Owner Occupancy Fraud and Mortgage Performance[*]

Ronel Elul[†]
Federal Reserve Bank of Philadelphia

Sebastian Tilson
Federal Reserve Bank of Philadelphia

March 16, 2015

Abstract

We use a matched credit bureau and mortgage dataset to identify occupancy fraud in residential mortgage originations, that is, borrowers who misrepresented their occupancy status as owner occupants rather than residential real estate investors. In contrast to previous studies, our dataset allows us to show that such fraud was broad based, appearing in the government-sponsored enterprise market and in loans held on bank portfolios as well. Mortgage borrowers who misrepresented their occupancy status performed worse than otherwise similar owner occupants and declared investors, defaulting at nearly twice the rate. In addition, these defaults are significantly more likely to be "strategic" in the sense that their bank card performance is better and utilization is lower.

Keywords: mortgages, mortgage default, consumer credit, household finance, misreporting, fraud

JEL Codes: D12, R3

---

[*] The authors thanks Tomasz Piskorski and participants at the Eastern Economic Association meetings for helpful comments

[†] Elul and Tilson are both from the Federal Reserve Bank of Philadelphia. Corresponding author: Ronel Elul, e-mail: ronel.elul@phil.frb.org.; tel: 215-574-3965.

I. Introduction

Policymakers and the popular press have cited anecdotal evidence to suggest that one of the contributing causes to the housing bubble was pervasive mortgage fraud.[1] Recent academic work has also verified the existence of mortgage fraud along several dimensions. Ben-David (2011) finds evidence of inflated prices. Griffin and Maturana (2015a) examine three dimensions of fraud among securitized nonagency loans: unreported second liens, owner occupancy misreporting, and appraisal overstatements. Piskorski, Seru, and Witkin (2015) study second lien misreporting and occupancy fraud in the nonagency securized market. Mian and Sufi (2015) argue that borrowers misstated their incomes on mortgage applications.

In this paper, we use a matched credit bureau and mortgage dataset to identify occupancy fraud in loans originated between 2005 and 2007. This occurs when mortgage borrowers claim on the mortgage application that they will be the owner occupants of the property, will not rent the property out to another individual or family, and do not intend to sell the property quickly. Borrowers may have an incentive to commit occupancy fraud because the benefits can be substantial: Banks often require declared residential mortgage investors to offer higher down payments and charge them higher interest rates because of the elevated default risk of investor loans (which we also document in this paper). In contrast to previous work, our data allow us to confirm that occupancy fraud was pervasive and did not just affect private-securitized loans. We show that more than half of all investors were fraudulent. And this applied to government-sponsored enterprise (GSE)–guaranteed, private securitized, and portfolio-held loans (by contrast, Federal Housing Administration (FHA) loans exhibited markedly lower fraud rates).

After we have identified these investors from the matched credit bureau and mortgage data, we compare the performance of the honest homeowners, the fraudulent investors, and the honest declared investors. We find that the fraudulent investors, after controlling for available characteristics, performed substantially worse than otherwise similar honest homeowners and declared investors.

Using the credit bureau data, we gain an understanding of the borrowers' other consumer liabilities, particularly bank cards. We find that the fraudulent investors who defaulted on their

[1] See the Financial Crisis Inquiry Report, 2011.

mortgages had significantly lower bank card utilization rates and were likelier to be current on these accounts relative to both honest homeowners and declared investors.

The remainder of the paper is organized as follows. Section II describes the related literature. Section III describes the data we have used. Section IV documents our definition of mortgage occupancy fraud. Section V provides descriptive statistics on occupancy fraud. Section VI covers loan performance. Section VII presents the results from estimating our econometric models. And Section VIII concludes.

II. Related Literature

This paper is not the first to examine the role of owner occupancy fraud and its impact on loan performance. Although they do not focused on fraud per se, Haughwout, Lee, Tracy, and van der Klaauw (2011) were among the first to use credit bureau data to explore the role of real estate investors during the mortgage boom and to show that the self-reported occupancy status may paint a misleading picture. They document significant increases in the share of purchase mortgages attributed to borrowers with multiple first lien mortgages in their credit files, with as many as half of all purchase mortgages attributable to investors in states that experienced the largest housing booms and busts. They also show that such investors account for a substantial share of defaults.

Several different strands of mortgage misrepresentation are explored in the literature. Garmaise (2015) explores the role of borrower misreporting of personal assets just above round number thresholds. He finds that borrowers who reported above-threshold assets were 25 percentage points more likely to default. Mian and Sufi (2015) explore the role of fraudulent income overstatement on mortgage applications. They compare the growth in income as implied by mortgage applications with the average Internal Revenue Service–reported income growth at the zip code level, and they find that there was substantial divergence between these two series. Income overstatement was higher in zip codes with low credit scores and low incomes; Mian and Sufi show that borrowers in these zip codes experienced some of the most significant increases in mortgage credit during the boom.

Piskorski et al. (2015) analyze privately securitized loans and find that second lien misrepresentation was widespread and occurred late in the intermediation process (e.g., by the underwriters of the residential mortgage-backed securities). More relevant to our paper, in their Internet Appendix, they detail additional analysis on the role of owner occupancy misrepresentation in their sample of privately securitized loans. They infer owner occupancy misrepresentation by comparing the property zip code reported by the residential mortgage backed securities (RMBS) trustee with 12 months of credit bureau–reported zip codes for the matched borrower. If none of these zip codes matches, then the authors conclude that this loan was characterized by owner occupancy fraud.

Note that this method of inference does not allow them to identify within–zip code misrepresentation, that is, fraudulent investors who misrepresent their owner occupancy status in the zip code in which they normally live. These “smart money” investors (Li, 2015) are likely aware of local trends and factors that should affect the value of local real estate, as opposed to distant speculators who trade on noise and create mispricing in local markets (Chinco and Mayer, 2015). We show, however, that occupancy fraud affected mortgages originated both to those living in the same zip code as well as those in different zip codes.

Griffin and Maturana (2015a) also examine three types of fraud (unreported second liens, owner occupancy misreporting, and appraisal overstatements) in privately securitized loans by matching to deeds data. They find that nearly half of the loans examined had at least one form of fraud and that these loans had 51% higher delinquency rates than otherwise comparable loans. They also explore the extent to which mortgage fraud and misrepresentation were responsible for the recent house price boom–bust cycle (2015b).

III. Data Description

We use a data set known as CRISM, or Credit Risk Insight Servicing McDash.[2] It is a match between loan-level mortgage data from McDash Analytics (formerly known as LPS) and credit bureau data from Equifax. Personally identifiable information has been removed. We restrict our data to borrowers who

(1) Are listed as the “primary” borrower in CRISM;

[2] See Beraja et al (2015) for more detail on the CRISM dataset.

(2) Are available and listed as primary borrowers in the Federal Reserve Bank of New York Consumer Credit Panel (FRBNY CCP); and

(3) Originated a first lien *purchase*[3] mortgage loan for a single-family unit in the McDash data set between 2005 and 2007.

Crucial to our occupancy fraud identification process, we also focus on CRISM-matched borrowers who also appear in the FRBNY CCP so we can use information on the borrowers' scrambled addresses. We also restrict to borrowers who have scrambled address, zip code, and state data from Equifax one quarter before and two quarters after their matched McDash mortgages originated. Our definition of occupancy fraud is discussed in detail in Section IV.

We focus on borrowers with self-reported McDash occupancy type as owner occupant and declared investors; we do not include second homes in our analysis. We also exclude a small number of loans with origination loan-to-value ratios (LTVs) exceeding 120%, loans whose investor type six months after origination was a Ginnie Mae buyout loan, local housing authority, federal home loan bank, and unknown. We also drop mortgages in which McDash does not list the state where the mortgaged property is located and mortgages with origination amounts exceeding $1 million. We also exclude Equifax borrowers whose address type is a post office box either one quarter before or two quarters after their matched McDash first lien originated for our fraud identification algorithm to work effectively.

We found 3,727,623 McDash mortgages meeting criterion 3 who were also matched to consumers in Equifax. Among these approximately 3.7 million mortgages loans in CRISM, about 10% (369,541) were also found in the FRBNY CCP. So, we focused on these 369,541 mortgage loans, matched to 366,065 borrowers, and after we apply the additional restrictions described, our final data set consists of 198,844 loans matched to 193,073 distinct borrowers.

Our house price index (HPI) data come from CoreLogic, and we use zip code–level house price indices for single-family detached homes (including distressed sales) when available and state-level indices otherwise. Our county-level unemployment rates come from the Bureau of Labor Statistics (BLS)**IV. Defining Occupancy Fraud**

[3] It is more difficult to misrepresent owner occupancy in a refinance because the lender often requires proof of residency, such as a utility bill.

A key aspect of our experimental design is the identification of fraudulent investors. We discuss our definition and compare it with others in the literature. Importantly, the CRISM data enable us to compare the self-reported occupancy type from the McDash Analytics loan-level data with information from the borrowers' Equifax matched credit bureau file. Our goal is to identify and classify borrowers who self-reported owner occupancy on their mortgage applications (judged by the McDash data) but who appear to be investors judging by their credit history information. In our owner occupancy fraud classification algorithm, we focus on three pieces of information:

1. The self-reported occupancy type;
2. The count of first-lien mortgages six months after their matched McDash mortgage is originated;
3. The borrowers' Equifax scrambled address from one quarter before and two quarters after when the McDash mortgage originated.

Using these data, we identify three types of borrowers:

1. **Honest owner occupants:** These are reported in the McDash data set as having originated a owner-occupied home purchase loan and do not meet the criteria for fraudulent investors in No. 2 below.
2. **Fraudulent investors:** These are self-reported owner occupants who did not change their Equifax scrambled addresses within the one quarter before and two quarters after window around the time their matched McDash mortgages originated and have more than one first-lien mortgage on their credit files six months after origination.
3. **Declared investors:** These are borrowers who are reported in the McDash data set as taking out a mortgage for an invesent property.

