FEDERAL RESERVE BANK
PHILADELPHIA

Working Papers
RESEARCH DEPARTMENT

Owner-Occupancy Fraud and Mortgage Performance

Ronel Elul
Federal Reserve Bank of Philadelphia Research Department

Aaron Payne
University of Pennsylvania

Sebastian Tilson

WP 23-01

PUBLISHED
January 2023

FRB CFI
PHILADELPHIA

ISSN: 1962-5361

DOI: https://doi.org/10.21799/frbp.wp.2023.01

Owner-Occupancy Fraud and Mortgage Performance[1]

Ronel Elul, Aaron Payne, and Sebastian Tilson[2]

January 2023

Abstract

We identify occupancy fraud — borrowers who misrepresent their occupancy status as owner-occupants rather than investors — in residential mortgage originations. Unlike previous work, we show that fraud was prevalent in originations not just during the housing bubble, but also persists through more recent times. We also demonstrate that fraud is broad-based and appears in government-sponsored enterprise and bank portfolio loans, not just in private securitization; these fraudulent borrowers make up one-third of the effective investor population. Occupancy fraud allows riskier borrowers to obtain credit at lower interest rates. These fraudulent borrowers perform substantially worse than similar declared investors, defaulting at a 75 percent higher rate. Their defaults are also likelier to be "strategic," suggesting that they pose a risk in the face of declining house prices.

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

JEL Codes: D12, R3

---

[1] The authors thank Neil Bhutta, Dan Gorin, Tomasz Piskorski, Felipe Severino, Stijn Van Nieuwerburgh and James Vickery, and participants at the Eastern Economic Association, AREUEA summer meetings, the Wharton Finance Macro lunch, and the FDIC Consumer Symposium for helpful comments and discussions. We are also grateful to Natalie Newton for excellent research assistance.

[2] Corresponding author: Ronel Elul, Research Department, Federal Reserve Bank of Philadelphia. Email: ronel.elul@phil.frb.org; Tel: 215-574-3965. Aaron Payne: payne.aaron182@gmail.com. Sebastian Tilson: sebastian.tilson@gmail.com. When this research was started, Payne and Tilson were research assistants at the Federal Reserve Bank of Philadelphia.

I. Introduction

Policymakers and the popular press have cited anecdotal evidence to suggest that one of the contributing causes of the housing bubble was pervasive mortgage fraud.[3] Academic work has also verified the existence of mortgage fraud during the housing bubble along several dimensions. These include occupancy fraud, inflated prices, unreported second liens, and income misstatement; we discuss the literature in the following section.

In this paper, we use a matched credit bureau and mortgage data set to study occupancy fraud, which occurs when mortgage borrowers claim on a new purchase mortgage application that they will be the owner-occupants of the property[4] but do not move from their old address and consequently hold multiple first liens (as recorded in the credit bureau data). We argue that the fraudulent purchasers that we identify are very likely to be investors and that accounting for fraud increases the size of the effective investor population by nearly 50 percent.

We demonstrate that occupancy fraud is much more widespread and persistent than shown in the previous literature. It was pervasive during the bubble and did not affect just private securitized loans, appearing in both government-sponsored enterprise (GSE)–guaranteed and portfolio-held loans. More significantly, we also show that fraud continues to remain a concern,[5] appearing even in recent originations.[6]

An important benefit for those undertaking investor fraud was obtaining loan terms that they could not have obtained, based on observables, by declaring themselves as investors, such as higher LTVs and lower interest rates. Fraudulent loans are also riskier: even after controlling for available characteristics, they perform substantially worse, defaulting at nearly twice the rate of similar declared investors. We also show that fraudulent investors are more "strategic" in their default decisions, further highlighting their risk in the face of declining prices. Our results support the hypothesis that these borrowers undertook fraud to obtain favorable loan terms despite their higher risk.

The remainder of the paper is organized as follows. Section II describes the related literature. Section III describes the data we use. Section IV documents our definition of mortgage occupancy fraud. Section V provides descriptive statistics for our sample. Section VI studies the determinants of fraud. Section VII presents the results from estimating the impact of fraud on mortgage default. In Section VIII, we consider two alternative

[3] See, for instance, the Financial Crisis Inquiry Commission's 2011 *Financial Crisis Inquiry Report*.
[4] This entails not renting out the property and not intending to sell the property quickly.
[5] News organizations have also recently suggested that fraud, including occupancy fraud, may pose a risk as the housing market weakens. See, for example, "As mortgage market cools, fraud risk heats up," by Kyle Campbell, *American Banker*, June 3, 2022.
[6] We study originations through the end of 2017. (We stop at this point to ensure that all loans in our sample have sufficient loan performance information before the onset of the COVID pandemic.)

hypotheses that could also lead to a positive relationship between fraud and default and show that they are unlikely to explain our results. Section IX examines the robustness of our results to additional specifications and sample limitations, and Section X concludes.

II. Related Literature

This paper is not the first to examine the role of fraud and its impact on loan performance.

Several other types of mortgage misrepresentation have been studied in the literature. Garmaise (2015) shows that borrowers who misreported personal assets were more likely to default. Mian and Sufi (2015) explore the role of fraudulent income overstatement on mortgage applications and argue that it was associated with above-average increases in mortgage credit during the housing bubble. Ben-David (2011) finds evidence of borrowers inflating prices during the housing bubble to obtain larger mortgages, and that these borrowers were likelier to default. Yavas and Zhu (2019) demonstrate widespread underreporting of second liens in portfolio mortgages and privately securitized mortgages originated in 2005 and 2006 (thus going beyond private mortgage-backed securities, or MBS, as do we).

More closely related to our work, a subset of the mortgage misrepresentation literature examines the role of occupancy fraud. Piskorski et al. (2015) find widespread evidence of second-lien and occupancy status misrepresentation in private 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. Similarly, Griffin and Maturana (2016) examine three types of fraud (unreported second liens, owner-occupancy misreporting, and appraisal overstatements) in private securitized loans by matching loans to deeds data. They find that nearly half of the loans examined had at least one form of fraud and that these loans had 50 percent higher delinquency rates than otherwise comparable loans.

Our paper adds to the literature in several important ways. First of all, we show that occupancy fraud was widespread, including in the large GSE market, going beyond the private-securitized market studied in Piskorski et al. (2015) and Griffin and Maturana (2016). More significantly, we show that occupancy fraud is not just associated with the housing bubble preceding the Great Recession, but has persisted to the present day.[7] The reason is likely that this type of fraud is difficult to detect until long after the mortgage has been originated.

[7] By contrast, the incidence of second liens has dropped dramatically, from 38 percent of all purchase mortgages in 2005, to just 7.8 percent in 2017, thus greatly lessening the scope for this type of fraud.

We echo previous work in demonstrating that misrepresentation is associated with elevated default risk. Because fraudulent investors default at higher rates and are more sensitive to house price declines, our work also underscores that occupancy misrepresentation continues to pose a significant risk to the housing market and the broader economy. Our results therefore suggest that, in order to deter such fraud, lenders and/or regulators might seek to audit loans after origination; we leave the design of such a policy to future work.

III. Data Description

We obtain an initial sample of loans from a dataset known as CRISM: Equifax Credit Risks Insight Servicing and Black Knight McDash Data (henceforth CRISM).[8] This data set combines loan-level mortgage data from Black Knight McDash (henceforth McDash) and some limited credit bureau information on the borrower from Equifax and represents a broad segment of the mortgage market (approximately two-thirds of all mortgages originated since 2005). We then further refine our sample by matching it to a more detailed sample of credit bureau data known as the Federal Reserve Bank of New York Consumer Credit Panel/Equifax Data (henceforth CCP); this is a 5 percent random sample of borrowers with a credit file (see Lee and van der Klaauw, 2010, for further details). Thus, our matched data constitutes a random sub-sample of mortgage loans covered in McDash with additional information on borrower characteristics drawn from credit bureau files.[9] We restrict our sample to borrowers who:

(1) are listed as the "primary" borrower in CRISM;
(2) are available and listed as primary borrowers in the CCP; and
(3) originated a first-lien *purchase* mortgage loan for a single-family unit in the McDash data set between January 2005 and December 2017.

We discuss our definition of occupancy fraud in detail in Section IV.

We focus on borrowers with self-reported McDash occupancy type as owner-occupant, declared investor, or second homebuyer, dropping those with a missing or unknown occupancy type.[10] We also restrict to borrowers who have scrambled address, zip code, and state data from the CCP one quarter before and four quarters after their matched McDash mortgages originated, as our identification of fraud relies on this information.

---

[8] See Beraja et al. (2019) for more detail on the CRISM data set.
[9] Personally identifiable information is not included in these anonymized datasets.
[10] We also drop the small number of loans with a McDash investor type six months after origination indicating Ginnie Mae buyout loan, local housing authority, Federal Home Loan Bank, or unknown.

We begin with 15,596,981 loans to primary borrowers in CRISM that meet the criteria above.[11] After we restrict to primary borrowers in the CCP that also have the credit bureau address data described above, we are left with 659,831 loans.

Our house price index (HPI) data come from CoreLogic Solutions (henceforth CoreLogic), and we use zip code-level house price indices for single-family detached homes (including distressed sales). Our county-level unemployment rates come from the Bureau of Labor Statistics (BLS).

IV. Defining Occupancy Fraud

A key aspect of our research design is the identification of fraudulent investors. We discuss our definition and compare it with others in the literature. Importantly, we are able to compare the self-reported occupancy type from the McDash loan-level data with information from the borrower's credit bureau file. Our goal is to identify and classify borrowers who self-report as owner-occupants on their purchase mortgage applications but who appear to be investors judging by their credit history information. We focus on three pieces of information:

- The self-reported occupancy type listed in McDash
- The count of first-lien mortgages four quarters (as reported in CCP) after their matched McDash mortgage is originated
- The borrowers' CCP scrambled address from one quarter before and four quarters after the McDash mortgage origination date

We first note that, intuitively, a reported homeowner is likely to be an investor if: (i) they have multiple first liens, and (ii) they did not move following the origination of their new purchase mortgage. We say a borrower did not move if they have a CCP scrambled address that is the same one quarter before and four quarters after their matched McDash mortgage originated. We define multiple liens as having more than one first-lien mortgage when observed four quarters after the matched first-lien purchase mortgage was originated.

As validation for our intuition, note from Table 2a that 96 percent of all declared owner-occupants move around their mortgage origination, whereas this is the case for only 38 percent of all declared investor mortgages. Furthermore, 17 percent of declared owner-occupants have multiple first liens in their credit bureau files four quarters following mortgage origination, compared with 42 percent of declared investors. This supports our

[11] We also drop loans that are missing variables from McDash that are required for our analysis.

intuition that these two criteria serve to distinguish our population of fraudulent investors from those we will characterize as honest homeowners. While we impose no further restrictions on our definition of fraud in the main body of the paper, one might be concerned that our fraudulent investor group also includes (undeclared) second homeowners. In section IX we show that our key results remain robust to an additional criterion: that the distance between the McDash and CCP zip codes be no greater than 75 miles.[12] However, we do not impose this restriction since, without it, the distribution of the zip code distances is already similar between fraudulent and declared investors (see Table 3). Furthermore, as will be clear from our regression results (Table 5), the interest rate benefit from owner-occupancy fraud is much smaller for a second homeowner than for an investor.

Formally, we define four types of borrowers:

Honest owner-occupants: These are reported in the McDash data set as having originated an owner-occupied home purchase loan and whose CCP scrambled addresses are different one quarter before and four quarters after their matched McDash mortgage origination.[13]

Fraudulent investors: These are reported in the McDash data set as having originated an owner-occupied home purchase loan and whose CCP scrambled addresses is the same one quarter before and four quarters after their matched McDash mortgage originated. The borrower's credit bureau file also reports more than one first-lien mortgage four quarters after the matched first lien was originated.

Declared investors: These are borrowers who are reported in the McDash data set as taking out a mortgage for the purchase of an investment property.

Second homebuyers: These are borrowers who are reported in the McDash data set as taking out a mortgage for the purchase of a second home.

Our final data set, after we restrict to these four borrower types, consists of 584,499 observations.
We plot the share of fraudulent and declared investors over time in Figure 1. Observe that while the share of both types of investors drops sharply beginning in 2008, the ratio between the two remains constant over time: in most years, fraudulent investors make up roughly one-third of the total pool of investors.

We compare our final sample of 584,499 loans to the broad McDash data set of first lien purchase mortgages originated in this period (over 15 million loans) in Table 2b, which confirms that characteristics are indeed balanced between the two.

---

[12] The intuition is that second homes are likelier to be located further from the borrower's primary residence. Indeed, in Table 3, we see that this is the case: over 70 percent of declared investors have a CCP zip code that is within 75 miles of their McDash zip code, compared with around 30 percent of declared second homeowners.

