have facilitated the expansion of mortgage credit, and thus provide new insights into one of the possible causes of the Great Recession.

Taken together, our empirical analysis suggests that borrower concerns about future credit access can mitigate the effects of adverse selection in limited-information documentation mortgage contracts. From a policy perspective, our results indicate that a blanket regulation mandating "qualified" mortgages (i.e., loans that require full documentation) may be overly restrictive and lead to credit rationing for the subset of the population that faces high information verification costs. Such regulations may have serious unintended consequences for the economy. Our analysis therefore suggests that regulators seeking to limit the potential of a future foreclosure crisis should rely on a more nuanced or targeted regulatory approach that limits the use of low-documentation loans by borrowers who have ex ante low information verification costs.

Initial submission: March 25, 2015; Accepted: November 28, 2015
Editors: Bruno Biais, Michael R. Roberts, and Kenneth J. Singleton

REFERENCES

Adelino, Manuel, Antoinette Schoar, and Felipe Severino, 2015a, Changes in buyer composition and the expansion of credit during the boom, Working Paper 20848, NBER.

Adelino, Manuel, Antoinette Schoar, and Felipe Severino, 2015b, House prices, collateral, and self-employment, *Journal of Financial Economics* 117, 288–306.

Agarwal, Sumit, Brent W. Ambrose, Souphala Chomsisengphet, and Anthony B. Sanders, 2012, Thy neighbor's mortgage: Does living in a subprime neighborhood affect ones probability of default? *Real Estate Economics* 40, 1–22.

Agarwal, Sumit, Brent W. Ambrose, and Vincent W. Yao, 2014, The limits of regulation: Appraisal bias in the mortgage market, Working paper, National University of Singapore.

Agarwal, Sumit, Brent W. Ambrose, and Yildiray Yildirim, 2015, The subprime virus, *Real Estate Economics* 43, 891–915.

Agarwal, Sumit, Gene Amromin, Itzhak Ben-David, Souphala Chomsisengphet, and Douglas D. Evanoff, 2014, Predatory lending and the subprime crisis, *Journal of Financial Economics* 113, 29–52.

Agarwal, Sumit, Itzhak Ben-David, and Vincent Yao, 2015, Collateral valuation and borrower financial constraints: Evidence from the residential real estate market, *Management Science* 61, 2220–2240.

Agarwal, Sumit, Yan Chang, and Abdullah Yavas, 2012, Adverse selection in mortgage securitization, *Journal of Financial Economics* 105, 640–660.

Agarwal, Sumit, and Douglas D. Evanoff, 2013, Loan product steering in mortgage market, Working paper, National University of Singapore.

Ai, Chunrong, and Edward C. Norton, 2003, Interaction terms in logit and probit models, *Economics Letters* 80, 123–129.

Ambrose, Brent W., and James N. Conklin, 2014, Mortgage brokers, origination fees, price transparency and competition, *Real Estate Economics* 42, 363–421.

Ben-David, Itzhak, 2011, Financial constraints and inflated home prices during the real estate boom, *American Economic Journal: Applied Economics* 3, 55–87.

Berkovec, Jim, Yan Chang, and Douglas A. McManus, 2012, Alternative lending channels and the crisis in U.S. housing markets, *Real Estate Economics* 40, S8–S31.

Buis, Maarten L., 2010, Stata tip 87: Interpretation of interactions in non-linear models, *The Stata Journal* 10, 305–308.

Burke, Jeremy M., Curtis R. Taylor, and Liad Wagman, 2012, Information acquisition in competitive markets: An application to the U.S. mortgage market, *American Economic Journal: Microeconomics* 4, 65–106.

Chomsisengphet, Souphala, and Anthony Pennington-Cross, 2006, The evolution of the subprime mortgage market, *Federal Reserve Bank of St. Louis Review* 88, 31–56.

Conklin, James N., 2016, Financial literacy, broker-borrower interaction, and mortgage default, *Real Estate Economics*.

Demyanyk, Yuliya, and Otto Van Hemert, 2011, Understanding the subprime mortgage crisis, *Review of Financial Studies* 24, 1848–1880.

Diamond, Douglas W., 1989, Reputation acquisition in debt markets, *Journal of Political Economy* 97, 828–862.

Diamond, Douglas W., 1991, Monitoring and reputation: The choice between bank loans and directly placed debt, *Journal of Political Economy* 99, 689–721.

Garmaise, Mark J., 2015, Borrower misreporting and loan performance, *Journal of Finance* 70, 449–484.

Green, Richard K., 2014, *Introduction to Mortgages and Mortgage Backed Securities* (Academic Press, Waltham, MA).

Griffin, John M., and Gonzalo Maturana, 2015, Who facilitated misreporting in securitized loans? Working paper, University of Texas at Austin.

Jiang, Wei, Ashlyn Aiko Nelson, and Edward Vytlacil, 2014, Liar's loan? Effects of origination channel and information falsification on mortgage delinquency, *Review of Economics and Statistics* 96, 1–18.

Keys, Benjamin J., Tanmoy Mukherjee, Amit Seru, and Vikrant Vig, 2009, Financial regulation and securitization: Evidence from subprime loans, *Journal of Monetary Economics* 56, 700–720.

LaCour-Little, Michael, and Jing Yang, 2013, Taking the lie out of liar loans: The effect of reduced documentation on the performance and pricing of alt-a and subprime mortgages, *Journal of Real Estate Research* 35, 507–553.

Mian, Atif, and Amir Sufi, 2009, The consequences of mortgage credit expansion: Evidence from the U.S. mortgage default crisis, *Quarterly Journal of Economics* 124, 1449–1496.

Mian, Atif, and Amir Sufi, 2015, Fraudulent income overstatement on mortgage applications during the credit expansion of 2002 to 2005, Working Paper 20947, NBER.

Pahl, Cynthia, 2007, A compilation of state mortgage broker laws and regulations 1996-2006, Community affairs report #2007-2, Federal Reserve Bank of Minneapolis.

Paley, Irina, and Konstantinos Tzioumis, 2011, Rethinking stated-income loans: Separating the wheat from the chaff, Working paper, Office of the Comptroller of the Currency, Washington, DC.

Pavlov, Andrey, and Susan Wachter, 2011, Subprime lending and real estate prices, *Real Estate Economics* 39, 1–17.

Piskorski, Tomasz, Amit Seru, and Vikrant Vig, 2010, Securitization and distressed loan renegotiation: Evidence from the subprime mortgage crisis, *Journal of Financial Economics* 97, 369–397.

Piskorski, Tomasz, Amit Seru, and James Witkin, 2015, Asset quality misrepresentation by financial intermediaries: Evidence from the RMBS market, *Journal of Finance* 70, 2635–2678.

Rajan, Uday, Amit Seru, and Vikrant Vig, 2010, Statistical default models and incentives, *American Economic Review (Papers and Proceedings)* 100, 506–210.

Saiz, Albert, 2010, The geographic determinants of housing supply, *Quarterly Journal of Economics* 125, 1253–1296.

U.S. Financial Crisis Inquiry Commission, 2011, *The Financial Crisis Inquiry Report: Final Report of the National Commission on the Causes of the Financial and Economic Crisis in the United States* (Government Printing Office, Washington, DC).

Williams, Richard, 2012, Using the margins command to estimate and interpret adjusted predictions and marginal effects, *Stata Journal* 12, 308–331.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Appendix S1: Internet Appendix.