We restrict our attention to borrowers in the McDash data with single-family property types so we can avoid situations in which our fraud classifier does not pick up an address change because of borrowers moving within a large, multifamily unit. Any concerns concerning the accuracy of the fraud classifier should bias downward the likelihood of finding that these borrowers behave differently.

Our methodology of identifying owner occupancy misrepresentation differs from other papers that have address the phenomenon, and we provide evidence that our definition has a number of benefits that improve on existing work. In particular, our paper borrows from Haughwout et al. (2011) in that our address change condition reduces the probability of double counting first-lien mortgages for borrowers who are in the process of selling and buying a house. Our matched sample covers a substantially broader set of mortgage types than the FRBNY CCP CoreLogic LoanPerformance Asset-Backed Security ABS data they used, which includes only privately securitized mortgages (primarily nonprime and jumbo mortgages.) CRISM data matching the entire universe of Equifax consumers with McDash Analytics mortgage loans were not available at the time. This is important because research conducted by Gao and Li (2012) showed that most residential real estate investors financing properties with mortgages are prime.

Also recall that Piskorski et al. (2015) classifies a loan as truly owner occupied if — for 12 months of data after the mortgage originates — one of the borrowers' zip codes from Equifax matches the property's zip code reported to the RMBS trustee. Our method enables us to identify fraudulent investors who purchase and finance a purportedly owner-occupied property in the same zip code. We show below that this represents a no-negligible portion of all fraudulent investors.

V. Descriptive Statistics

In this section, we compare origination characteristics by borrower type, that is, honest owner occupants, fraudulent investors, and declared investors. Table 2 features origination characteristics and others for easy comparison across the borrower types.

Incidence of Occupancy Fraud

In Table 2a we show the share of borrowers by vintage half years and the intended investor type of the mortgage who are classified as misrepresenting their occupancy status according to our definition of occupancy fraud. Similar to Piskorksi (2015) we find a significant drop in the share of owner occupancy misrepresentation among private-securitized loans from the first half of 2007 to the second half 2007, but we find slight increases in the share of owner occupancy misrepresentation among other types of loans such as FHA, GSE, and loans held on the bank's portfolio. Overall, our estimate of the share of borrowers misrepresenting their occupancy status peaks for the 2005 second half vintage at 7.1% while it declines to 5.4% for the 2007 second half vintage.*Originating FICO Scores*

Similar to Gao and Li (2012), who find that most *declared* residential real estate investors are prime, we both confirm their finding that self-declared investors are largely prime and add that the fraudulent investors we identified from the self-declared owner occupants are overwhelmingly prime as well. In Table 2 we see that the subprime share was highest among the honest homeowners, but the subprime share for the fraudulent investors was 50% less for the fraudulent investors relative to the honest homeowners. In line with Gao and Li (2012), declared investors were largely prime.

The Geography of Mortgage Occupancy Fraud

Because many jumbo mortgages are located in states such as California with high costs of living, it would be advantageous to understand the geographic distribution of the occupancy fraud across the U.S. In Figure 1, we plot on a national heat map the state-level mortgage occupancy fraud rate for self-reported owner-occupied mortgages for loans that were originated between 2005 and 2007. The geographic patterns are informative: In the continental U.S., it appears that occupancy fraud rates were highest in the bubble states of California, Nevada, Arizona, and Florida. Nationally, the occupancy fraud rate ranged from an estimated low of 1.54% in Kansas to an estimated high 15.30% in Hawaii. The second, third, fourth, and fifth highest fraud rates belong to California (14.20%), Nevada (11.20%), Florida (9.64%), and Arizona (9.07%), respectively.

More broadly, the origination characteristics appear to suggest that the fraudulent investors took on substantially riskier mortgages than declared investors and honest homeowners. We will see later in this paper how they performed on their debt obligations as the housing boom came to an end and house prices began their collapse across the country.

Originating Loan-to-Value Ratios

Not surprisingly, based on the reported first-lien origination loan-to-value (LTV) ratio, honest homeowners put down the lowest down payments, and declared investors put down the highest down payments on their properties. Fraudulent investors had a mean origination LTV ratio of 79.5%, closer to honest homeowners' mean LTV (81%) than to the LTV ratio of the declared investors (75.5%). This higher LTV represented a substantial advantage for borrowers who misrepresented their occupancy type because originators tend to require higher down payments from declared investors to compensate for the known additional risk of default associated with real estate investors.

Second Liens and Combined Loan-to-Value Ratios

Our credit bureau data allow us to check for the incidence of both closed-end and revolving second liens. We focus on the presence of second liens around the time the Equifax borrowers originated their matched McDash first-lien purchase mortgages; specifically, we wait two quarters to capture the second liens to allow time for them to appear in the credit bureau data. We find that the fraudulent investors behaved much more similar to the declared investors than the honest homeowners in terms of the incidence of second liens. We find that 28.9% of the honest homeowners had second liens around first lien origination, and 50.9% of fraudulent investors and 50.7% of declared investors had second liens around origination (see Table 2).

The widespread incidence of second liens around origination implies that the LTV ratios calculated from the matched McDash first liens are an underestimate of the true equity positions of the borrowers. Because the credit bureau data have not only the count of second liens but also the balance associated with them, we add these balances to the origination amount and divide by the appraisal amount to get an estimate of the combined LTV ratio for each property. We find that the equity positions were worse than the first-lien LTV ratios implied, particularly for the fraudulent investors (see Table 2), who had the highest combined loan-to-value ratios (CLTV) at origination at 86% relative to 85% for honest homeowners and 83.2% for declared investors.

Incidence of Adjustable-Rate, Interest-Only, and Jumbo Mortgages

Among the self-reported owner occupants, 12% of those we identify as honestly representing their owner occupancy financed their homes with an adjustable-rate mortgage (ARM). However, we find that fraudulent investors were 8.3 percentage points more likely to have entered into an ARM. Fraudulent investors' higher preference for ARMs more closely resembled that of declared investors (Table 2), who financed their properties with ARMs at a rate of 18.5% on average. This is consistent with the possibility that fraudulent investors were intending to hold these properties for a short period of time, thus making them less sensitive to changes in interest rates. An alternative explanation is that taking out an ARM is motivated by a desired by investors to conserve liquidity.

Similarly, the share of mortgages that were interest only at origination was substantially higher for the fraudulent investors among the self-declared pool of owner occupants across both prime and subprime borrowers. At its peak, fraudulent investors in the 2006 vintage were more than twice as likely as honest homeowners to have had an interest-only mortgage. Interestingly, the interest-only share for the 2007 vintage of declared investors was more similar to that of the honest homeowners (Figure 2).

We also find that fraudulent investors identified from the pool of self-declared owner occupants were more likely to take out a jumbo mortgage, that is, one with an origination amount that exceeded the GSEs' (e.g., Fannie Mae, Freddie Mac) conforming loan limit — the maximum value of a mortgage that they can buy from the originator. In fact, we found that this is true for loans that originated in both the bubble states where the housing boom and bust was accentuated and in states where the boom and bust was not as pronounced. And indeed, the contrast with honest homeowners is more noticeable in nonbubble states.

Investor Type

Both fraudulent and declared investors are much less likely to have FHA-guaranteed loans. This is likely because of the stricter enforcement of FHA owner occupancy requirements. In addition, fraudulent investors are less likely to have GSE loans, even less than declared investors. In part, this is associated with the fact that fraudulent investors are more likely to have jumbo loans. Thus, they are also likelier to have private securities mortgages, or loans held in portfolio.

VI. Loan Performance

Cross-Sectional View of Mortgage Payment Performance

We investigate the delinquency and default behavior of these borrowers by examining the rate at which borrowers became 60 or more days past due as of December 2008. For loans that originated between 2005 and 2007, the fraudulent investors identified from the pool of self-reported owner occupants became seriously delinquent or defaulted at more than twice the rate of honest homeowners and declared investors. Table 3a summarizes the delinquency rates by year of origination and borrower type.

The differences in loan performance were particularly striking for borrowers in the prime (with FICO scores between 680 and 739) and super prime (with FICO scores between 740 and 850) credit score categories. As of December 2008, fraudulent investors went into serious delinquency at more than four times the rate of honest homeowners among the pool of super prime–declared owner occupants (Figure 4). Among borrowers with originating FICO scores between 680 and 739 as of December 2008, 7.9% of the honest homeowners had gone into serious delinquency or default, while 21.5% of the fraudulent investors identified from the population of self-declared owner occupants had gone into serious delinquency or default. The differences are also substantial among the nonprime borrowers (with FICO scores between 620 and 679) and even among the subprime (with FICO scores between 550 and 619) borrowers (Table 3b).

House prices peaked in early 2007 and began to fall until early 2011. In one set of analyses, we followed our borrowers until December 2008, the first year and a half of the collapse in house prices. By December 2008, 35% of honest homeowners and 43% of fraudulent investors were underwater with their mortgages, that is, the outstanding value of the mortgage exceeded the estimated value of the property (Figure 5). Not surprisingly, this is associated with much higher default rates. Among fraudulent investors with mark-to-market updated LTV ratios as of December 2008 between 100% and 110% (those with slight negative equity), 22% were seriously delinquent or in foreclosure. This compares with 12% of honest homeowners who were seriously delinquent or in default or foreclosure. For borrowers with deeper negative equity (in excess of 110% updated LTV), serious delinquency or default rates were 83% higher for fraudulent investors than honest owner occupants (15 percentage points higher default rate).

Strategic Default: Evidence from Other Consumer Liabilities

We now present evidence from the mortgage borrowers' matched credit bureau data to argue that these fraudulent owner occupants may have acted strategically in their default decisions. That is, the borrowers may have defaulted on their mortgages, not because of an inability to pay, but rather an unwillingness to pay, driven by substantial declines in home value that caused many borrowers into negative equity on their properties. We first capture each borrower's bank card utilization rate as of December 2008 as a proxy for the borrowers' liquidity (Elul et al., 2010) and calculate the utilization rate along three different dimensions: all national borrowers, borrowers with their mortgaged properties located in bubble states, and those with properties located in nonbubble states. We also divided those who became seriously delinquent or defaulted and those who remained current or at most 30 days past due within each geographic group.