[13] In identifying honest homeowners and fraudulent investors, we drop borrowers whose address is associated with a post office box, as this would make it difficult to distinguish true non-movers (this represents approximately 4 percent of these borrowers).

The most significant set of borrowers not classified above (and therefore dropped in our sample) are those who have only a single mortgage in their credit bureau (like honest homeowners) but do not move following their mortgage origination (like fraudulent investors).[14] As our classification gives conflicting guidance regarding this group, we drop them for the body of our analysis; however, we examine this group in detail in the Section IX and argue that they are likely a combination of frauds and honest homeowners.

We now discuss how our methodology of identifying owner-occupancy misrepresentation relates to that of other papers that have addressed the phenomenon.

Haughwout et al. (2011) also use multiple first liens to identify investors in credit bureau data, although they cannot explicitly tie this to fraud, as they do not have mortgage data to identify the reported occupancy type. Griffin and Maturana (2016) identify fraudulent borrowers as those for whom the tax mailing address (from deeds records) differs from that of the property on the loan. Similarly, Piskorski et al. (2015) identify fraudulent borrowers as those for whom the credit bureau zip code (interpreted as representing the borrower's true mailing address) is not the same as the loan property zip code.

Our approach combines these two types of information: the number of first liens and the CCP address. We do not have access to the property tax mailing address in our data set, so we cannot take the same approach as Griffin and Maturana (2016); our "no-move" restriction captures a similar intuition but relies on the credit file address. Unlike Piskorski et al. (2015), we use not just zip codes but more precise information on the (scrambled) CCP address; this also allows us to identify fraud that takes place within a zip code.

Finally, both Griffin and Maturana (2016) and Piskorski et al. (2015) confine their analysis to private securitized loans (primarily subprime and jumbo mortgages). By contrast, in using our credit bureau data on additional liens and addresses, we are able to study the extent of fraud across the entire universe of mortgage and loan types. As we show below, this substantially increases the total amount of fraud. In particular, we find significant incidence of fraud among prime GSE-guaranteed loans and also those held on bank portfolios. It also allows us to extend our analysis far beyond the housing boom and bust studied previously.

Note that our approach to identifying fraud relies on the credit bureau accurately updating borrower addresses and capturing first-lien mortgages. We now discuss corroborating evidence from other data sets to support these assumptions.

---

[14] This represents approximately 12 percent of those who identify themselves as owner-occupants.

First, using our definition of moving, we compute a moving rate in the overall CCP for our time period of 11.9 percent, which can be compared to a 12.3 percent rate in the American Community Survey (ACS).[15] In Figure 2, we also plot the state-year moving rates in the two data sets and show that they are highly correlated. To validate the credit bureau data on additional first-lien mortgages that we use to identify fraud, we compare our data set to the 2016 Survey of Consumer Finances (SCF). For consumers in our data set who purchased an investment property or second home in 2015, 50 percent did not have any other first liens. For purchase mortgages on investment properties and second homes in the SCF, the comparable figure (those who did not have a mortgage on their principal residence or other properties) is 42 percent.[16] While the two data sets have different designs, this does suggest that our data set indeed records multiple first liens reasonably accurately.[17]

Despite the evidence we provide, there may still be concerns about measurement error in our fraud definition. We discuss in Sections VII and VIII how these concerns might affect the interpretation of our estimates.

V. Descriptive Statistics

In this section, we compare descriptive characteristics by borrower type: honest owner-occupants, fraudulent investors, declared investors, and second homebuyers. A broad set of summary statistics is given in Table 3, broken into pre- and post-housing bubble originations.[18]

We observe from Table 3 that, while there is indeed high representation of private securitization among fraudulent mortgages during the bubble years of 2005–2007, were we to restrict attention to private securitized mortgages (as in the previous literature), we would have accounted for less than half of all fraudulent loans. As we see in Figure 4, there is a sharp drop in the share of owner-occupancy misrepresentation among private securitized loans in the second half of 2007 as standards tightened in this market (see also Piskorski et al., 2015). However, the share of fraud stays more steady for the other investor types. Overall, our estimate of the share of borrowers misrepresenting their occupancy status peaks in the first half of 2006 at 6.8 percent, and remains in the range of 2–3 percent for the post-bubble period (Figure 4).

Figures 3a and 3b show heat maps with the state-level fraud rates, broken into bubble and post-bubble periods. The areas with the highest bubble-era fraud rates were California and Washington, D.C., with fraud rates in

---

[15] For further discussion of moving in the CCP and how it compares to other data sources, see DeWaard et al. (2018).
[16] The SCF reports investment properties and second homes together. In addition, we restrict attention to 2015 originations, as this is the most recent full-year in the SCF data.
[17] Most notably, the SCF is a household-level survey, while the CCP and CRISM are individual-level. See also Dettling et al. (2015).
[18] See Table 1 for descriptions of variables.

excess of 13 percent (and exceeding the share of declared investors). Other states with high fraud rates include Hawaii, Nevada, Florida, and Arizona (Figure 3a). Many of these correspond to "bubble states," indicating that fraud was likelier in areas that experienced greater-than-average house price growth during the boom period and larger price declines during the bust. We also see in Figure 3b that, while fraud rates declined post-bubble, many of the same areas continue to see higher-than-average incidence.

Next, we show that fraudulent investors were riskier, in a number of dimensions observable *ex-ante*, than declared investors. Indeed, in our multivariate analysis below, we show that one key benefit of fraud was to obtain better terms than declared investors would receive for risky loans.

Fraudulent investors are likelier to have low credit scores than declared investors, both in the bubble and post-bubble periods, although on average they are still generally prime. In post-bubble years, fraudulent investors are much likelier to have high-LTV mortgages (origination LTV≥90 percent) than declared investors. Fraudulent investors are also much likelier to have additional second liens.[19] Fraudulent investors are more likely than any other borrower type to use the 2/28 ARMs that featured prominently in the housing boom and bust.

Finally, fraudulent investors have much larger mortgages than declared investors: the rate of jumbo loans is at least three times that of declared investors in both the bubble and post-bubble periods. This is consistent with two motivations for fraud that we discuss below: first, that it allowed frauds to circumvent tighter underwriting guidelines imposed by lenders on declared investors, and second, that the benefit from obtaining lower interest rates is greater for these larger loans.

In addition to being observably riskier at origination, fraudulent investors are also riskier ex-post. Their default rates are more than double those of declared investors in both periods. For bubble-era originations, their default rates are substantially higher than any other borrower types. Coupled with the fact that fraudulent investors take out larger loans, high default risk implies a high dollar share of defaults: frauds account for over 11 percent of all dollars in default over our entire sample, despite making up less than 4 percent of originations. In Figure 5 we break down the default rates by origination vintage and two key risk characteristics: FICO score at origination[20] and LTV at origination, confirming that, even within risk categories, fraudulent investors are riskier than declared investors. Our multivariate analysis will explore this further.

Finally, we will show in our analysis below that fraudulent investors' elevated default risk is driven by strategic motives, namely greater sensitivity to house price declines. The full analysis will consider several indicators of

---

[19] Furthermore, as we show below, once we restrict attention to declared investors with multiple first liens, the incidence of high-LTV loans is also much higher for fraudulent investors in the bubble years.
[20] FICO at origination is from McDash.

strategic behavior. However, from the summary statistics, we can compare borrowers' bankcard utilization rates, which can be viewed as an indicator of (il)liquidity (see Elul et al., 2010). Outside of default, fraudulent investors have a similar rate of high utilization (defined as utilization greater than or equal to 80 percent) as declared investors. However, for those in default, the difference is striking: Fraudulent investors' utilization rates are significantly *lower* than declared investors. This suggests that their default decisions were less likely to be driven by an inability to pay, and indeed, our analysis below suggests that negative equity played a greater role in their default decisions.

VI. Estimations and Results — Determinants of Fraud

The Extensive Margin

In order to systematically compare fraudulent and declared investors, we estimate probit models with fraud status as the outcome, restricting the sample to just these two borrower types. More formally, the probability that mortgage i is identified as fraudulent is modeled as $\Pr(fraud) = \Pr\,(y \leq Y_i)$, where y is normally distributed with mean 0 and variance 1, and

$$Y_i = \beta X_i \qquad (1)$$

X_i are loan (borrower) characteristics. We also include fixed effects for the property state and origination half-year vintage.[21] Standard errors are clustered by the county of the originated property. Table 4 column (1) reports the marginal effects from this regression. Fraudulent borrowers have lower FICO origination scores than do declared investors; a score below 660 raises the likelihood of being fraudulent by 10–20 percentage points. Frauds are also much likelier to have high LTVs at origination and larger loan amounts. These results suggest that a key motivation to commit fraud was to receive mortgages that would not have been available had they declared themselves as investors. Additionally, note that interest rates for fraudulent mortgages are lower than those for declared investors with similar characteristics; we explore this in more detail below. Frauds are more prevalent in private MBS and bank portfolio loans (relative to the omitted category, GSE loans) but are less likely to be FHA/VA loans. In column (2), we add a control for the observable characteristics that identify fraud, i.e., indicators for having multiple first liens and moving around origination; we see that frauds are no longer less frequent among FHA/VA loans since, by construction, they have multiple liens, making it hard to qualify for an FHA or VA loan. Fraudulent investors also exhibit other characteristics that suggest they are

[21] We include these fixed effects to account for the possibility that there may be cross-state or time heterogeneity in fraud prevalence that is correlated with cross-state or time heterogeneity in risk characteristics.

more constrained than declared investors: they are more likely to have second liens and tend to have mortgages with longer terms. To investigate whether or not frauds are attracted to "bubbles" (i.e., areas where house prices have recently appreciated), in column (3) we substitute state fixed effects with an indicator for "bubble states" (Arizona, California, Florida, and Nevada)[22] and interact this with an indicator for whether the loan was originated in 2005–2007 (the "bubble years") or 2008–2017. Consistent with the descriptive statistics, fraud is more common in the bubble states, especially during the bubble years.

Fraud and Interest Rates

The previous results suggest that fraudulent investors receive lower interest rates than declared investors. We now examine this in more detail. We estimate multivariate models of the interest rate at origination, controlling for various borrower, mortgage, and property characteristics. For mortgage *i,* receiving interest rate Y_i at origination (in percentage points), we estimate linear regression models of the form:

$$Y_i = \beta X_i + \gamma_{Borrower\,Type_i} + \epsilon_i \qquad (2)$$

where X_i is a set of mortgage and borrower characteristics at the time of origination, *Borrower Type*$_i$ is one of: honest homeowner, fraudulent investor, declared investor, or second homeowner, as described above, and $\gamma_{Borrower\ Type}$ is a fixed effect for *i*'s Borrower Type. We also include controls for the state of origination and the origination half-year vintage.

Our main results can be found in column (1) of Table 5. The reported covariates have the expected signs:[23] Borrowers with higher FICO scores at origination, or lower LTV ratios, pay lower interest rates. As is well-known, borrowers with small loans ($200,000 or below) pay higher rates. Turning to the covariate of interest, borrower type, we see that declared investors pay interest rates that are 26 basis points higher than fraudulent investors, while honest homeowners and second homeowners pay rates that are 9 and 6 basis points lower, respectively. One key difference between fraudulent investors and other borrower types is that the former are, by construction, required to have multiple first liens and to not move around the time of origination; to the extent that these two characteristics are correlated with unobservable determinants of interest rates that are visible to the lender, we may be overstating the difference in interest rates between fraudulent and declared

[22] These are the states that experienced the greatest degree of home price appreciation from the end of the 2001 recession through the end of 2005 (leaving aside Hawaii, which we exclude from this group because of its unique characteristics).
[23] The other covariates are not reported and can be found in the online appendix. They include mortgage term, interest rate type (FRM or ARM with varying initial rate periods), and investor type.

investors. Therefore, in column (2), we add controls for these characteristics;[24] however, our results do not significantly change.

We now interact the borrower type with key mortgage characteristics from X_i. This allows us to determine how the premium paid by different borrower types varies with risk characteristics; in addition, by interacting the borrower type with the investor type (GSE, portfolio, private securitization) we are also able to assess the degree to which different types of lenders are able to identify the *ex-ante* risk of fraud.[25] In Table 6, we report the marginal effects from estimating this model. Each cell represents the premium paid (or discount, if negative) by the borrower type in that column, for the risk characteristic in that row, relative to that paid by a fraudulent investor with the same risk characteristic. As in the previous table, column (2) adds controls for the whether the borrower has multiple first liens and moved; the results are once again not significantly different.

The degree of variation across risk characteristics is not dramatic. This suggests that the primary motivations for fraud were those explored in our extensive margin analysis: to obtain access to loan terms that would be otherwise unavailable if frauds instead declared as investors. We do note that for loans with high origination LTV (over 80 percent, and particularly over 90 percent), the premium paid by declared investors was particularly large, at over 30 basis points. Another interesting result is that the premium paid by declared investors is no smaller for portfolio loans than for the other investor types, which suggests that these lenders were not any better at identifying fraud. This further supports our previous findings that fraud is not associated only with securitization.

VII. Estimation and Results — Fraud and Default Behavior

Our previous analyses show that fraudulent investors are observably riskier at origination; from the summary statistics, we also see that they default at higher rates. In order to isolate the association between fraud and default risk, we estimate multivariate probit models of default where we control for risk characteristics. That is,

---

[24] While these characteristics (and the declared occupancy type) are used to define fraudulent investors, their coefficients can still be identified separately from the coefficients on borrower type. Note that the declared investor and second homeowner borrower types include observations with both single first liens and multiple first liens, as well as those that move and do not move following mortgage origination. Thus, these two borrower types together are sufficient to identify the "move" and "multiple firsts" indicators. In addition, honest homeowners can also have single or multiple first liens, which adds further observations than can be used to identify the "multiple firsts" coefficient. However, honest homeowners cannot be used to separately identify the "move" coefficient since, among declared homeowners, moving is perfectly colinear with fraud in our sample.

[25] While we cannot control for lender-specific fixed effects in this dataset, Griffin and Maturana (2016) show that, for loans in private mortgage-backed securities, there is very little variation in owner-occupancy misreporting across lenders, suggesting that decisions concerning fraud were made by the borrowers (possibly in conjunction with brokers, as our results below suggest).

the probability that mortgage i is in default[26] is modeled as $\Pr(Default) = \Pr(y \leq Y_i)$, where y is normally distributed with mean 0 and variance 1,

$$Y_i = \beta X_i + \delta Z_i + \gamma_{Borrower\ Type_i} \qquad (3)$$

We include a variety of mortgage and borrower characteristics at origination in X_i, and Z_i includes the following variables that capture the evolution of the mortgage and borrower risk in the two years following origination: the change in the local unemployment rate and the updated LTV (i.e., an estimate of what the LTV would be when the house price at origination is updated using the local house price index). We again include controls for the origination state and half-year vintage.

In the first specification of Table 7a, we pool all origination vintages together. This table reports only the key covariates — the full set can be found in the online appendix. Coefficients have the expected signs: higher origination FICO scores are associated with lower default risk; higher LTV ratios, both at origination and two years following origination, are associated with higher risk. We also create buckets for the change in county unemployment rates in the two years following origination: larger increases (i.e., higher buckets) are associated with higher default rates. Turning to the coefficients of interest, we see that, relative to fraudulent investors (the base category), all of the other borrower types have a substantially lower risk of default: a difference of 3.5–4.5 percentage points, which is significant relative to the sample average default rate of 6.5 percentage points. Another way to quantify the risk associated with fraud is to observe that the additional risk of a fraudulent loan is roughly equivalent to the difference in default rates between a near-prime borrower (FICO between 660 and 700) and a prime one (FICO above 700).

In column (2), we add indicator variables for the two criteria that we use to define fraud: having multiple first liens, and moving. Having just one first lien is indeed associated with significantly lower default risk (likely because of the lower debt burden relative to those with multiple first liens). Nevertheless, the higher risk of fraud remains even when we control for these two characteristics.

Since there are significant differences in default rates between the two periods of our sample (pre- and post-bubble) and underwriting tightened following the end of the housing bubble (eliminating many other types of fraud), we estimate our baseline default model separately for 2005–2007 and 2008–2017 originations. Column (3) confirms, not surprisingly, that the risk associated with fraudulent loans is larger for bubble-era originations: declared investors have default rates that are 6.4 percentage points lower than fraudulent investors (relative to

[26] Recall that a loan is deemed to be in "default" if it is ever seriously delinquent (60 or more days past due) or in default at some point in the two years following origination.

an overall default rate of 11 percent for this time period), with similar effects for the other borrower types. However, fraudulent loans continue to be riskier even in post-bubble originations (Column (4)): declared investors are 2.3 percentage points less likely to default than fraudulent investors (the overall default rate in this period is 4 percent). This suggests that occupancy fraud might continue to pose a risk to the mortgage market.

Although we provided evidence from external data to validate the credit bureau criteria that we use to identify fraudulent investors, one might nevertheless be concerned about measurement error in this bureau data. If such measurement error were uncorrelated with default outcomes, then this would lead to some honest homeowners being mistakenly characterized as fraudulent, and vice versa. For example, if some address changes took longer than four quarters to update in the credit bureau data, then some honest homeowners would be characterized as fraudulent. The result would be an attenuation in the difference between the default rates of fraudulent investors and the other borrower types. However, one might be more concerned that there are unobserved borrower characteristics leading to a correlation between fraud and default. In Section VIII we consider two alternative hypotheses that could lead to such a relationship. The first relates the incidence of multiple first liens to a weakening housing market, and the second hypothesis relates the not moving to default. We show that these hypotheses are unlikely to explain our key results.

Strategic Default

Having shown that fraudulent investors are significantly more likely to default than declared investors (and other borrowers), we now show that these default decisions are particularly sensitive to house prices, i.e., these borrowers are more "strategic." This will highlight the risk that they could pose in the face of declining house prices.

The literature has approached the question of strategic default in several ways. In Bhutta et al. (2017), default is characterized as strategic if it is triggered by negative equity.[27] We first examine this approach by adding interactions of several key risk characteristics (coded as categorical variables) with the borrower type to the models in Table 7a. These interacted covariates are: origination FICO score, updated LTV, and the two-year change in unemployment.

The marginal effects for these variables, by borrower type (relative to fraudulent investors), are reported in Table 7b. All of the coefficients are negative, reflecting the fact that fraudulent investors have higher default

[27] In order to reduce other sources of variation, Bhutta et al. (2017) restrict attention to borrowers who had zero equity at origination; unfortunately, this strategy is not feasible for us since our analysis also encompasses post-bubble originations. However, in unreported results we estimate a model that interacts origination LTV with the subsequent change in house prices and find results that are qualitatively similar to the ones here.

risk than other borrower types (as shown in Table 7a). For FICO and unemployment, however, there is only modest variation across the categories, suggesting that the extra risk of fraud is not differentially associated with, for instance, the intensity of local unemployment rate changes. Updated LTV is an exception, however: we find that fraudulent investors are substantially more likely to default than other borrower types when their updated LTV is above 100 percent (i.e., when equity is negative). For instance, when equity is positive (updated LTV<100 percent), fraudulent investors are only 3 percentage points more likely to default than declared investors, whereas for the highest category (updated LTV ≥120 percent), the difference is 12 percentage points. In columns (2) and (3), we show that the updated LTV gradient is large both in the bubble years and post-bubble originations, supporting the continued sensitivity of frauds to house price declines.

Our second test of strategic default is inspired by Elul et al. (2010). In that paper, high bankcard utilization rates at the time of mortgage default are associated with illiquidity; that is, such default that is *not* strategic, since it reflects an attempt by the borrower to drawn down other sources of credit to try to avoid default or, alternatively, captures the accumulation of past liquidity shocks. We obtain bankcard utilization from the borrower's matched credit bureau record: For loans that do not default in our two-year observation window, bankcard utilization is measured two years following origination; for defaulting mortgages, we measure utilization at the time of first default. We have included non-defaulting mortgages in this analysis in order to capture any differences across borrower types that could impact baseline utilization rates. Strategic behavior would be suggested by a borrower having relatively lower utilization in case of default. More precisely, we estimate probit regressions where the probability that borrower *i* has high utilization (i.e., a ratio of aggregate bankcard balances to credit limit of 80 percent or higher either two years following origination or at the time of first default) is modeled as $\Pr(High\ Utilization) = \Pr\,(y \leq Y_i)$, where y is normally distributed with mean 0 and variance 1, and

$$Y_i = \beta X_i + \delta D_i + \gamma_{Borrower\ Type_i} + \alpha_{Borrower\ Type_i} D_i \qquad (4)$$

X_i includes mortgage and borrower characteristics at origination; Z_i includes the dynamic variables (updated LTV and the unemployment change); in contrast to the previous analysis, these are now to two years following origination only for non-defaulting borrowers, and to the time of first default for defaulting borrowers. D_i is an indicator variable denoting whether or not the loan is ever delinquent in the two years following origination. As fraudulent investors are the base borrower type, we conjecture that $\alpha>0$ for the other borrower types. That is, the other borrower types are likelier to have high utilization in case of default.[28]

[28] Fourteen percent of our borrowers do not have valid bankcard utilization data and are dropped from the estimation sample for this model.

Table 8a reports the marginal effects for uninteracted covariates. Not surprisingly, borrowers who default have higher utilization, i.e., $\delta>0$. Borrowers who are ex-ante more constrained, such as those with FHA loans, lower FICO scores, or higher LTV at origination, have higher utilization. Borrowers with smaller mortgages also have higher utilization, likely because they are less wealthy. Changes in the economic environment since origination also play a role: Larger changes in unemployment rates and higher updated LTV ratios are also associated with modestly high utilization. Columns (3) and (4) confirm that these results hold for both parts of our sample.

Next, we consider the interaction of borrower type and default from these estimations. These marginal effects are reported in Table 8b. This table reports the marginal effect on utilization of changing from fraudulent investors (the base borrower type) to one of the other borrower types, separately for borrowers who do and do not default. In addition, we also report a chi-squared statistic that tests whether the marginal effects are equal across defaulters and non-defaulters.

We see that, generally, non-defaulting fraudulent investors have utilization rates that are close to the other borrower types; this is also consistent with our summary statistics. In particular, the γ for declared investors is statistically indistinguishable from zero.

The situation is quite different in default, however. We find that α for the other borrower types is significantly above zero. In particular, fraudulent investors who default are 7 percentage points *less* likely to have high bankcard utilization than similar declared investors. And the chi-squared statistic confirms that the relative difference for these two borrower types between the likelihood of high utilization in and out of default is statistically significant. This difference is even more striking for honest homeowners, as the latter would be expected to try harder to avoid default (i.e., they behave less strategically). In column (2) of Table 8b, we show that our results do not change significantly after controlling for multiple first liens and moving.

Columns (3) and (4) again confirm that the relatively more strategic default behavior of fraudulent investors holds both in the housing bubble and bust as well as for more recent originations.

VIII. Alternative Hypotheses

As we discussed in Section VII, measurement error in the bureau variables that we use to identify fraud that is uncorrelated with default outcomes would likely attenuate the difference in default risk between fraudulent investors and the other borrower types. However, one might be more concerned that there are unobserved borrower characteristics leading to a correlation between fraud and default.

In this section we consider two alternative hypotheses through which such a correlation could arise. These hypotheses model what we characterized as occupancy fraud as the ex-post result of negative mortgage-market outcomes. Each hypothesis individually addresses one of the two features that we used to characterize fraud: having multiple first liens and not moving. After investigating them, however, we show that neither of them overturns our key result regarding the higher default risk of fraudulent investors.

Accidental Fraud

The first alternative hypothesis that we consider is that fraud was "accidental," in that these borrowers did not set out to commit occupancy fraud but were unable to sell their original home, for example, because of a real estate market that was worse than anticipated when the transaction commenced. This could explain the presence of multiple liens, and could also be consistent with higher default rates because these borrowers would have an ex-post debt burden that is higher than originally anticipated.

To address this alternative hypothesis, we first construct measures of housing market strength and examine how well they predict the incidence of fraud. First, using the Multiple Listings Service data from CoreLogic Solutions (henceforth MLS), we construct a semiannual measure of the average days on market (DOM) for single-family properties offered for sale in the county associated with the consumer address in the CCP data (prior to the new purchase mortgage origination date).[29] The available MLS data enables us to construct this measure for 1,911 counties, covering roughly 85 percent of our sample. We create buckets for this measure to allow for a nonlinear effect. Our summary statistics, in Table 3, show that fraudulent investors do indeed have a somewhat longer DOM than do honest homeowners (or declared investors). This will be explored further in our multivariate regressions. In addition, another measure of the strength of the housing market that we consider is the change in the house price index in the year *following* mortgage origination; a decline in house prices is identified with a weak market.[30]

We now use this DOM measure as a regressor in a linear probability model for each of the two criteria defining occupancy fraud: having multiple liens four quarters following the CRISM loan origination and not moving around the loan origination. We also consider fraud itself as a dependent variable. We restrict attention to borrowers who are declared owner-occupants in the McDash mortgage data since we are interested in distinguishing between frauds and "accidental" frauds.

[29] In computing the average county DOM, we drop properties listed for rental, apartments, commercial/industrial/business properties, condos, farms, land/plots, mobile homes, multifamily units, duplexes, triplexes, fourplexes, timeshares, townhouses, listings with a negative DOM, and those with a DOM greater than 540.