On average, there was very little difference in the bank card utilization or default rates among the three borrower types (honest homeowners, fraudulent investors, and declared investors). Utilization and default rates are slightly lower among fraudulent and declared investors, consistent with their slightly higher credit scores (Table 2). But there is significant heterogeneity in the bank card utilization rate among borrowers who became seriously delinquent or defaulted as of December 2008. At the mean, fraudulent borrowers who had become seriously delinquent or in default on their mortgages had much more liquidity as measured by bank card utilization rate than honest homeowners in default identified from the pool of declared owner occupants and declared investors, and they are likelier to be current on their bank cards (Tables 4a and 4b).

As shown in Figure 5, fraudulent investors were also more likely to have negative equity in their homes as of December 2008. And those with negative equity were much more likely to be seriously delinquent or in default as of December 2008. Not only were "underwater" borrowers much more likely to be in serious delinquency or in default, but the fraudulent investors also had the highest default rates relative to the honest homeowners and the declared investors. Figures 6a and 6b show the share of borrowers in serious delinquency or in default stratified by updated LTV ratios exceeding 100%, that is, borrowers who now owed more on their mortgages than their houses were worth on the market. We note that the delinquency rate monotically increases as the updated LTV increase within each borrower type group.

Taken together, the evidence on bank card utilization and negative equity at default point to the possibility that fraudulent investors who we identified from the pool of declared owner occupants may have acted strategically in defaulting, that is, their default decisions are more likely to be driven by considerations of home equity rather than liquidity.

VII. Estimations and Results

Probability of Default

We first estimate a probit model for the probability that a loan defaults as of December 2008. We include a variety of mortgage origination characteristics, local house price and unemployment dynamics, and data drawn from other parts of the borrowers' credit histories as explanatory variables. We cannot identify and control for lender-specific fixed effects, but Griffin and Maturana (2015a) show that there is very little variation in owner occupancy misreporting across lenders, suggesting that it is likely that these decisions were made by the borrowers, perhaps in conjunction with brokers.

In the first specification in Table 5, we estimate the probability of default by December 2008 with a variety of origination characteristics known to influence the probability of default. In addition, we include fixed effects for the investor type of the loans six months after it was originated (FHA, GSE, portfolio, or private label); private label is the excluded category. State and origination year fixed effects are also included (2005 is the excluded origination year). Percent changes in zip code–level house price and unemployment dynamics are captured as well from origination to December 2008. All covariates have the expected signs and are highly statistically and economically significant. The second specification reports the marginal effects for this specification.

Of particular interest is the occupancy fraud dummy variable, which we note is both highly statistically significant and economically significant in explaining the probability of mortgage default. The marginal effects computed in specification (2) imply that, holding all else constant, a fraudulent investor is approximately 6 percentage points more likely to default than an otherwise similar borrower, relative to an average default rate of 10 percentage points.

In the third column, we include an additional dummy variable for whether the borrower has multiple first-lien mortgages in his or her credit bureau file six months after his or her matched McDash first-lien mortgage was originated. The coefficient is highly statistically significant in explaining mortgage default. This may be because, with several mortgages, borrowers are particularly vulnerable to falling prices. Recall from Table 2 that declared investors are very likely to have multiple first liens. And indeed, this indicator now accounts for all of the elevated default risk of these loans. On the other hand, although fraudulent investors all have multiple liens (by construction) and including this indicator reduces the explanatory power of the fraudulent investor dummy by approximately one-third, significant default risk remains.[4]

In the fourth specification, we estimate the probability of default with a probit model similar to specification (3) but we also include a dummy variable for whether the borrower had a second lien mortgage reported to the credit bureau six months after his or or her matched McDash first-lien mortgage was originated. Like the coefficient on the multiple first-lien dummy, the coefficient is also highly statistically significant in explaining mortgage default but is half as large as that of the coefficient for multiple first-liens. Inclusion of the second liens flag slightly reduces the explanatory power of the fraudulent investor dummy variable.

In the fifth specification, we estimate the probability of default with a probit model similar to specification (1) but where we allow for the possibility of interaction between our two types of investors — fraudulent and declared — and whether the loans was FHA guaranteed, GSE, private securitized, or held in portfolio. For declared investors, we find that the interaction effects are not statistically significant, that is, declared investors have higher default rates regardless of investor type. This is not the case for fraudulent investors; however, fraudulent investors with GSE — and even more so FHA — investor types are significantly less likely to default.

[4] In this specification, the fraud indicator gives the additional default risk associated with this group on top of that associated with having multiple liens.

In the sixth specification, we estimate a model similar to that of column 1 but remove the state fixed effects and replace them with a dummy variable for bubble states, that is, California, Arizona, Nevada, and Florida. The bubble state dummy flag has the expected strongly positive sign. Although the interaction with the fraudulent investor dummy is negative, it is insignificant. By contrast, the interaction for the declared investors is strongly negative, indicating that the latter groups' contribution to default in the bubble states was smaller.

Probability of High Bank Card Utilization

In Table 6a, we estimate two probit models for the probability — among borrowers who did not yet become seriously delinquent or default on their first-lien mortgage — that their total bank card utilization rate as of December 2008 (reported in their credit bureau files) exceeds 80%. Specification (2) reports the marginal effects from specification (1). We include origination characteristics in both specifications (1) and (3), but we drop the state fixed effects from specification (3) and instead include a bubble state dummy variable. In both specifications, we find that after controlling for other characteristics (e.g., origination FICO score), occupancy fraud is not statistically significant; that is, for mortgages that are current as of December 2008, there is no significant difference in bank card utilization rate between honest homeowners and fraudulent investors. Declared investors are modestly more likely to have higher utilization.

As far as the other covariates, higher origination FICO scores are associated with lower utilization. Higher unemployment rates and higher current LTV ratios are associated with higher utilization, likely reflecting local economic stresses. ARM borrowers have higher utilization; this is consistent with earlier work showing that they are borrowing constrained (Johnson and Li, 2014).

In Table 6b, we estimate the similar probit models for the probability of high credit utilization, but we focus on the group of borrowers who eventually had a seriously delinquency or default as measured by becoming 60 days or more past due on their first-lien mortgages. Utilization is measured here at the time of the first default. In specification (2), we report the marginal effects from the estimation of specification (1). The results are similar to those in Table 6a, with several notable exceptions. First, fraudulent investors are significantly less likely to have high utilization rates, reflecting, we suggest, a more strategic approach to default. In particular, they are 4 percentage points less likely to have had high bank card utilization rates, relative to an average of

31% for the entire sample. In addition, high current LTV is associated with lower incidence of high utilization, in contrast to the previous table. This is consistent with the "double-trigger" theory of mortgage default (see, for example, Elul et al., 2010).

Determinants of Fraudulent Investors

In Table 7, we estimate two probit models for the probability of a self-declared owner occupant being a fraudulent investor. Recall our definition of a fraudulent investor: These are self-reported owner occupants who did not change their Equifax scrambled addresses within the one quarter before and two quarters after window around the time their matched McDash mortgages originated and have more than one first-lien mortgage on their credit files six months after origination.

Confirming our summary statistics, we find that FHA loans are 4.7 percentage points less likely to be fraudulent, relative to an overall rate of 6.1 percentage points. GSE-guaranteed loans are also modestly less likely to be fraudulent.

Fraudulent investors are also associated with various indicators of housing bubbles such as higher house price appreciation. Similarly, in specification (3), we replace the state fixed effects with a bubble state dummy and find a significant positive coefficient on the bubble state dummy.

Fraudulent investors are also associated with higher origination amounts, interest-only loans, ARMs, low or unknown documentation loans, and broker-originated loans (as in the previous literature). They are also significantly more common in 2007, consistent with Haughwout et al. (2011).

VIII. Conclusion

Using a matched credit bureau and mortgage data set to identify occupancy fraud in residential mortgages originated between 2005 and 2007, we find that such fraud was widespread. In contrast to previous studies, our data set allows us to show that occupancy fraud was common in the GSE market and in loans held in portfolio in addition to the private label market. We found that mortgage borrowers who misrepresented their occupancy status performed worse than otherwise similar owner occupancy occupants and declared investors, with an incidence of default at nearly twice that of honest owner occupants or declared investors. Fraudulent investors' bank card utilization rates and default rates relative to those of honest owner occupants and declared investors imply that the fraudulent investors' mortgage defaults may have been strategic. Our results and estimates are large and economically significant and demonstrate one important role that occupancy fraud played during U.S. housing boom and bust.

References

Ben-David, Itzhak. “Financial Constraints and Inflated Home Prices during the Real Estate Boom.” *American Economic Journal: Applied Economics*, 3(3): 55–87, 2011.

Beraja, Martin, Andreas Fuster, Erik Hurst, and Joseph Vavra. “Regional Heterogeneity and Monetary Policy.” Federal Reserve Bank of New York Staff Report No. 731, June 2015

Chinco, Alex, and Chris Mayer. “Misinformed Speculators and Mispricing in the Housing Market.” Forthcoming. 2015.

Elul, Ronel, Nicholas S. Souleles, Souphala Chomsisengphet, Dennis Glennon, and Robert Hunt. “What ‘Triggers’ Mortgage Default?” *American Economic Review,* 100(2): 490–494, 2010.

Gao, Zhenyu, and Wenli Li. “Real Estate Investors and the Boom and the Bust of the U.S. Housing Market.” Working Paper, September 2012.

Garmaise, Mark J. “Borrower Misreporting and Loan Performance.” *The Journal of Finance*, 70 (1): 449–484, 2015.

Griffin, John M., and Gonzalo Maturana. “Who Facilitated Misreporting in Securitized Loans?” *The Journal of Finance,* 2015a.

Griffin, John M., and Gonzalo, Maturana. “Did Dubious Mortgage Origination Practices Distort House Prices?” Working Paper, May 2015b.

Haughwout, Andrew, Donghoon Lee, Joseph Tracy, and Wilbert van der Klaauw. “Real Estate Investors, the Leverage Cycle, and the Housing Market Crisis.” Federal Reserve Bank of New York Staff, Report 514, September 2011.