[30] We choose one year following origination, as this coincides with the period over which fraud is defined.

The results are reported in Table 9a (a full set of covariates is also included but not reported — the coefficients can be found in the online appendix). Column (1) reports the results for a model of multiple liens. In column (2) we add our measures of market strength: DOM and the 1-year ahead HPI change. Both of these enter with the expected sign — very high DOM raises the likelihood of having multiple mortgages, while ex-post HPI increases lower it. However, neither of these has a significant impact on the explanatory power of the model (the coefficients for the other covariates in the model are also not impacted). In columns (3) and (4), we consider the second criterion characterizing fraud — not moving. Once again, our two measures of a weak housing market are associated with an increased likelihood of not moving; however, they do not significantly increase the explanatory power of the model. Finally, columns (5) and (6) again demonstrate that while these two measures of market weakness do enter with the correct signs, they have limited explanatory power for fraud.

In Table 9b, we examine the effect of these measures on our probit models of mortgage default from Table 7a. We report the coefficients only for the borrower types; the others can be found in the online appendix. Column (1) of Table 9b estimates the baseline model for the subsample of borrowers for whom we can compute DOM; we confirm that the coefficients are close to those in column (1) of Table 7a.[31] In column (2), we add the DOM measure and 1-year-forward change in house prices and observe that increased DOM, and decreased prices, are indeed associated with a modestly higher incidence of default; this is likely because in a weaker market, it is more difficult (and possibly less compelling) for troubled borrowers to sell their house instead of defaulting. Nevertheless, the impact of fraud itself is essentially unchanged.[32]

In short, while a slower housing market is modestly associated with a higher incidence of multiple liens and higher default rates, we conclude that these effects are largely independent of fraud.

Default and Moving

The second alternative hypothesis that we examine focuses on the other criterion used to define fraud — not moving around the time of the purchase mortgage origination. More precisely, it considers the possibility that the event of default itself is correlated with behavior that leads the credit bureau to not update the borrower's address. This could occur, for example, if the borrower defaulted soon after originating the mortgage, and then moved back into their original residence before the bureau could update the address. As fraud is identified by

[31] Any small differences are due to the restriction here to counties for which we can compute DOM.
[32] Column (3) repeats this same estimation while including controls for moving and having a single first lien; these results are not significantly different from the corresponding ones from column (2) of Table 7a.

not moving within the year following the mortgage origination date, this could then lead to an association between fraud and default.

To address this concern, we first drop defaults that occur within the first year following mortgage origination (the window we use to identify address changes and thus define fraud). For honest homeowners, fraudulent and declared investors, these constitute 40–48 percent of all defaults; by contrast, only 25 percent of defaults by second homeowners occur in the first year following origination.

The result of default regressions on this sample is reported in the top panel of Table 9c; as can be seen by comparing the results to columns (1), (3) and (4) of Table 7a, this does not qualitatively change our main results — fraud remains associated with significantly higher default rates.[33] In addition, in other analysis (not reported) we dropped defaults that occur in the first two years following origination and considered defaults in years 2–4: Once again the results are qualitatively unaffected.

To confirm the lack of a strong relationship between moving and default, in the lower panel of Table 9c, we also compare the unconditional default rates for investors and second homeowners across moving status, as these two groups include both movers and non-movers.

IX. Robustness

In this section we consider the robustness of our results to various alternative specifications.

Single-Lien Non-Movers

Recall that we define fraudulent investors as those who satisfy three criteria: (i) They are declared owner-occupants, (ii) they have multiple first liens one year following mortgage origination, and (iii) they do not move around the time of the mortgage origination. Declared owner-occupants who move around the mortgage origination are classified as honest homeowners, regardless of how many liens they have. However, there is also a subset of declared owner-occupants whom we cannot classify, as they have only a single lien, but they do not move around the mortgage origination. These make up over 12 percent of declared owner-occupants. In this section, we examine this category of borrowers and study the robustness of our key result to including them.

[33] We do note that the relative impact of fraud is now smaller, particularly for the 2005–2007 vintages. This reflects the fact that, as is well-known, many mortgages that were originated during this time period defaulted very quickly (see Mayer et al., 2009, and Elul, 2016, for example).

One possible explanation for this category is lags in updating addresses in the credit bureau data. Although we showed earlier that the moving rate in the bureau data aligns well with that from other data sets such as the ACS, it may well be that bureaus are slow in updating addresses for some borrowers, so that they appear not to have moved around the date of mortgage origination. Note that this need be the case for only a relatively small share (i.e., 12 percent of owner-occupants) to be consistent with the size of this category. This would also not skew aggregate moving statistics, as these slow-to-be-recorded moves would appear in the moves for later years.[34] Furthermore, the address data itself could be correct, for example, if the borrower is undertaking an extended renovation before moving into their new home. In either case, they should properly be considered as honest homeowners.

Another possible explanation is that they indeed have only a single first lien, and they indeed are not moving around the mortgage origination. However, we have already seen that in both our data and in other data sets such as the SCF, there are many investors who have only a single first lien. Thus, these borrowers may nevertheless be fraudulent investors who do not intend to live in the house as owner-occupants, despite their declaration.

As we could not distinguish between these two possible explanations, we did not include in our main analysis this group of non-movers with a single first lien. However, in what follows, we examine our key result when we include this category of borrowers. Our findings suggest that this category may consist of a mixture of honest homeowners and fraudulent investors. Examining some key summary statistics in Table 10, we see that the characteristics of these borrowers lie along several dimensions between those of honest homeowners and fraudulent investors. These include the share in bubble states, origination LTV, and their bankcard utilization rate in case of default. Most notably, their delinquency rate lies between the two categories. In other ways, however, they appear to be much closer to honest homeowners. In particular, they do not have the larger loan amounts that characterize fraudulent investors, and have lower average FICO scores at origination than do fraudulent investors.

In columns (1) and (2) of Table 11, we report the results of default regressions corresponding to columns (1) and (2) of Table 7a. The relative default risk for these single mortgage non-movers is modestly higher than that for honest homeowners, although they are still significantly less risky than fraudulent investors. Again, this is consistent with our hypothesis that the single-lien non-mover group is a mixture of fraudulent investors and

[34] It is also possible that these borrowers instead actually have other first liens that are not recorded by the bureaus, and for this reason, they are actually fraudulent investors. But this seems less likely, because we have already shown that the share of multiple liens in the CCP data aligns well with other data sets, and this other (unrecorded) first lien would have already been present before the origination of the purchase mortgage in question.

honest homeowners; lacking a principled way to classify this group, we therefore feel the most prudent approach is to exclude them from our main analysis.

Fraud and Second Homeowners

In an earlier version of this paper, we included an additional criterion to characterize fraud, namely that the distance between the centroids of the property (McDash) zip code and the borrower zip code in CCP be no more than 75 miles apart. This was designed to more sharply distinguish fraudulent investors from second homeowners. We did not include this criterion in the construction in the body of this paper, as the share of fraudulent investors with zip distance greater than 75 miles is very similar to that for declared investors, and very different to that for second homeowners, suggesting that our simpler classification was accurate. To confirm this decision, in column (3) of Table 11 we estimate our baseline default model while dropping fraudulent investors who have a zip distance greater than 75 miles and show that our results are essentially unaffected.

Multiple First Liens

We imposed the requirement that fraudulent investors have multiple first liens, in order to ensure that they are indeed investors rather than homeowners. At the same time, there may as a result be other differences that are correlated with this characteristic. While we do include a control for having a single first lien in many of the models we estimate in the body of the paper, we show in column (4) of Table 11 that our key default result is essentially unaffected if we restrict the entire sample to borrowers with multiple first liens — this column should be compared with column (2) of Table 7a. We also re-estimated the models found in other tables with this restriction and confirmed little impact. The results are not reported.

X. Conclusion

We identify widespread occupancy fraud in residential mortgages, both during the housing bubble and also in recent origination vintages. In contrast to previous studies, we are also able to show that occupancy fraud was common in the GSE market and in loans held in portfolio, not just in private MBS. We find that mortgage borrowers who misrepresented their occupancy status performed worse than otherwise similar declared investors. Their default decisions are also more strategic than other borrower types. Our results are economically significant and suggest that such fraud may also pose a risk in future boom-bust cycles. One area for future research is understanding whether behavioral characteristics may help explain why some borrowers were led to speculate in housing markets through fraud.

References

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

Beraja, Martin, Andreas Fuster, Erik Hurst, and Joseph Vavra. “Regional Heterogeneity and the Refinancing Channel of Monetary Policy.” *Quarterly Journal of Economics*, 134(1), pp. 109–182, 2019.

Bhutta, Neil, Jane Dokko, and Hui Shan. Consumer Ruthlessness and Mortgage Default during the 2007 to 2009 Housing Bust. *Journal of Finance*, 72(6), pp. 2433–2466, 2017.

DeWaard, Jack, Janna E. Johnson, and Stephan D. Whitaker. “Internal Migration in the United States: A Comparative Assessment of the Utility of the Consumer Credit Panel.” Federal Reserve Bank of Cleveland Working Paper 1804R, 2018.

Dettling, Lisa J., Sebastian J. Devlin-Foltz, Jacob Krimmel, Sarah J. Pack, and Jeffrey P. Thompson. Comparing Micro and Macro Sources for Household Accounts in the United States: Evidence from the Survey of Consumer Finances. Federal Reserve Board Finance and Economics Discussion Series Working Paper 2015-086, 2015.

Elul, Ronel, Nicholas S. Souleles, Souphala Chomsisengphet, Dennis Glennon, and Robert Hunt. “What ‘Triggers’ Mortgage Default?” *American Economic Review*, 100(2), pp. 490–4, 2010.

Elul, Ronel. “Securitization and Mortgage Default,” *Journal of Financial Services Research*, 49(3), pp 281–309, 2016.

Garmaise, Mark J. “Borrower Misreporting and Loan Performance.” *Journal of Finance*, 70(1), pp. 449–84, 2015.

Griffin, John M., and Gonzalo Maturana. “Who Facilitated Misreporting in Securitized Loans?” *Review of Financial Studies*, 29(2), pp. 384–419, 2016.

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.

Lee, Donghoon, and Wilbert van der Klaauw. “An Introduction to the FRBNY Consumer Credit Panel.” Federal Reserve Bank of New York Staff Report 479, November 2010.

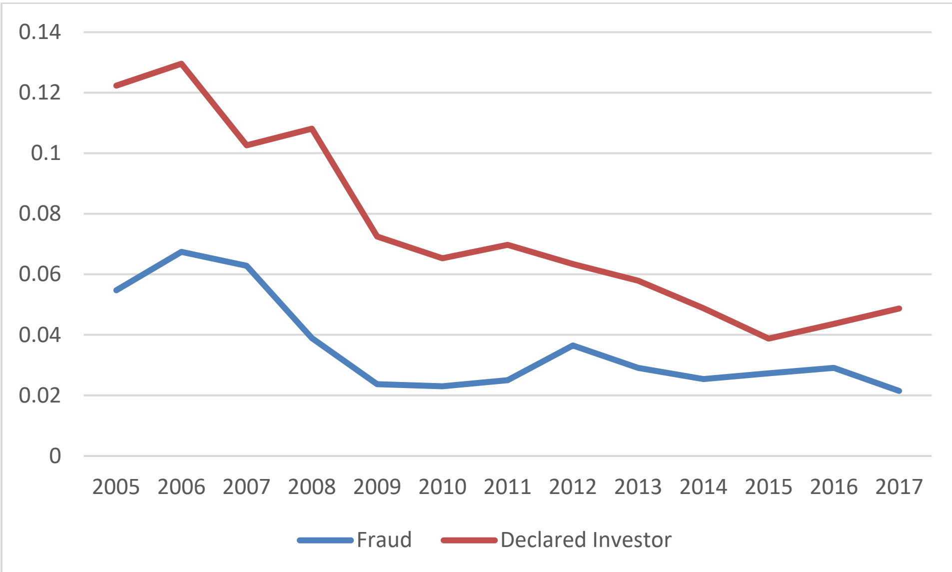
Mayer, Christopher, Karen Pence, and Shane M. Sherlund. “The Rise in Mortgage Defaults,” *Journal of Economic Perspectives*, 23(1), pp. 27–50, 2009.

Mian, Atif, and Amir Sufi. “Fraudulent Income Overstatement on Mortgage Applications During the Credit Expansion of 2002 to 2005.” NBER Working Paper No. 20947, 2015.

Piskorski, Tomasz, Amit Seru, and James Witkin. “Asset Quality Misrepresentation by Financial Intermediaries: Evidence from the RBMS Market.” *Journal of Finance*, 70(6), pp. 2635–78, 2015.