Johnson, Kathleeen W., and Geng Li. “Are Adjustable-Rate Mortgage Borrowers Borrowing Constrained?” *Real Estate Economics,* 42 (2): 457–471, 2014.

Li, Wenli. “Smart Money or Dumb Money: Investors’ Role in the Housing Bubble.” Federal Reserve Bank of Philadelphia *Business Review,* 2015Q3.

Mian, Atif, and Amir Sufi. “Fraudulent Income Overstatement on Mortgage Applications During the Credit Expansion of 2002 to 2005.” NBER Working Paper 20947, 2015.

Piskorski, Tomasz, Amit Seru, and James Witkin. “Asset Quality Misrepresentation by Financial Intermediaries: Evidence from the RBMS Market.” *The Journal of Finance,* 70 (6): 2635–2678, 2015.

Table 1: Variable Descriptions
Source: Authors' calculations of McDash Analytics, Federal Reserve Bank of New York Consumer Credit Panel/Equifax, and Equifax Credit Risk Insight Servicing data

Variable	Description
Occupancy Fraud Flag	1, if McDash Analytics–reported occupancy type = 1, first mortgage count (from Equifax) >1 six months after origination, and (Equifax) scrambled address one quarter before McDash loan originates is equal to (Equifax) scrambled address two quarters after McDash loan originates
Descriptive Statistics Tables	
Borrower Classification Share	Share of borrowers classified by type (Equifax, McDash Analytics, CRISM)
Borrower Classification Share by Origination Dollars	Share of borrowers classified by type by dollars originated (Equifax, McDash Analytics, CRISM)
Borrower Classification Share as of December 2008 Defaults	Among borrowers who defaulted by December 2008, share of borrowers classified by type (Equifax, McDash Analytics, CRISM)
First Default	Chronologically first McDash Analytics as_of_mon_id where mba_stat in ('6','9','F','R','L') (McDash Analytics)
Bubble State Share	Share of mortgaged properties in California, Nevada, Arizona, and Florida (McDash Analytics)
Subprime Share	Share of borrowers with origination FICO scores <660 (McDash Analytics)
FICO (Origination) (Mean)	Mean number of borrowers with originating FICO scores <660 (McDash Analytics)
LTV Ratio (Origination) (Mean)	Mean LTV ratio of borrowers (McDash Analytics)
CLTV Ratio (Origination) (Mean)	Balance of all mortgages on the property, divided by the property's appraised value at origination, in percent (Equifax, McDash Analytics)
Percent Change in HPI from Origination to December 2008 (Mean)	Percentage change in the property's zip code–level CoreLogic house price index from origination to December 2008; if zip code level is not available, the state level is used
Share of Borrowers with Second Liens Around Origination	Share of borrowers with second liens two quarters after origination (Equifax)
Interest Rate (Origination) (Mean)	Mean interest rate at origination (McDash Analytics)
Share Broker Originated	Share of borrowers whose loan source type from McDash Analytics at origination is broker (McDash Analytics)
ARM Share	Share of borrowers with an ARM at origination (McDash Analytics)
Interest-Only Share	Share of borrowers with an interest-only mortgage at origination (McDash Analytics)
Jumbo Share	A mortgage whose origination amount exceeding the GSE's conforming loan limit in the origination year (McDash Analytics)
Investor Type: PLS Share	McDash Analytics–reported investor type six months after originations = Private label security/All (McDash Analytics)
Investor Type: GSE Share	McDash Analytics–reported investor type six months after originations = GSE investor/All (McDash Analytics)

Investor Type: Portfolio Share	McDash Analytics-reported investor type six months after originations = Mortgages retained on banks' balance sheet/All (McDash Analytics)
Bank Card Utilization	Total bank card balance/Total bank card limit in past three months, as of December 2008 (Equifax)
Share Bank Card Utilization >80%	1 if bank card utilization is greater than 0.80 as of December 2008 (Equifax)
Bank Card Default Rate	Number of bankcard accounts – Number of bankcard always paid as agreed (i.e., never delinquent)/Number of bankcard accounts (Equifax)
Share of Borrowers with at Least One "Default" (60+ days past due) through July 2015	Share of borrowers with at least one McDash Analytics mba_stat variable in ('6','9','F','R','L') through July 2015 (McDash Analytics)
Updated LTV Ratio (December 2008) (Mean)	Principal balance (as of December 2008)/([Origination amount/LTV ratio] * [1+ Zip code–level HPI appreciation from origination to December 2008]), mean (McDash Analytics, CoreLogic)
Updated LTV Ratio at First Default (Mean)	Principal balance (at first default)/([Origination amount/LTV ratio] * [1 + Zip code–level HPI appreciation from origination to first default]), mean (McDash Analytics, CoreLogic)
	Regressions
Multiple First Liens (Origination) Flag	1, if count of first-lien mortgage from Equifax six months after the Equifax borrower's CRISM-matched McDash loan originated > 1
Second Liens (Orig) Flag	1, if borrower has a second lien two quarters after origination
Declared Investor Flag	1, if McDash Analytics occupancy type = 3
Interest Rate (Origination)	Interest rate at origination (McDash Analytics)
FICO (Origination)	Originating FICO credit score (McDash Analytics)
Origination Amount (Log)	Natural logarithm of the (McDash Analytics) origination amount
LTV Ratio (Origination)	LTV ratio at origination from McDash Analytics
LTV Ratio (Origination) >80 Flag	1, if McDash Analytics LTV_ratio (at origination) >80
Interest-Only Flag	1, if interest-only mortgage at origination (McDash Analytics)
Jumbo Flag	1, if origination amount exceeding the GSE's conforming loan limit in the origination year (McDash Analytics)
ARM Flag	1, if ARM at origination (McDash Analytics)
Low Doc Flag	1, if McDash Analytics flags the loan as low document type at origination
Unknown Doc Flag	1, if McDash Analytics flags the loan as having an unknown document type at origination
Correspondent Flag	1, if McDash Analytics marks the loan source type at origination as "correspondent" (loan_src_type=7)
Broker Flag	1, if McDash Analytics marks the loan source type at origination as "broker" (loan_src_type=2)
% Change 2-Year Lagged HPI	Percentage change in the property's zip code–level CoreLogic house price index two years before the McDash Analytics loan originating; if zip code level is not available, the state level is used
% Change HPI (Origination) to December 2008	Percentage change in the property's zip code–level CoreLogic house price index from origination to December 2008; if zip code level is not available, the state level is used

Unemployment Rate at Close Date	Property's zip code–level unemployment rate in the month it closed (BLS)
% Change Unemployment (Origination) to December 2008	Percentage change in the property's zip code-level unemployment rate from origination to December 2008 (BLS)
FHA Flag	1, if investor type (six months after origination) is FHA (McDash Analytics)
GSE Flag	1, if investor type (six months after origination) is GSE (McDash Analytics)
Portfolio Flag	1, if investor type (six months after origination) is portfolio (McDash Analytics)
Bubble State Flag	1, if prop_state is California, Nevada, Arizona, and Florida (McDash Analytics)
Updated LTV Ratio (December 2008)	Principal balance (as of December 2008)/((Origination amount/LTV ratio) * [1 + Zip code–level HPI appreciation from origination to December 2008]) (McDash Analytics, CoreLogic)
Updated LTV at Default	Principal balance (at First Default)/((Origination amount /LTV ratio) * (1 + Zip code–level HPI appreciation from origination to first default) (McDash Analytics, CoreLogic)
% Change Unemployment Until Default	Percentage change in the property's zip code–level unemployment rate from origination to its first default (McDash Analytics, BLS)

Table 2: Summary Statistics by Borrower Type
Source: Authors' calculations of McDash Analytics, Federal Reserve Bank of New York Consumer Credit Panel/Equifax, and Equifax Credit Risk Insight Servicing data

Characteristic	Honest Owner Occupant	Fraudulent Investor	Declared Investor
Sample Size (Loans)	174,901	12,128	11,815
Share	88.0%	6.1%	5.9%
Share by Origination Dollars	87.5%	8.1%	4.4%
Share as of December 2008 Defaults	82.6%	11.7%	5.7%
Bubble State	19.6%	39.2%	28.6%
Subprime	27.0%	18.0%	11.7%
FICO (Origination)	703	713	726
LTV Ratio (Origination)	81%	78%	75%
CLTV Ratio (Origination)	85%	86%	83%
LTV > 80% or LTV = 80 + 2nd lien near origination	43.1%	34.2%	27.6%
Percent Change in HPI from Origination to December 2008	-12.4%	-19.2%	-14.9%
Multiple First Liens	15.4%	100%	62.1%
Second Liens around Origination	28.9%	50.9%	50.7%
Interest Rate (Origination)	6.4%	6.5%	6.9%
Broker Originated	20.0%	27.1%	24.0%
ARM	12.0%	20.3%	18.5%
Interest Only	15.0%	29.0%	18.2%
Jumbo	10.0%	20.7%	5.4%
Investor Types: FHA	12.3%	2.2%	1.9%
Investor Type: PLS	24.6%	40.3%	38.2%
Investor Type: GSE	54.0%	45.0%	52.0%
Investor Type: Portfolio	9.1%	12.4%	8.0%
Bank Card Utilization (December 2008)	37.5%	36.8%	34.6%
Bank Card Utilization >80%	19.6%	18.9%	17.6%
Bank Card Default (December 2008)	14.2%	13.0%	11.5%
Serious Delinq/Default (60+ DPD) Through July 2015	27%	43%	27%

Table 2a: Fraud Share of Borrowers by Vintage Half Year and Intended Investor Type
Source: Authors' calculations of McDash Analytics, Federal Reserve Bank of New York Consumer Credit Panel/Equifax, and Equifax Credit Risk Insight Servicing data

	Fraud Share of Borrowers by Vintage and Investor Type				
	All	**FHA**	**GSE**	**Private Securitized**	**Portfolio**
2005 Second Half	7.1	1.2	5.9	9.6	9.6
2006 First Half	6.5	1.1	4.7	10.1	8.7
2006 Second Half	6.4	1.2	5.1	10.5	7.4
2007 First Half	6.0	1.4	5.4	11.3	6.9
2007 Second Half	5.4	1.4	5.6	8.2	9.1