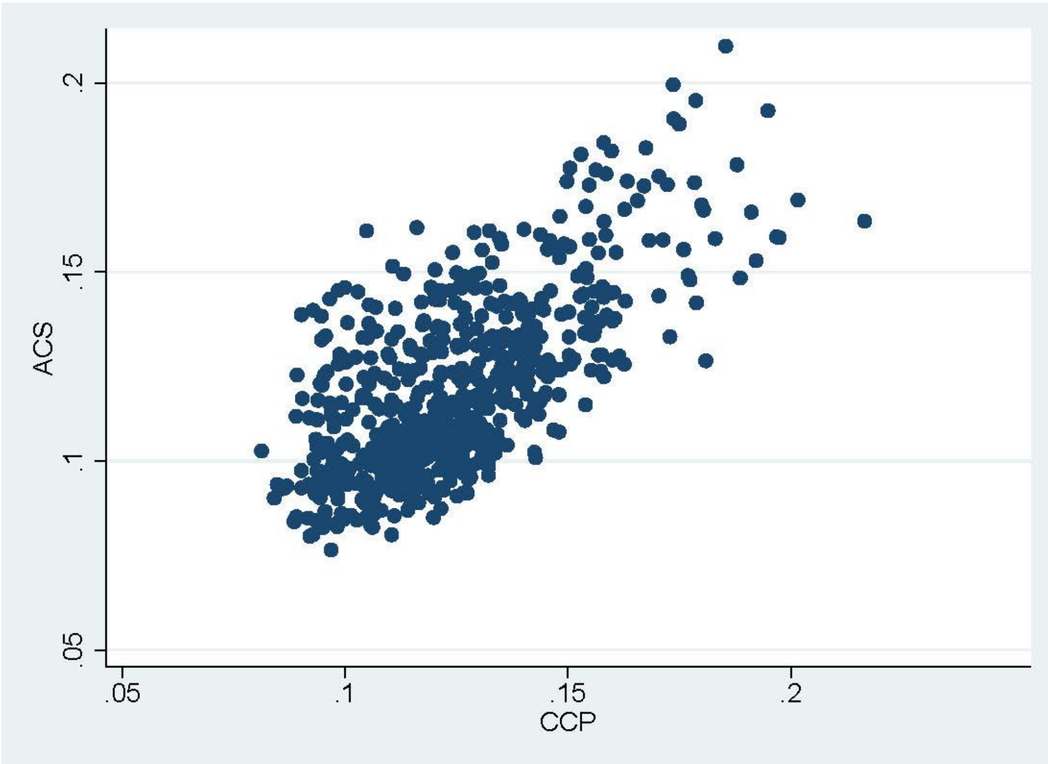
Yavas, Abdullah, and Shuang Zhu. “Misreporting of Second Liens in Portfolio Mortgages and Privately Securitized Mortgages.” Mimeo, 2019.

Figure 1: Investor Shares over Time



Note: Borrower types are as defined in Section IV.
Source: Authors' calculations of McDash, CCP, and CRISM data.

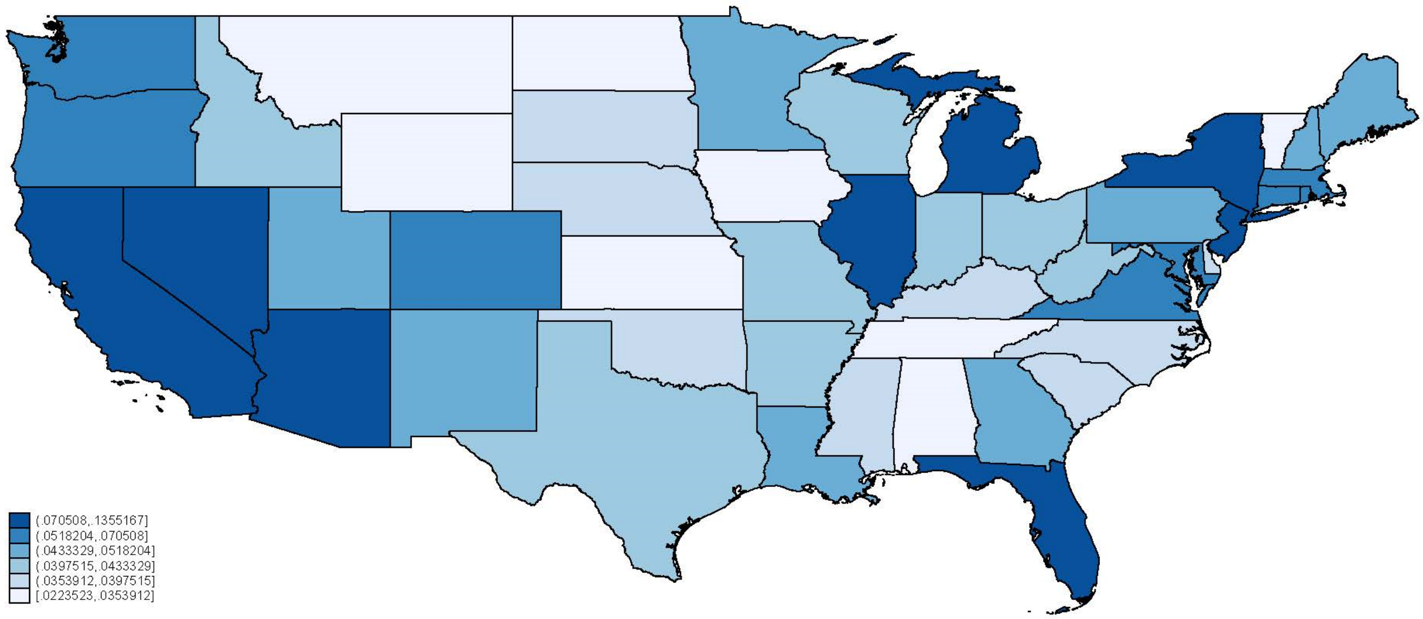
Figure 2: State-Year Moving Rates — CCP vs. ACS



Note: This figure plots the state-year moving rates from the CCP against those from the American Community Survey, where moving in the CCP is defined as a change in scrambled address from December of the previous year to December of the current year.
Source: Authors' calculations of CCP and ACS data.

Figure 3a: Geography of Occupancy Fraud — 2005 to 2007

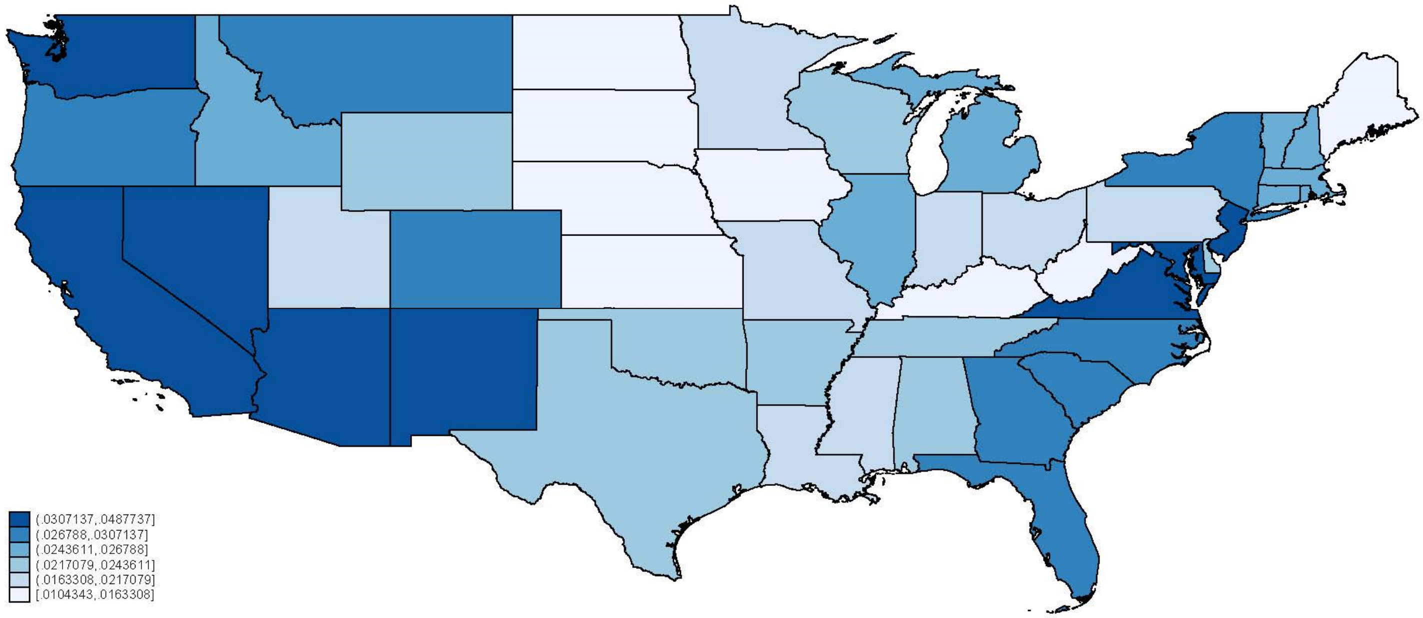
State-level mortgage occupancy fraud rate as a share of purchase mortgage originations: 2005–2007.



Source: Authors' calculations of McDash, CCP, and CRISM data. Alaska (0.06) and Hawaii (0.14) not shown.

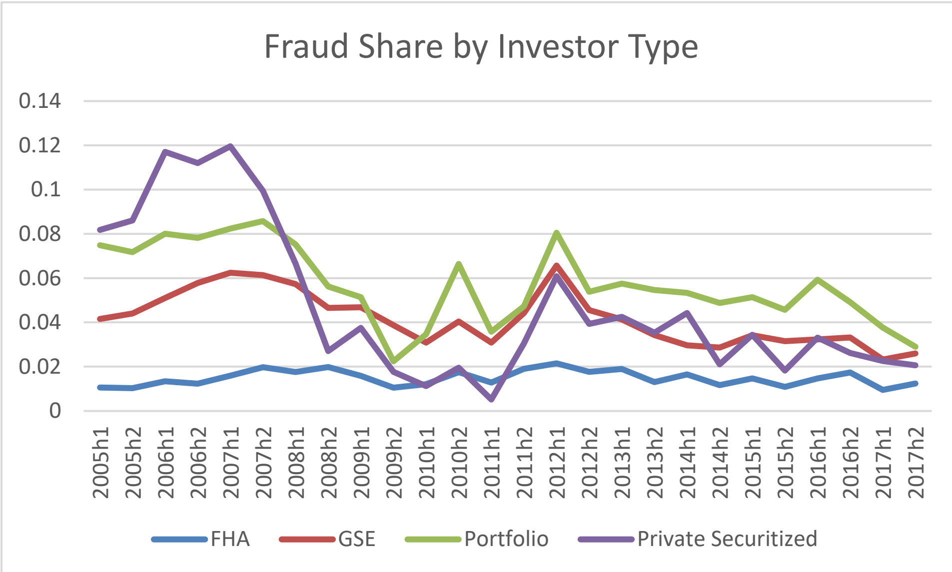
Figure 3b: Geography of Occupancy Fraud — 2008 to 2017

State-level mortgage occupancy fraud rate as a share of purchase mortgage originations: 2008–2017



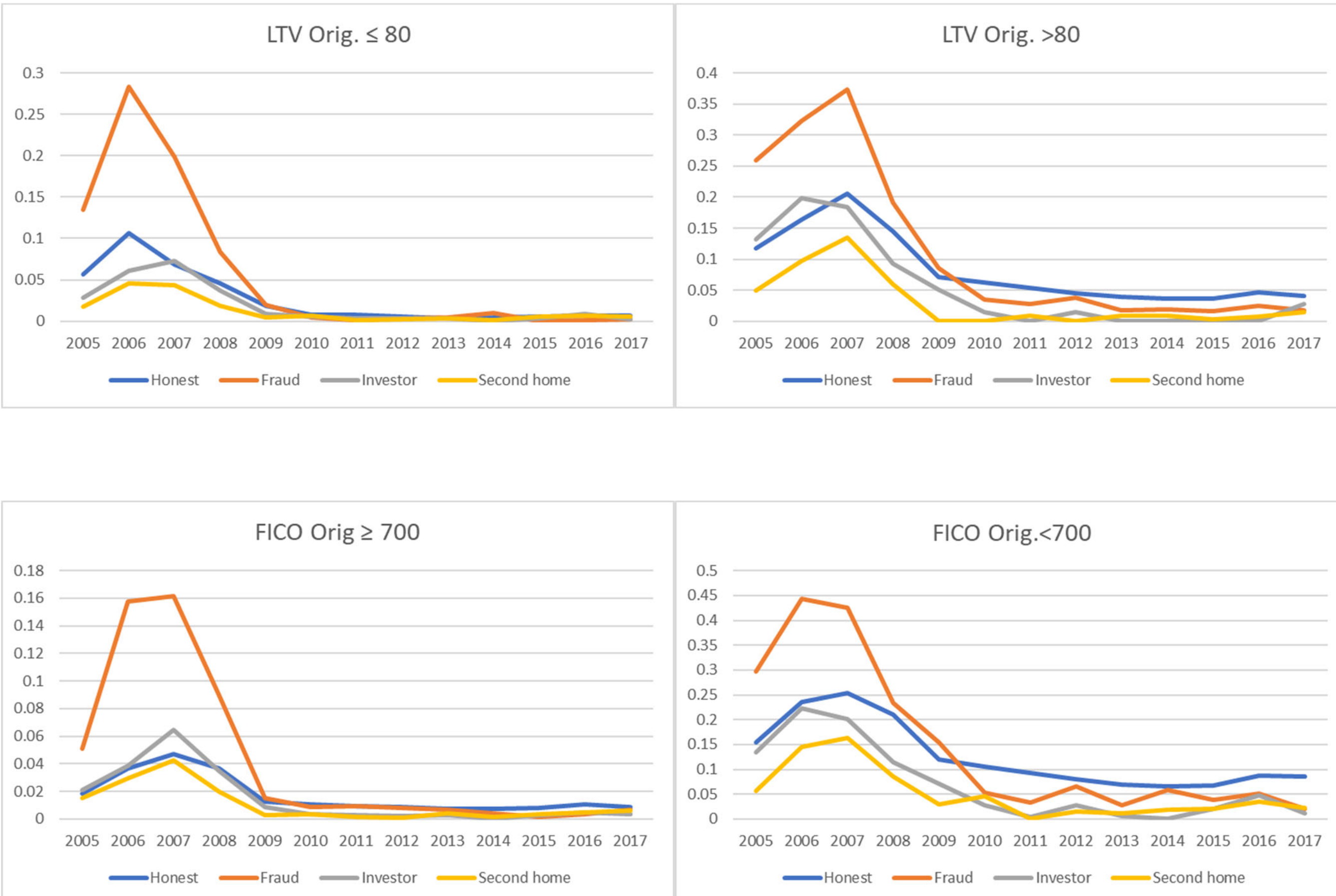
Source: Authors' calculations of McDash, CCP, and CRISM data. Alaska (0.04) and Hawaii (0.05) not shown.

Figure 4: Fraud Share by Origination Vintage and Investor Type



Source: Authors' calculations of McDash, CCP, and CRISM data.

Figure 5: Two-Year Default Rate by Origination Year and Risk Characteristics



Source: Authors' calculations of McDash, CCP, and CRISM data.

Table 1: Descriptions of Variables

Variable	Description
Borrower Type	Honest homeowner, fraudulent investor, declared investor, or second home
Default	60+ days delinquent in any month within two years of origination date, McDash data
Bubble State	McDash property state is California, Nevada, Arizona, or Florida
FICO (Origination)	McDash origination FICO score
LTV Ratio (Origination)	LTV ratio of CRISM mortgage at origination
% Change HPI: Two years following origination	Percentage change in the property's zip code-level CoreLogic house price index over the two years following the mortgage origination date; if zip code level is not available, the county level is used, and if this is also unavailable, the state level is used
% Change 2-Year Lagged HPI	Percentage change in the property's zip code-level CoreLogic house price index two years before the McDash loan origination date; if zip code level is not available, county level is used, and if this is also unavailable, the state level is used
Second Lien	Borrowers have a second lien (HELOC or closed-end home equity loan) in CCP four quarters after CRISM mortgage origination
Interest Rate (Origination)	Interest rate observed when mortgage first enters the McDash data
Investor Type	McDash-reported investor type six months following origination: FHA/VA, GSE (FNMA/FHLMC), Private Securitized, or Portfolio.
Interest Rate Type	Fixed Rate vs. ARM; for ARMs, loans have either 1-year, 2-year, 3-year, 5-year, 7-year, or 10-year introductory fixed periods
Bankcard Utilization ≥80%	1 if bankcard utilization is greater than or equal to 0.80 as of December 2008
Bankcard Utilization (No Default)	Total bankcard balance/Total bankcard limit (for bankcards with an update in the previous 3 months) two years following mortgage origination (for mortgages that do not default within two years). From CCP.
Bankcard Utilization (Default)	Bankcard utilization at quarter of first default (for mortgages that default within two years of origination) From CCP.
Updated LTV Ratio (two years following Origination)	Origination amount/(LTV at origination× [1+ zip code-level HPI change over two years following origination])
Multiple First Liens	More than one first-lien mortgage in CCP four quarters following the CRISM mortgage origination date
Unemployment Rate at Close Date	Property's zip code-level unemployment rate at origination (BLS)
Change Unemployment (two years following origination)	Percentage change in the property's county-level unemployment rate over two years following origination (BLS)
Mortgage Term	Years until mortgage maturity (at origination): 15/20 years, 30 years, or 40 years
Days on Market	Average days-on-market (DOM) for single-family properties offered for sale in the county associated with the CCP address. From CoreLogic

Source: Variables based on authors' calculations of McDash, CCP, CRISM, CoreLogic, and BLS data.

Table 2a: Occupancy Type — Moving and Multiple Mortgages (2005–2017 Originations)

	Moved w/in 4 quarters of purchase mortgage origination	Multiple first liens 4 quarters following purchase mortgage origination
Declared Owner-Occupant	0.96	0.17
Declared Investor	0.38	0.42
Declared Second Homeowner	0.44	0.48

Source: Authors' calculations of McDash and CCP data.

Table 2b: Comparison of McDash and Paper Sample

	Full McDash: 2005-07	Paper: 2005-07	Full McDash: 2008-17	Paper: 2008-17
Number loans	4,768,672	179,272	10,828,309	405,227
60+ days past due w/in 2yr	0.13	0.11	0.04	0.04
Bubble state	0.21	0.19	0.22	0.22
Declared owner-occupied	0.90	0.88	0.95	0.94
Declared investor	0.10	0.12	0.05	0.06
Second home	0.04	0.05	0.03	0.03
FICO at orig < 660	0.26	0.24	0.16	0.15
FICO at orig in [660,700)	0.18	0.17	0.19	0.18
FICO at orig in [700,750)	0.25	0.25	0.24	0.24
FICO at orig in [750,800)	0.27	0.28	0.33	0.34
FICO at orig >= 800	0.05	0.05	0.08	0.08
FICO_at orig (avg.)	706	709	726	727
LTV at orig ≤70	0.16	0.16	0.11	0.11
LTV at orig (70,80]	0.48	0.49	0.24	0.25
LTV at orig (80,90]	0.07	0.07	0.10	0.10
LTV at orig > 90	0.29	0.29	0.54	0.54
Updated LTV 2 yrs from orig. < 80	0.38	0.38	0.40	0.39
Updated LTV [80,90)	0.20	0.21	0.27	0.26
Updated LTV [90,100)	0.17	0.17	0.20	0.21
Updated LTV [100,120)	0.18	0.18	0.12	0.13
Updated LTV >= 120	0.07	0.06	0.01	0.01
Loan amt at orig ≤ 200k	0.60	0.60	0.58	0.57
Loan amt at orig (200k,359650]	0.26	0.27	0.28	0.29
Loan amt at orig (359650,417k]	0.05	0.05	0.06	0.06
Loan amt at orig (417k,700k]	0.07	0.06	0.06	0.06
Loan amt at orig > 700k	0.02	0.02	0.02	0.02
Loan amt at orig (avg,$)	222,898	221,161	225,210	227,443
HPI% chg. 2yr from orig	-0.05	-0.05	0.06	0.05
HPI% chg. 2yr prior to orig	0.18	0.17	0.01	0.00
Interest rate at orig (%)	6.51	6.45	4.48	4.52
Brokered	0.19	0.18	0.27	0.29
FRM	0.74	0.78	0.97	0.97
ARM: 1yr intro rt	0.01	0.01	0.00	0.00
ARM: 2yr intro rt	0.07	0.07	0.00	0.00
ARM: 3yr intro rt	0.02	0.02	0.00	0.00
ARM: 5yr intro rt	0.08	0.08	0.01	0.01
ARM: 7yr intro rt	0.02	0.02	0.01	0.01
ARM: 10yr intro rt	0.02	0.02	0.00	0.00
Interest-only loan	0.14	0.13	0.01	0.01
Option ARM	0.06	0.03	0.00	0.00
Low/no documentation	0.19	0.19	0.10	0.11
Unknown documentation	0.42	0.43	0.11	0.07
FHA/VA	0.11	0.12	0.48	0.47
GSE	0.54	0.56	0.44	0.45
Portfolio	0.10	0.09	0.06	0.06
Private securitized	0.25	0.22	0.02	0.02

Source: Summary statistics for first lien purchase originations. Authors' calculations of McDash and CoreLogic data. FICO score at origination is from McDash.

Table 3: Summary Statistics by Borrower Type

	2005–2007 Orig.				2008–2017 Orig.			
	Honest	Fraud	Investor	Second	Honest	Fraud	Investor	Second
Number Loans	137,604	11,029	21,303	9,336	354,218	11,402	25,997	13,610
Share (Count)	0.77	0.06	0.12	0.05	0.87	0.03	0.06	0.03
Share by Origination Dollars	0.78	0.08	0.09	0.06	0.87	0.04	0.05	0.04
Share of Delinq/Defaults – Count (2 yrs ahead)	0.76	0.13	0.09	0.02	0.94	0.02	0.03	0.01
Share of Delinq/Defaults – $ (2 yrs. ahead)	0.74	0.17	0.07	0.03	0.93	0.03	0.03	0.01
Serious Delinquency/Default (2 yrs. Ahead)	0.11	0.24	0.09	0.05	0.05	0.04	0.02	0.01
Moved around Mortgage Origination	1.00	0.00	0.54	0.42	1.00	0.00	0.36	0.36
CCP zip ≤ 75 Miles from McDash zip	0.93	0.64	0.69	0.28	0.97	0.74	0.73	0.35
Multiple First Liens (4Q after Orig.)	0.19	1.00	0.44	0.47	0.10	1.00	0.50	0.39
Bubble State	0.18	0.35	0.17	0.29	0.21	0.32	0.34	0.36
FICO at Orig. <660	0.26	0.23	0.16	0.07	0.17	0.08	0.06	0.02
FICO at Orig. in [660,700)	0.17	0.20	0.17	0.13	0.19	0.13	0.09	0.06
FICO at Orig. in [700,750)	0.24	0.27	0.27	0.27	0.25	0.25	0.21	0.20
FICO at Orig. in [750,800)	0.27	0.25	0.34	0.42	0.32	0.43	0.49	0.52
FICO at Orig. ≥800	0.05	0.04	0.06	0.10	0.07	0.11	0.15	0.21
FICO Score at Orig. (Avg.)	705	706	722	743	723	744	755	766
LTV at Orig. ≤70	0.14	0.15	0.21	0.23	0.09	0.19	0.24	0.28
LTV at Orig. (70,80]	0.47	0.60	0.49	0.57	0.21	0.37	0.59	0.57
LTV at Orig. (80,90)	0.06	0.06	0.09	0.09	0.10	0.10	0.04	0.10
LTV at Orig. ≥90	0.32	0.19	0.21	0.10	0.60	0.35	0.13	0.05
LTV at Orig. (Avg.)	82	79	78	75	88	81	75	74
Updated LTV 2 yrs from orig. <80	0.37	0.34	0.46	0.50	0.35	0.55	0.71	0.78
Updated LTV in [80,90)	0.20	0.20	0.22	0.22	0.28	0.22	0.12	0.14
Updated LTV in [90,100)	0.18	0.16	0.16	0.13	0.22	0.14	0.08	0.05
Updated LTV in [100,120)	0.19	0.19	0.12	0.11	0.14	0.08	0.08	0.02
Updated LTV ≥120	0.07	0.12	0.04	0.04	0.01	0.01	0.01	0.00
Loan amount at Orig. ≤200k	0.59	0.45	0.73	0.56	0.57	0.43	0.69	0.52
Loan amount at Orig. (200k,359650]	0.27	0.32	0.21	0.28	0.29	0.30	0.22	0.27
Loan amount at Orig. (359650,417k]	0.05	0.08	0.03	0.05	0.06	0.13	0.05	0.09
Loan amount at Orig. (417k,700k]	0.06	0.11	0.02	0.07	0.06	0.08	0.03	0.07
Loan amount at Orig. >700k	0.02	0.04	0.01	0.04	0.02	0.05	0.01	0.06
Loan amount at Orig. (avg, $)	223,518	274,454	168,253	244,189	226,544	300,686	177,787	284,320
HPI % Chg.: 2 years from orig.	-0.05	-0.10	-0.03	-0.05	0.05	0.05	0.05	0.06
HPI % Chg.: 2 years prior to Orig.	0.17	0.22	0.18	0.23	0.00	-0.01	-0.03	0.01
Second Lien (4Q after Orig.)	0.31	0.51	0.39	0.41	0.07	0.29	0.25	0.23
Interest Rate at Orig.	6.42	6.73	6.64	6.28	4.49	4.50	5.07	4.34
Brokered	0.18	0.25	0.12	0.14	0.30	0.24	0.20	0.15
FRM	0.78	0.61	0.84	0.75	0.97	0.94	0.96	0.89
ARM: 1-year intro rate	0.01	0.02	0.01	0.03	0.00	0.00	0.00	0.01
ARM: 2-year intro rate	0.07	0.14	0.04	0.02	0.00	0.00	0.00	0.00
ARM: 3-year intro rate	0.02	0.04	0.02	0.02	0.00	0.00	0.01	0.00
ARM: 5-year intro rate	0.08	0.14	0.07	0.12	0.01	0.03	0.01	0.04
ARM: 7-year intro rate	0.02	0.02	0.01	0.03	0.01	0.02	0.01	0.04
ARM: 10-year intro rate	0.02	0.03	0.01	0.04	0.00	0.01	0.00	0.02
Interest-Only Loan	0.13	0.25	0.09	0.17	0.00	0.02	0.01	0.03
Option ARM	0.03	0.08	0.04	0.01	0.00	0.00	0.00	0.00
Low/No Documentation	0.20	0.24	0.13	0.17	0.10	0.12	0.17	0.14
Unknown Documentation	0.39	0.41	0.61	0.49	0.07	0.10	0.06	0.10
FHA/VA	0.13	0.03	0.16	0.00	0.52	0.26	0.12	0.00
GSE	0.56	0.49	0.60	0.71	0.40	0.62	0.81	0.83
Portfolio	0.10	0.12	0.05	0.10	0.06	0.11	0.05	0.13
Private Securitized	0.22	0.36	0.20	0.20	0.02	0.02	0.01	0.04
Card Util. ≥80% (not in default w/in 2 yrs)	0.22	0.15	0.19	0.16	0.22	0.13	0.13	0.11
Card Util. ≥80% (in default w/in 2 yrs.)	0.56	0.34	0.46	0.39	0.69	0.34	0.48	0.29
Days on Market (County Average)	162	186	161	176	167	157	160	154