Table 3a: Percent Seriously Delinquent or in Default as of December 2008 by Vintage
Source: Authors' calculations of McDash Analytics, Federal Reserve Bank of New York Consumer Credit Panel/Equifax, and Equifax Credit Risk Insight Servicing data

	Origination Year		
Borrower Type	**2005**	**2006**	**2007**
Honest Owner Occupant	9.0%	12.8%	7.5%
Fraudulent Investor	17.3%	26.4%	15.8%
Declared Investor	9.6%	13.6%	6.8%

Table 3b: Percent Seriously Delinquent or in Default as of December 2008 by Originating FICO Scores
Source: Authors' calculations of McDash Analytics, Federal Reserve Bank of New York Consumer Credit Panel/Equifax, and Equifax Credit Risk Insight Servicing data

Originating FICO Score	Honest Owner Occupant (%)	Fraudulent Investor (%)	Declared Investor (%)
Deep Subprime (350–549)	32.9	36.8	51.6
Subprime (550–619)	26.6	43.1	28.8
Nonprime (620–679)	16.7	36.4	20.9
Prime (680–739)	7.9	21.5	12.0
Super Prime (740–850)	2.2	9.3	4.2

Table 4a: Summary Statistics for Borrowers Serious Delinquency or Default as of December 2008
Source: Authors' calculations of McDash Analytics, Federal Reserve Bank of New York Consumer Credit Panel/Equifax, and Equifax Credit Risk Insight Servicing data

	Honest Owner Occupant (%)	**Fraudulent Investor (%)**	**Declared Investor (%)**
Share in Serious Delinquency or Default (December 2008)	9.8	20.0	10.0
Investor Type: PLS Share	49.5	63.8	67.5
Updated LTV Ratio (December 2008) (Mean)	108.8	113.5	105.6
Share Broker Originated	26.5	34.5	21.7
Bank Card Utilization (December 2008) (Mean)	75	56	65
Bank Card Default (December 2008) (Mean)	42	28	35
Share Bank Card Utilization >80%	55	38	46

Table 4b: Summary Statistics for Borrowers at First Serious Default
Source: Authors' calculations of McDash Analytics, Federal Reserve Bank of New York Consumer Credit Panel/Equifax, and Equifax Credit Risk Insight Servicing data

	Honest Owner Occupant (%)	**Fraudulent Investor (%)**	**Declared Investor (%)**
Share with a First Default Through July 2015	26.7	42.8	26.7
Investor Type: PLS Share	36.4	52.4	51.4
Updated LTV Ratio at First Default (Mean)	101.9	104.3	98.4
Share Broker Originated	23.5	31.9	26.6
Bank Card Utilization (First Default) (Mean)	66.3	51.5	57.8
Bank Card Default (First Default) (Mean)	29.1	18.9	23.8
Share Bank Card Utilization >80%	31.6	24.9	27.4

Table 5: Mortgage Default as of December 2008

The table reports estimations of probit models of mortgage default on or before December 2008 (including loans that terminated prior to December 2008). Specification (3) adds the multiple first liens six months after origination flag. Specification (4) adds a dummy for seconds liens six months after origination. Specification (5) adds interactions between investor type and fraud/declared investors to the baseline specification. Specification (6) uses a bubble state dummy instead of state fixed effects. Source: Authors' calculations of McDash Analytics, Federal Reserve Bank of New York Consumer Credit Panel/Equifax, and Equifax Credit Risk Insight Servicing data. Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1.

Variables	(1) Probit	(2) Marginal Effects	(3) Probit	(4) Probit	(5) Probit	(6) Probit
Occupancy Fraud Flag	0.369***	0.063***	0.235***	0.227***	0.406***	0.393***
	(0.016)	(0.003)	(0.019)	(0.019)	(0.023)	(0.022)
Multiple First Liens Flag (Orig)			0.177***	0.168***		
			(0.013)	(0.013)		
Second Lien Flag (Orig)				0.086***		
				(0.011)		
Declared Investor Flag	0.060***	0.010***	-0.031	-0.043*	0.075***	0.139***
	(0.021)	(0.004)	(0.022)	(0.022)	(0.028)	(0.026)
Interest Rate (Orig)	0.120***	0.021***	0.121***	0.123***	0.121***	0.121***
	(0.004)	(0.001)	(0.004)	(0.004)	(0.004)	(0.004)
FICO (Orig)	-0.007***	-0.001***	-0.007***	-0.007***	-0.007***	-0.007***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Origination Amount (Log)	-0.045***	-0.008***	-0.064***	-0.062***	-0.045***	-0.022**
	(0.011)	(0.002)	(0.011)	(0.011)	(0.011)	(0.010)
LTV Ratio (Orig)	0.017***	0.003***	0.018***	0.018***	0.017***	0.017***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Interest-Only Flag	0.309***	0.053***	0.300***	0.294***	0.309***	0.300***
	(0.014)	(0.002)	(0.014)	(0.014)	(0.014)	(0.013)
Jumbo Flag	-0.057***	-0.010***	-0.056***	-0.064***	-0.060***	-0.063***
	(0.019)	(0.003)	(0.019)	(0.019)	(0.019)	(0.018)
ARM Flag	0.163***	0.028***	0.159***	0.149***	0.161***	0.158***
	(0.014)	(0.002)	(0.014)	(0.014)	(0.014)	(0.014)
Low Doc Flag	0.175***	0.030***	0.172***	0.169***	0.174***	0.175***
	(0.012)	(0.002)	(0.012)	(0.012)	(0.012)	(0.012)
Unknown Doc Flag	0.068***	0.012***	0.065***	0.063***	0.067***	0.067***
	(0.011)	(0.002)	(0.011)	(0.011)	(0.011)	(0.011)
Correspondent Flag	0.001	0.000	-0.002	0.001	0.001	0.001
	(0.011)	(0.002)	(0.011)	(0.011)	(0.011)	(0.011)
Broker-Originated Flag	0.149***	0.026***	0.142***	0.139***	0.149***	0.153***
	(0.012)	(0.002)	(0.012)	(0.012)	(0.012)	(0.012)
% Change HPI (Orig) to Dec '08	-2.065***	-0.355***	-2.059***	-2.062***	-2.066***	-1.864***
	(0.054)	(0.009)	(0.054)	(0.054)	(0.054)	(0.043)
% Change Unemp (Orig) to Dec '08	0.080***	0.014***	0.080***	0.082***	0.079***	0.085***
	(0.018)	(0.003)	(0.018)	(0.018)	(0.018)	(0.014)
FHA Flag	-0.349***	-0.063***	-0.338***	-0.332***	-0.339***	-0.343***
	(0.019)	(0.003)	(0.019)	(0.019)	(0.019)	(0.019)
GSE Flag	-0.322***	-0.059***	-0.317***	-0.316***	-0.311***	-0.328***
	(0.013)	(0.003)	(0.013)	(0.013)	(0.014)	(0.013)
Portfolio Flag	-0.175***	-0.034***	-0.169***	-0.170***	-0.183***	-0.175***
	(0.016)	(0.003)	(0.016)	(0.016)	(0.018)	(0.016)
Fraudulent Investor#FHA Flag					-0.390***	
					(0.112)	
Fraudulent Investor#GSE Flag					-0.099***	
					(0.036)	
Fraudulent Investor#Portfolio Flag					0.058	
					(0.049)	

Declared Investor#GSE Flag					-0.046	
					(0.044)	
Declared Investor#Portfolio Flag					0.035	
					(0.070)	
Bubble State Flag						0.150***
						(0.016)
Fraudulent Investor#Bubble State Flag						-0.044
						(0.032)
Declared Investor#Bubble State Flag						-0.202***
						(0.041)
Origination Year: 2006	0.086***	0.015***	0.078***	0.079***	0.085***	0.093***
	(0.012)	(0.002)	(0.012)	(0.012)	(0.012)	(0.011)
Origination Year: 2007	-0.041***	-0.007***	-0.049***	-0.048***	-0.041***	-0.036***
	(0.013)	(0.002)	(0.013)	(0.013)	(0.013)	(0.013)
Observations	161,698	161,698	161,698	161,698	161,687	161,698
State FE	Yes	Yes	Yes	Yes	Yes	No

Table 6a: Probit Models of High Bank Card Utilization (Borrowers who did not Default as of Dec 2008)
These are probit models for the probability of a borrower having bank card utilization (greater than 80%) as of December 2008 among borrowers who did not default on their mortgage. A variety of mortgage characteristics are included as controls in addition to other covariates. Specification (3) uses bubble state fixed effect instead of state fixed effects. Source: Authors' calculations of McDash Analytics, Federal Reserve Bank of New York Consumer Credit Panel/Equifax, and Equifax Credit Risk Insight Servicing data. Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1.