Source: Authors' calculations of McDash, CCP, CRISM, and CoreLogic data. FICO score at origination is from McDash.

Table 4: Determinants of Fraud vs. Declared Investor

	(1)	(2)	(3)
FICO Orig. in [660,700)	-0.079***	-0.102***	-0.069***
	(0.007)	(0.004)	(0.007)
[700,750)	-0.133***	-0.142***	-0.120***
	(0.006)	(0.004)	(0.006)
[750,800)	-0.192***	-0.180***	-0.180***
	(0.006)	(0.004)	(0.006)
≥800	-0.198***	-0.171***	-0.188***
	(0.008)	(0.005)	(0.008)
LTV Orig. in (70,80]	0.038***	0.018***	0.032***
	(0.004)	(0.003)	(0.004)
(80,90)	0.197***	0.070***	0.189***
	(0.007)	(0.005)	(0.007)
≥90	0.310***	0.159***	0.301***
	(0.006)	(0.004)	(0.006)
Orig amt. in (200k,359650]	0.138***	0.115***	0.151***
	(0.004)	(0.003)	(0.004)
(359650,417k]	0.275***	0.189***	0.302***
	(0.008)	(0.004)	(0.007)
(417k,700k]	0.297***	0.182***	0.317***
	(0.009)	(0.005)	(0.009)
>700k	0.340***	0.197***	0.392***
	(0.015)	(0.007)	(0.014)
Lagged 2yr HPI Change	0.026**	0.008	-0.157***
	(0.012)	(0.009)	(0.010)
Interest Rate	-0.138***	-0.096***	-0.087***
	(0.002)	(0.002)	(0.002)
Has Second Lien	0.046***	-0.033***	0.047***
	(0.003)	(0.002)	(0.004)
FHA	-0.133***	0.155***	-0.122***
	(0.005)	(0.006)	(0.005)
Portfolio	0.054***	0.035***	0.042***
	(0.007)	(0.005)	(0.007)
Private Securitized	0.109***	0.046***	0.065***
	(0.007)	(0.004)	(0.006)
Term is 15/20 years	-0.081***	-0.062***	-0.065***
	(0.006)	(0.004)	(0.006)
Term is 40 years	0.276***	0.100***	0.265***
	(0.019)	(0.010)	(0.020)
ARM	0.008	-0.000	-0.001
	(0.006)	(0.004)	(0.006)
Bubble State			0.014***
			(0.004)
Bubble Year			0.185***
			(0.006)
Bubble State×Years			0.233***
Controls for Single 1st & Moving	N	Y	N
State and Orig. Vintage Fixed Effects	Y	Y	N
Additional Covariates	Y	Y	Y
Observations	69731	69731	69731

Note: Marginal effects from probit models for the probability that a declared investor or fraudulent investor is fraudulent. "Bubble Years" denotes 2005–2007 originations. Other covariates are not reported and are available in the online appendix. Standard errors are in parentheses (clustered at the county level); *** p<0.01, ** p<0.05, * p<0.1.
Source: Authors' calculations of McDash, CCP, CRISM, CoreLogic, and BLS data.

Table 5: Fraud and Interest Rates

	2005–2017 Orig. (1)	2005–2017 Orig. (2)
Owner-Occupant	-0.094***	0.092***
	(0.003)	(0.005)
Investor	0.263***	0.351***
	(0.004)	(0.004)
Second Home	-0.065***	0.019***
	(0.004)	(0.005)
Have Second Lien	0.030***	0.019***
	(0.002)	(0.002)
FICO Orig.∈[660,700)	-0.168***	-0.169***
	(0.002)	(0.002)
FICO Orig.∈[700,750)	-0.262***	-0.264***
	(0.002)	(0.002)
FICO Orig.∈[750,800)	-0.349***	-0.351***
	(0.002)	(0.002)
FICO Orig. ≥800	-0.369***	-0.371***
	(0.003)	(0.003)
LTV Orig.∈(70,80]	0.064***	0.062***
	(0.002)	(0.002)
LTV Orig.∈(80,90)	0.214***	0.213***
	(0.003)	(0.003)
LTV Orig. ≥90	0.227***	0.227***
	(0.002)	(0.002)
Orig. Amt.∈(200K,359650]	-0.130***	-0.129***
	(0.001)	(0.001)
Orig. Amt.∈(359,650,417K]	-0.162***	-0.162***
	(0.003)	(0.003)
Orig. Amt.∈(417K,700K]	-0.112***	-0.113***
	(0.003)	(0.003)
Orig. Amt.>700K	-0.187***	-0.189***
	(0.005)	(0.005)
Moved		-0.155***
		(0.004)
Single First Lien		-0.042***
		(0.002)
State and Orig. Vintage Fixed Effects	Y	Y
Additional Covariates	Y	Y
Observations	584499	584499

Note: OLS regression models for the interest rate at the time of origination (or when first available). All specifications include origination half-year and state fixed effects; other covariates are not reported and are available in the online appendix. Standard errors (clustered at the county level) are in parentheses; *** p<0.01, ** p<0.05, * p<0.1.
Source: Authors' calculations of McDash, CCP, CRISM, CoreLogic, and BLS data.

Table 6: Interest Rates — Borrower Type Interactions

	2005–2017 Origs.			
	(1)		(2)	
Honest Homeowner				
FICO Orig. <660	-0.198***	(0.009)	-0.062***	(0.010)
[660,700)	-0.104***	(0.008)	0.031***	(0.009)
[700,750)	-0.074***	(0.007)	0.062***	(0.008)
[750,800)	-0.028***	(0.006)	0.108***	(0.007)
≥800	-0.023*	(0.012)	0.114***	(0.013)
LTV Orig.≤70	-0.041***	(0.009)	0.095***	(0.010)
(70,80]	-0.019***	(0.006)	0.115***	(0.007)
(80,90)	-0.116***	(0.011)	0.021*	(0.012)
≥90	-0.132***	(0.006)	0.004	(0.007)
Orig. Amt.≤200k	-0.112***	(0.005)	0.025***	(0.006)
(200k,359650]	-0.055***	(0.006)	0.079***	(0.007)
(359650,417k]	-0.054***	(0.010)	0.078***	(0.011)
(417k,700k]	-0.005	(0.012)	0.129***	(0.012)
>700k	0.051***	(0.018)	0.183***	(0.018)
FHA	0.039***	(0.009)	0.176***	(0.010)
GSE	-0.107***	(0.005)	0.029***	(0.006)
Portfolio	-0.366***	(0.011)	-0.231***	(0.012)
Private Securitized	-0.240***	(0.009)	-0.108***	(0.010)
Declared Investor				
FICO Orig.<660	0.112***	(0.011)	0.199***	(0.011)
[660,700)	0.258***	(0.010)	0.341***	(0.010)
[700,750)	0.247***	(0.008)	0.327***	(0.009)
[750,800)	0.264***	(0.007)	0.340***	(0.008)
≥800	0.274***	(0.014)	0.349***	(0.014)
LTV Orig.≤70	0.084***	(0.011)	0.169***	(0.011)
(70,80]	0.149***	(0.008)	0.232***	(0.008)
(80,90)	0.307***	(0.014)	0.380***	(0.015)
≥90	0.316***	(0.009)	0.394***	(0.010)
Orig. Amt.≤200k	0.258***	(0.006)	0.333***	(0.006)
(200k,359650]	0.208***	(0.008)	0.296***	(0.008)
(359650,417k]	0.138***	(0.015)	0.231***	(0.015)
(417k,700k]	0.234***	(0.018)	0.314***	(0.018)
>700k	0.099***	(0.032)	0.175***	(0.032)
FHA	-0.036***	(0.012)	0.088***	(0.012)
GSE	0.377***	(0.007)	0.434***	(0.007)
Portfolio	0.348***	(0.015)	0.394***	(0.015)
Private Securitized	0.468***	(0.011)	0.518***	(0.012)
Second Home				
FICO Orig.<660	-0.076***	(0.019)	-0.008	(0.019)
[660,700)	-0.071***	(0.014)	-0.005	(0.014)
[700,750)	-0.101***	(0.011)	-0.040***	(0.011)
[750,800)	-0.071***	(0.009)	-0.012	(0.009)
≥800	-0.062***	(0.015)	-0.002	(0.015)
LTV Orig.≤70	-0.032***	(0.011)	0.033***	(0.011)
(70,80]	-0.007	(0.007)	0.056***	(0.007)
(80,90)	-0.118***	(0.015)	-0.059***	(0.015)
≥90	-0.133***	(0.014)	-0.071***	(0.014)
Orig. Amt.≤200k	-0.094***	(0.008)	-0.030***	(0.009)
(200k,359650]	-0.027***	(0.010)	0.035***	(0.010)
(359650,417k]	-0.063***	(0.016)	-0.002	(0.016)
(417k,700k]	-0.146***	(0.019)	-0.092***	(0.019)
>700k	-0.184***	(0.025)	-0.138***	(0.025)
FHA	-0.069***	(0.008)	-0.006	(0.008)
GSE	-0.069***	(0.008)	-0.006	(0.008)
Portfolio	-0.155***	(0.016)	-0.093***	(0.016)
Private Securitized	-0.112***	(0.014)	-0.052***	(0.014)
Single 1st & Moving Controls	N		Y	
State and Vintage F.E.	Y		Y	
Additional Covariates	Y		Y	
Observations	584499		584499	

Note: This table reports the results of probit regressions of origination interest rate, interacting selected risk characteristics with borrower type. The table shows marginal effects relative to the base borrower type (fraudulent investor). Columns (2) and (4) include controls for moving and multiple mortgages. All other covariates are as in Table 5 and are reported in the online appendix. Standard errors are in parentheses to the right of the marginals (clustered at the county level); *** p<0.01, ** p<0.05, * p<0.1.
Source: Authors' calculations of McDash, CCP, CRISM, CoreLogic, and BLS data.