VARIABLES	(1) Probit	(2) Marginal Effects	(3) Probit
Occupancy Fraud Flag	-0.010	-0.002	-0.009
	(0.024)	(0.005)	(0.024)
Interest Rate (Orig)	0.075***	0.015***	0.076***
	(0.007)	(0.001)	(0.007)
FICO (Orig)	-0.009***	-0.002***	-0.009***
	(0.000)	(0.000)	(0.000)
Origination Amount (Log)	-0.050***	-0.010***	-0.045***
	(0.012)	(0.003)	(0.011)
LTV Ratio (Orig)	0.005***	0.001***	0.005***
	(0.001)	(0.000)	(0.001)
Interest-Only (Orig) Flag	0.137***	0.028***	0.141***
	(0.017)	(0.004)	(0.017)
Jumbo Flag	-0.040	-0.008	-0.047*
	(0.025)	(0.005)	(0.025)
ARM Flag	0.129***	0.026***	0.130***
	(0.021)	(0.004)	(0.021)
Low Doc Flag	0.030**	0.006**	0.027*
	(0.015)	(0.003)	(0.015)
Unknown Doc Flag	-0.006	-0.001	-0.010
	(0.012)	(0.003)	(0.012)
Correspondent Flag	-0.004	-0.001	-0.004
	(0.013)	(0.003)	(0.013)
Broker Flag	0.027*	0.006*	0.026*
	(0.015)	(0.003)	(0.015)
Declared Investor Flag	0.056**	0.011**	0.059**
	(0.025)	(0.005)	(0.025)
Updated LTV Ratio (Dec '08)	0.004***	0.001***	0.004***
	(0.000)	(0.000)	(0.000)
% Change Unemp (Dec '08)	0.044**	0.009**	0.068***
	(0.022)	(0.004)	(0.017)
FHA Flag	-0.061***	-0.013***	-0.058**
	(0.023)	(0.005)	(0.023)
GSE Flag	-0.136***	-0.029***	-0.138***
	(0.017)	(0.004)	(0.017)
Portfolio Flag	-0.066***	-0.014***	-0.065***
	(0.021)	(0.005)	(0.021)
Bubble State Flag			-0.075***
			(0.019)
Origination Year: 2006	0.001	0.000	-0.001
	(0.015)	(0.003)	(0.014)
Origination Year: 2007	0.032**	0.007**	0.030**
	(0.015)	(0.003)	(0.015)
Observations	97,335	97,335	97,335
State FE	Yes	Yes	No

Table 6b: Probit Models of High Bank Card Utilization (at the time of first mortgage default)

These are probit models for the probability of a borrower having high bank card utilization (defined as greater than 80%) at the first time a borrower who defaults on their mortgage (defined as being 60 days or more past due). A variety of mortgage origination characteristics are included as controls in addition to other covariates. Specification (3) uses bubble state fixed effects instead of state fixed effects. Source: Authors' calculations of McDash Analytics, Federal Reserve Bank of New York Consumer Credit Panel/Equifax, and Equifax Credit Risk Insight Servicing data. Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1.

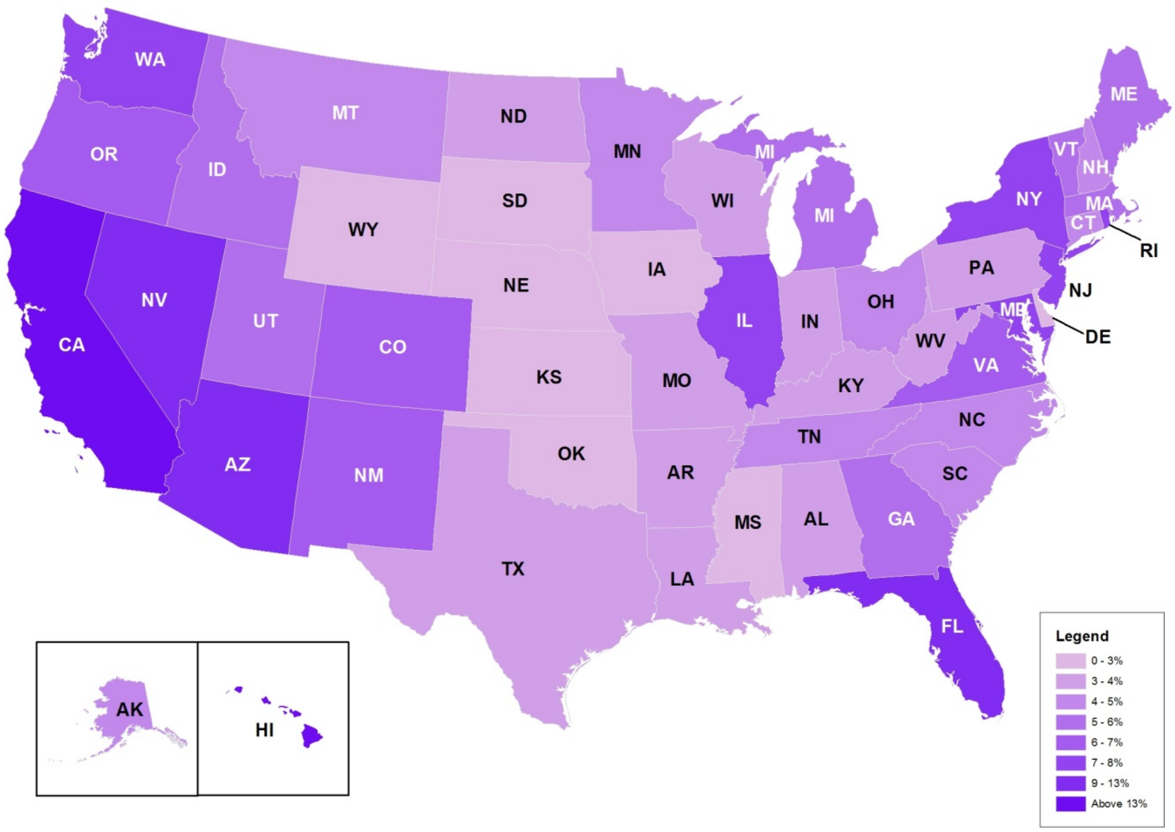
Variables	(1) Probit	(2) Marginal Effects	(3) Probit
Occupancy Fraud Flag	-0.122***	-0.042***	-0.122***
	(0.022)	(0.008)	(0.022)
Interest Rate (Orig)	-0.018***	-0.006***	-0.016***
	(0.006)	(0.002)	(0.006)
FICO (Orig)	-0.002***	-0.001***	-0.002***
	(0.000)	(0.000)	(0.000)
Origination Amount (Log)	0.123***	0.042***	0.114***
	(0.015)	(0.005)	(0.014)
LTV Ratio (Orig)	0.004***	0.001***	0.004***
	(0.001)	(0.000)	(0.001)
Interest-Only Flag	-0.008	-0.003	-0.019
	(0.018)	(0.006)	(0.018)
Jumbo Flag	-0.022	-0.008	-0.043*
	(0.025)	(0.009)	(0.024)
ARM Flag	-0.019	-0.007	-0.018
	(0.019)	(0.007)	(0.019)
Low Doc Flag	0.015	0.005	0.014
	(0.017)	(0.006)	(0.017)
Unknown Doc Flag	0.016	0.006	0.017
	(0.015)	(0.005)	(0.015)
Correspondent Flag	-0.013	-0.004	-0.020
	(0.016)	(0.006)	(0.016)
Broker Flag	-0.009	-0.003	-0.012
	(0.016)	(0.006)	(0.016)
Declared Investor Flag	-0.006	-0.002	-0.001
	(0.029)	(0.010)	(0.029)
Updated LTV at Default	-0.001***	-0.000***	-0.001***
	(0.000)	(0.000)	(0.000)
% Change Unemp until Default	0.007	0.002	0.076***
	(0.024)	(0.008)	(0.018)
FHA Flag	-0.015	-0.005	-0.011
	(0.026)	(0.009)	(0.026)
GSE Flag	0.023	0.008	0.030
	(0.018)	(0.006)	(0.018)
Portfolio Flag	-0.029	-0.010	-0.024
	(0.022)	(0.007)	(0.022)
Bubble State Flag			-0.152***
			(0.019)
Origination Year: 2006	0.021	0.007	0.010
	(0.016)	(0.005)	(0.016)
Origination Year: 2007	0.047***	0.016***	0.037**
	(0.018)	(0.006)	(0.018)
Observations	47,222	47,222	47,222
State FE	Yes	Yes	No

Table 7: Models for the Determinants of Fraudulent Investors

These are probit models for the probability that a self-declared owner occupant is a fraudulent investor. A variety of mortgage origination characteristics are included as controls in addition to other covariates. Specification (3) uses bubble state fixed effects instead of state fixed effects. Source: Authors' calculations of McDash Analytics, Federal Reserve Bank of New York Consumer Credit Panel/Equifax, and Equifax Credit Risk Insight Servicing data. Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1.

Variables	(1) Probit	(2) Marginal Effects	(3) Probit
Interest Rate (Orig)	0.051***	0.006***	0.053***
	(0.005)	(0.001)	(0.005)
FICO (Orig)	0.001***	0.000***	0.001***
	(0.000)	(0.000)	(0.000)
Origination Amount (Log)	0.169***	0.020***	0.210***
	(0.013)	(0.002)	(0.012)
LTV Ratio (Orig)	-0.001*	-0.000*	-0.001***
	(0.000)	(0.000)	(0.000)
Interest Only Flag (Orig)	0.196***	0.023***	0.207***
	(0.014)	(0.002)	(0.014)
Jumbo Flag	-0.083***	-0.010***	-0.100***
	(0.020)	(0.002)	(0.020)
ARM Flag	0.227***	0.027***	0.232***
	(0.016)	(0.002)	(0.016)
Low Doc Flag	0.165***	0.020***	0.162***
	(0.015)	(0.002)	(0.015)
Unknown Doc Flag	0.099***	0.012***	0.093***
	(0.012)	(0.001)	(0.012)
Correspond Flag	0.008	0.001	0.003
	(0.013)	(0.002)	(0.013)
Broker Flag	0.153***	0.018***	0.158***
	(0.013)	(0.002)	(0.013)
% Change 2 Year Lagged HPI	0.198***	0.024***	0.268***
	(0.041)	(0.005)	(0.036)
Unemp Rate at Origination	0.019***	0.002***	0.028***
	(0.005)	(0.001)	(0.004)
FHA Flag	-0.539***	-0.049***	-0.531***
	(0.031)	(0.002)	(0.030)
GSE Flag	-0.096***	-0.012***	-0.101***
	(0.015)	(0.002)	(0.015)
Portfolio Flag	-0.011	-0.001	-0.014
	(0.018)	(0.002)	(0.018)
Bubble State Flag			0.200***
			(0.014)
Origination Year: 2006	0.020	0.002	0.022*
	(0.013)	(0.002)	(0.013)
Origination Year: 2007	0.105***	0.013***	0.118***
	(0.017)	(0.002)	(0.016)
Observations	156,416	156,416	156,416
State FE	Yes	Yes	No

Figure 1: The Geography of Occupancy Fraud* (State-Level Mortgage Occupancy Fraud Rate Among Self-Reported Owner Occupied Properties for 2005—2007 Properties)
Source: Authors' calculations of McDash Analytics, Federal Reserve Bank of New York Consumer Credit Panel/Equifax, and Equifax Credit Risk Insight Servicing data



*The authors would like to thank Adam Scavette (Federal Reserve Bank of Philadelphia) for his help in creating this map.

Figure 2: Interest-Only Mortgage Share by Borrower Type by Year of Origination
Source: Authors' calculations of McDash Analytics, Federal Reserve Bank of New York Consumer Credit Panel/Equifax, and Equifax Credit Risk Insight Servicing data

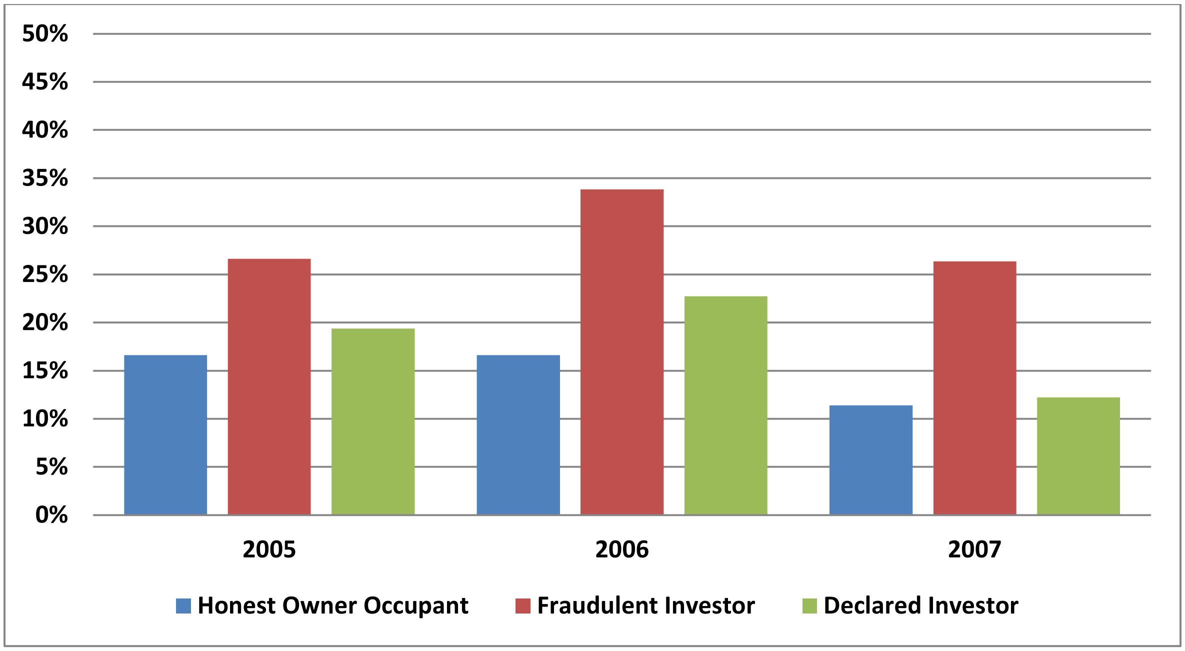


Figure 3: Percent Seriously Delinquent or Foreclosed as of December 2008 by Borrower Type
Source: Authors' calculations of McDash Analytics, Federal Reserve Bank of New York Consumer Credit Panel/Equifax, and Equifax Credit Risk Insight Servicing data

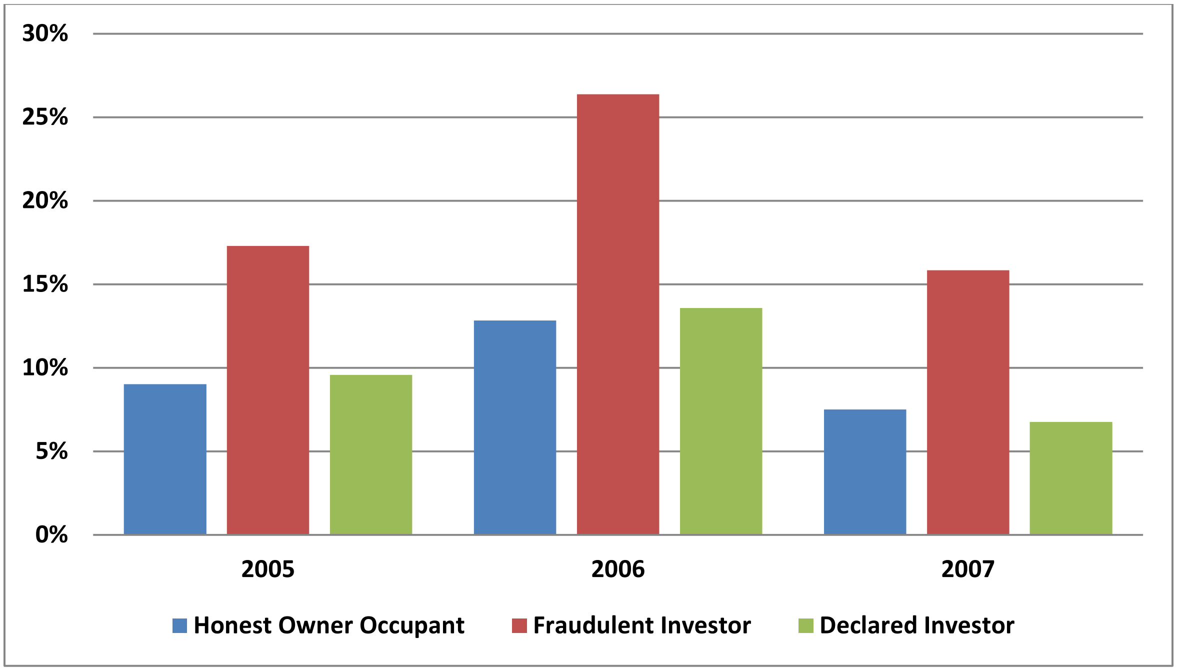


Figure 4: Percent of Borrowers Seriously Delinquent or in Default or Foreclosure as of December 2008: Super Prime (740–850) Borrowers
Source: Authors' calculations of McDash Analytics, Federal Reserve Bank of New York Consumer Credit Panel/Equifax, and Equifax Credit Risk Insight Servicing data

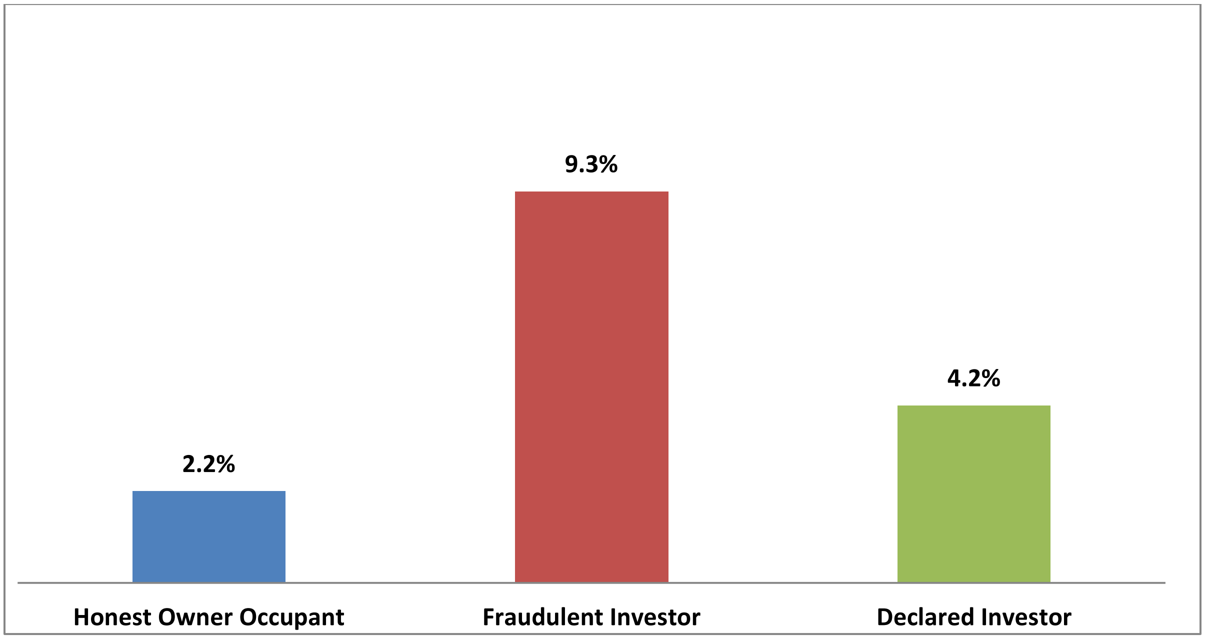


Figure 5: Share of Borrowers Underwater as of December 2008 by Origination Year
Source: Authors' calculations of McDash Analytics, Federal Reserve Bank of New York Consumer Credit Panel/Equifax, and Equifax Credit Risk Insight Servicing data

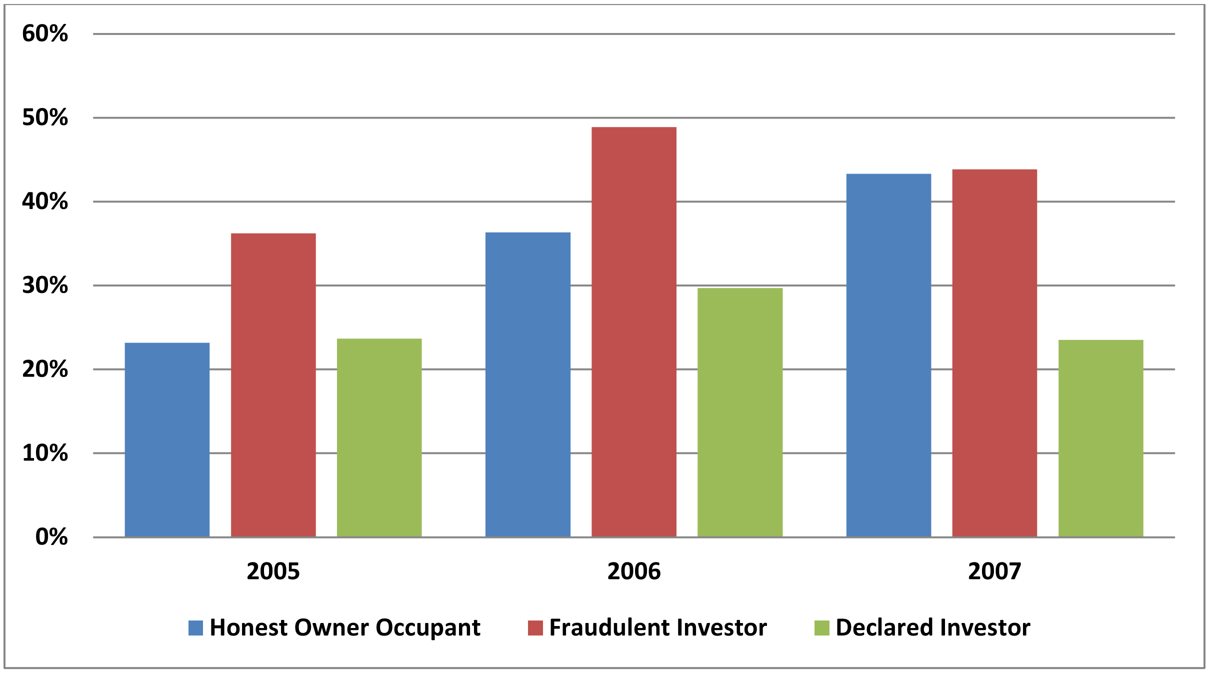


Figure 6a: Share of Borrowers Seriously Delinquent or in Default as of December 2008 for Borrowers with Updated Loan-to-Value Ratio Between 100% and 110%
Source: Authors' calculations of McDash Analytics, Federal Reserve Bank of New York Consumer Credit Panel/Equifax, and Equifax Credit Risk Insight Servicing data

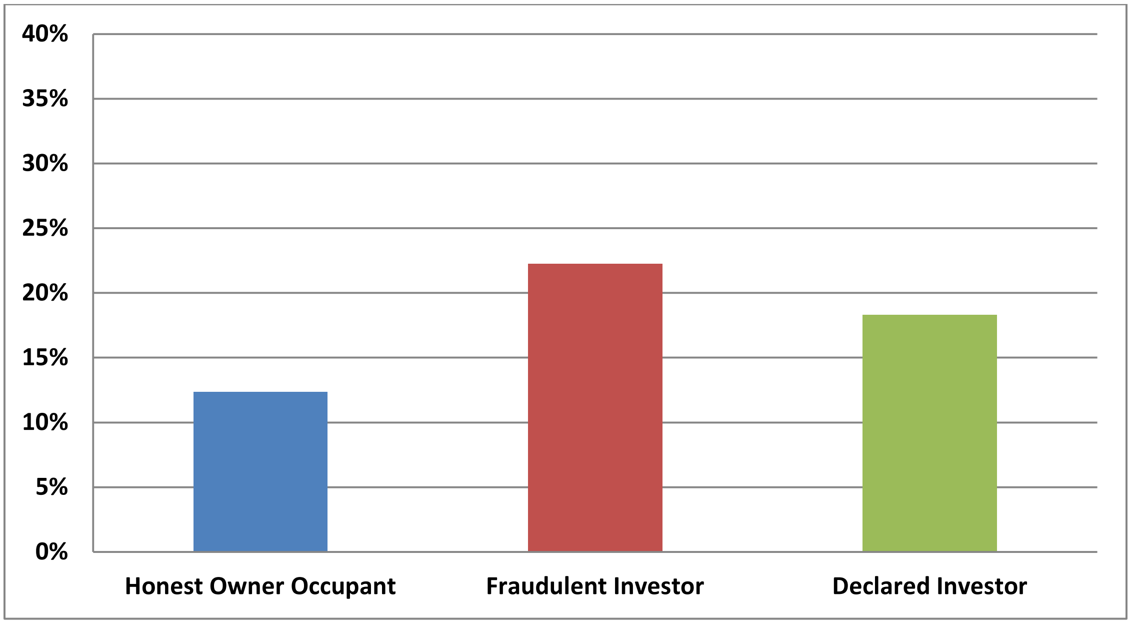


Figure 6b: Share of Borrowers Seriously Delinquent or in Default as of December 2008 for Borrowers with Updated Loan-to-Value Ratio Between 110% and 125%
Source: Authors' calculations of McDash Analytics, Federal Reserve Bank of New York Consumer Credit Panel/Equifax, and Equifax Credit Risk Insight Servicing data

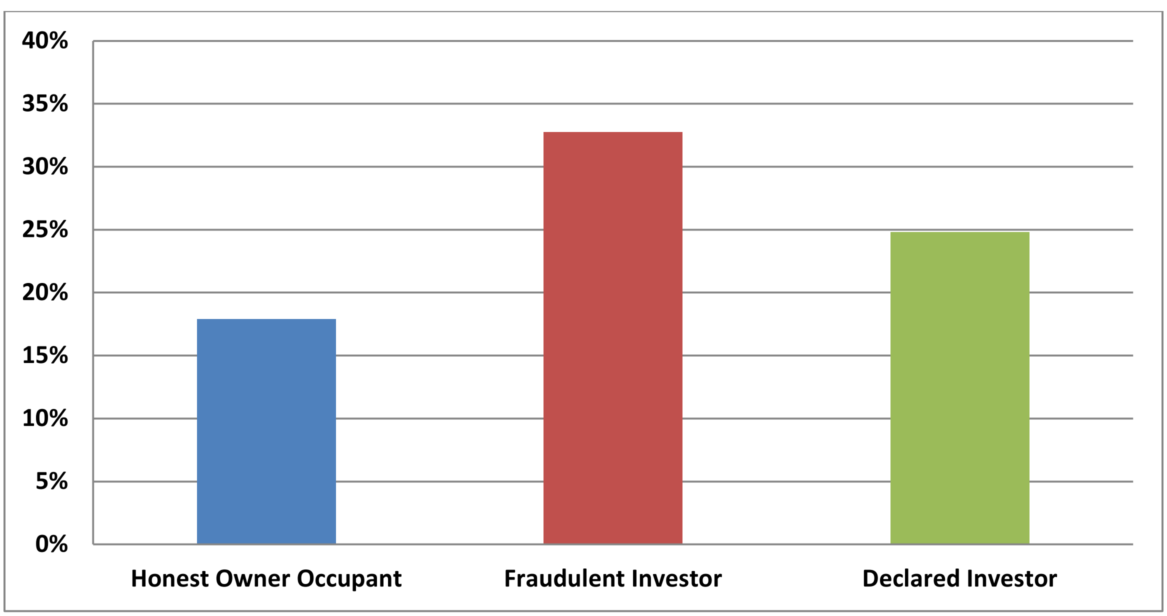


Figure 7a: Total Investor Share by Investor Type in the Bubble States
Source: Authors' calculations of McDash Analytics, Federal Reserve Bank of New York Consumer Credit Panel/Equifax, and Equifax Credit Risk Insight Servicing data

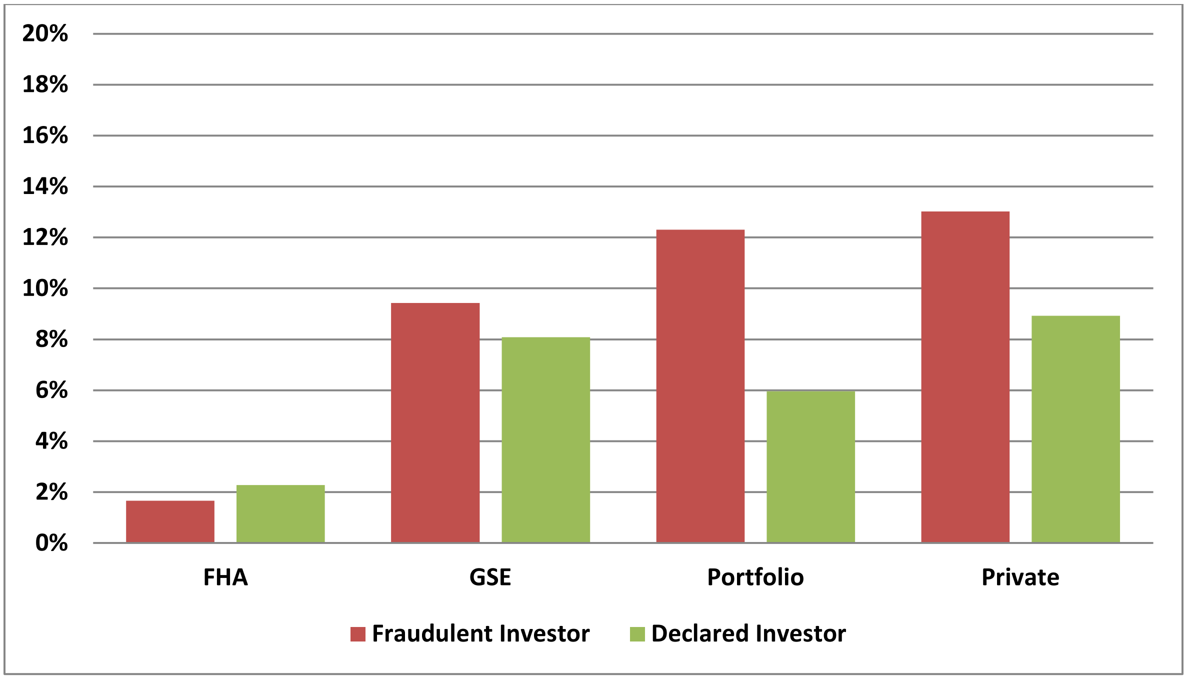


Figure 7b: Total Investor Share by Investor Type Nationally Excluding Bubble States

Source: Authors' calculations of McDash Analytics, Federal Reserve Bank of New York Consumer Credit Panel/Equifax, and Equifax Credit Risk Insight Servicing data