Table 7a: Fraud and Mortgage Default

	(1) 2005–17	(2) 2005–17	(3) 2005–07	(4) 2008–17
Owner-occupant	-0.035***	-0.029***	-0.064***	-0.009***
	(0.002)	(0.003)	(0.003)	(0.002)
Investor	-0.043***	-0.037***	-0.061***	-0.023***
	(0.002)	(0.003)	(0.004)	(0.003)
Second home	-0.044***	-0.038***	-0.074***	-0.016***
	(0.003)	(0.003)	(0.005)	(0.003)
Have Second Lien	0.008***	0.007***	0.012***	0.006***
	(0.001)	(0.001)	(0.002)	(0.001)
Orig. Int. Rate	0.028***	0.028***	0.043***	0.022***
	(0.001)	(0.001)	(0.001)	(0.001)
FICO Orig. in [660,700)	-0.070***	-0.071***	-0.093***	-0.060***
	(0.002)	(0.002)	(0.003)	(0.002)
FICO Orig. in [700,750)	-0.101***	-0.101***	-0.132***	-0.085***
	(0.002)	(0.002)	(0.003)	(0.001)
FICO Orig. in [750,800)	-0.122***	-0.123***	-0.170***	-0.099***
	(0.002)	(0.001)	(0.003)	(0.001)
FICO Orig. ≥800	-0.127***	-0.128***	-0.180***	-0.102***
	(0.002)	(0.002)	(0.003)	(0.002)
Orig. Amt. in (200k,359650]	0.001	-0.000	0.008***	-0.001
	(0.001)	(0.001)	(0.002)	(0.001)
Orig. Amt. in (359650,417k]	0.008***	0.006**	0.021***	0.002
	(0.003)	(0.003)	(0.005)	(0.003)
Orig. Amt. in (417k,700k]	-0.003*	-0.005***	0.009**	-0.011***
	(0.002)	(0.002)	(0.003)	(0.002)
Orig. Amt. >700k	-0.019***	-0.020***	-0.018***	-0.020***
	(0.003)	(0.003)	(0.006)	(0.003)
LTV Orig. in (70,80]	0.008***	0.007***	0.018***	0.000
	(0.002)	(0.002)	(0.003)	(0.002)
LTV Orig. in (80,90)	0.010***	0.011***	0.011***	0.008***
	(0.002)	(0.002)	(0.004)	(0.002)
LTV Orig. ≥90	0.007***	0.008***	0.013***	0.006***
	(0.002)	(0.002)	(0.004)	(0.002)
Updated LTV in [80, 90)	0.014***	0.013***	0.018***	0.009***
	(0.001)	(0.001)	(0.002)	(0.001)
Updated LTV in [90,100)	0.025***	0.024***	0.041***	0.014***
	(0.001)	(0.001)	(0.003)	(0.001)
Updated LTV in [100,120)	0.037***	0.036***	0.069***	0.018***
	(0.002)	(0.002)	(0.004)	(0.002)
Updated LTV ≥120	0.089***	0.089***	0.149***	0.042***
	(0.005)	(0.005)	(0.007)	(0.004)
Unemp Bucket 1	0.002*	0.002*	0.005	-0.000
	(0.001)	(0.001)	(0.003)	(0.001)
Unemp Bucket 2	0.006***	0.005***	0.009***	-0.000
	(0.002)	(0.002)	(0.004)	(0.002)
Unemp. Bucket 3	0.015***	0.015***	0.024***	0.009***
	(0.002)	(0.002)	(0.004)	(0.003)
FHA/VA	0.019***	0.020***	0.018***	0.017***
	(0.001)	(0.001)	(0.003)	(0.001)
Portfolio	0.016***	0.016***	0.022***	0.014***
	(0.002)	(0.002)	(0.003)	(0.002)
Private Securitized	0.014***	0.014***	0.015***	0.014***
	(0.001)	(0.001)	(0.003)	(0.002)
Single First Lien		-0.012***		
		(0.001)		
Moved		0.004**		
		(0.002)		
State and Vintage F.E.	Y	Y	Y	Y
Additional Covariates	Y	Y	Y	Y
Observations	584499	584499	179272	405227

Notes: Marginal effects from probit models of mortgage default 2-yrs ahead. All models include origination half-year and state fixed effects. Unemployment buckets are: 2005–07 originations: ≤1.4%, 1.5–2.1%, 2.2–3.3%, ≥3.4%; 2008–17 originations: ≤0.5%, 0.6%–2.1%, 2.2–3.3%, ≥3.4%. Standard errors are in parentheses (clustered at the county level); *** p<0.01, ** p<0.05, * p<0.1.
Source: Authors' calculations of McDash, CCP, CRISM, CoreLogic, and BLS.

Table 7b. Marginal Effects on Default: Updated LTV, Origination FICO Score, and Unemployment Change

	2005–2017 (1)		2005–2007 (2)		2008–2017 (3)	
Honest Homeowner						
FICO Orig. <660	-0.038***	(0.006)	-0.075***	(0.008)	0.024***	(0.007)
in [660,700)	-0.041***	(0.005)	-0.082***	(0.008)	-0.004	(0.005)
in [700,750)	-0.035***	(0.003)	-0.063***	(0.006)	-0.013***	(0.003)
in [750,800)	-0.020***	(0.002)	-0.045***	(0.005)	-0.007***	(0.002)
≥800	-0.012***	(0.004)	-0.030***	(0.010)	-0.003	(0.003)
Updated LTV <80	-0.017***	(0.003)	-0.041***	(0.005)	0.002	(0.003)
in [80, 90)	-0.022***	(0.004)	-0.054***	(0.007)	0.003	(0.004)
in [90,100)	-0.026***	(0.004)	-0.054***	(0.008)	0.001	(0.005)
in [100,120)	-0.056***	(0.006)	-0.092***	(0.009)	-0.005	(0.006)
≥120	-0.095***	(0.011)	-0.115***	(0.014)	-0.055***	(0.021)
Unemp. Bucket 1	-0.026***	(0.003)	-0.057***	(0.005)	0.006*	(0.003)
Bucket 2	-0.028***	(0.005)	-0.064***	(0.011)	0.001	(0.005)
Bucket 3	-0.047***	(0.006)	-0.073***	(0.009)	-0.022**	(0.010)
Bucket 4	-0.041***	(0.004)	-0.064***	(0.006)	-0.024***	(0.006)
Declared Investor						
FICO Orig. <660	-0.072***	(0.007)	-0.089***	(0.010)	-0.039***	(0.009)
in [660,700)	-0.052***	(0.005)	-0.081***	(0.009)	-0.027***	(0.006)
in [700,750)	-0.035***	(0.004)	-0.058***	(0.007)	-0.018***	(0.004)
in [750,800)	-0.018***	(0.003)	-0.038***	(0.006)	-0.007***	(0.002)
≥800	-0.008*	(0.005)	-0.016	(0.012)	-0.003	(0.004)
Updated LTV <80	-0.028***	(0.003)	-0.043***	(0.006)	-0.015***	(0.004)
in [80, 90)	-0.032***	(0.004)	-0.052***	(0.008)	-0.019***	(0.005)
in [90,100)	-0.028***	(0.005)	-0.046***	(0.009)	-0.013**	(0.006)
in [100,120)	-0.067***	(0.007)	-0.089***	(0.011)	-0.031***	(0.006)
≥120	-0.118***	(0.014)	-0.144***	(0.020)	-0.082***	(0.022)
Unemp. Bucket 1	-0.034***	(0.003)	-0.056***	(0.006)	-0.015***	(0.004)
Bucket 2	-0.041***	(0.006)	-0.052***	(0.013)	-0.020***	(0.006)
Bucket 3	-0.061***	(0.007)	-0.087***	(0.011)	-0.036***	(0.010)
Bucket 4	-0.051***	(0.005)	-0.067***	(0.008)	-0.039***	(0.007)
Second Home						
FICO Orig. <660	-0.096***	(0.010)	-0.143***	(0.014)	-0.019	(0.018)
in [660,700)	-0.056***	(0.006)	-0.094***	(0.010)	-0.016*	(0.009)
in [700,750)	-0.033***	(0.004)	-0.058***	(0.008)	-0.012**	(0.005)
in [750,800)	-0.021***	(0.003)	-0.046***	(0.006)	-0.007**	(0.003)
≥800	-0.012***	(0.004)	-0.038***	(0.011)	-0.000	(0.004)
Updated LTV <80	-0.030***	(0.004)	-0.055***	(0.007)	-0.010**	(0.004)
in [80, 90)	-0.040***	(0.005)	-0.067***	(0.010)	-0.017***	(0.007)
in [90,100)	-0.049***	(0.006)	-0.086***	(0.011)	-0.013	(0.010)
in [100,120)	-0.068***	(0.009)	-0.117***	(0.014)	0.001	(0.016)
≥120	-0.114***	(0.019)	-0.160***	(0.027)	-0.052	(0.040)
Unemp. Bucket 1	-0.043***	(0.004)	-0.084***	(0.008)	-0.002	(0.007)
Bucket 2	-0.051***	(0.007)	-0.090***	(0.016)	-0.015*	(0.009)
Bucket 3	-0.061***	(0.008)	-0.092***	(0.012)	-0.032**	(0.014)
Bucket 4	-0.055***	(0.006)	-0.079***	(0.009)	-0.039***	(0.009)
Single 1st & Move Controls	N		N		N	
State and Orig. Vintage F.E.	Y		Y		Y	
Additional Covariates	Y		Y		Y	
Observations	584499		179272		405227	

Note: This table estimates probit models of default, like models (1), (3) and (4) of Table 7a, where we now interact borrower type with FICO at origination, updated LTV, and the 2-year-ahead unemployment change. This table reports the marginal effects of changing the borrower type variable category. State and origination half-year fixed effects included. Standard errors are in parentheses to the right of the marginals (clustered at the county level); * p<0.10, ** p<0.05, *** p<0.010.
Source: Authors' calculations of McDash, CCP, CRISM, CoreLogic, and BLS.

Table 8a: High Bankcard Utilization, Default and Fraud

	(1)	(2)	(3)	(4)
	2005–17 Orig.		2005–07	2008–17
Mortgage Default w/in 2 yrs	0.218***	0.218***	0.168***	0.254***
	(0.004)	(0.004)	(0.006)	(0.005)
FHA	0.056***	0.056***	0.045***	0.054***
	(0.002)	(0.002)	(0.003)	(0.002)
Portfolio	0.009***	0.008***	0.010***	0.008**
	(0.002)	(0.002)	(0.004)	(0.003)
Private Securitized	0.025***	0.026***	0.022***	0.031***
	(0.002)	(0.002)	(0.003)	(0.004)
FICO Orig. in [660,700)	-0.101***	-0.100***	-0.115***	-0.097***
	(0.002)	(0.002)	(0.004)	(0.003)
FICO Orig. in [700,750)	-0.194***	-0.193***	-0.193***	-0.195***
	(0.002)	(0.002)	(0.004)	(0.003)
FICO Orig. in [750,800)	-0.274***	-0.273***	-0.257***	-0.280***
	(0.002)	(0.002)	(0.004)	(0.003)
FICO Orig. ≥800	-0.300***	-0.299***	-0.276***	-0.308***
	(0.003)	(0.003)	(0.004)	(0.003)
Origination LTV (70,80]	0.009***	0.010***	0.021***	0.003
	(0.002)	(0.002)	(0.003)	(0.002)
Origination LTV (80,90)	0.027***	0.027***	0.031***	0.027***
	(0.002)	(0.002)	(0.005)	(0.003)
Origination LTV ≥90	0.037***	0.037***	0.044***	0.036***
	(0.002)	(0.002)	(0.005)	(0.003)
Orig. Amt. in (200k,359650]	-0.013***	-0.012***	-0.013***	-0.012***
	(0.001)	(0.001)	(0.002)	(0.001)
Orig. Amt. in (359650,417k]	-0.019***	-0.018***	-0.016***	-0.019***
	(0.002)	(0.002)	(0.004)	(0.003)
Orig. Amt. in (417k,700k]	-0.035***	-0.034***	-0.032***	-0.034***
	(0.003)	(0.003)	(0.005)	(0.003)
Orig. Amt >700k	-0.047***	-0.046***	-0.041***	-0.045***
	(0.005)	(0.005)	(0.006)	(0.007)
Updated LTV in [80, 90)	0.010***	0.011***	0.007***	0.008***
	(0.002)	(0.002)	(0.003)	(0.002)
Updated LTV in [90,100)	0.016***	0.017***	0.013***	0.014***
	(0.002)	(0.002)	(0.004)	(0.002)
Updated LTV in [100,120)	0.018***	0.019***	0.020***	0.014***
	(0.002)	(0.002)	(0.004)	(0.003)
Updated LTV ≥120	0.026***	0.026***	0.035***	0.015**
	(0.004)	(0.004)	(0.006)	(0.006)
Unemp. Bucket 2	0.009***	0.008***	0.010**	0.006***
	(0.002)	(0.002)	(0.004)	(0.002)
Unemp Bucket 3	0.012***	0.012***	0.015***	0.003
	(0.003)	(0.003)	(0.005)	(0.005)
Unemp Bucket 4	0.010***	0.010***	0.008	0.006
	(0.004)	(0.004)	(0.005)	(0.005)
Single First Lien		0.013***		
		(0.002)		
Moved		0.007**		
		(0.003)		
State and Vintage F.E.	Y		Y	
Additional Covariates	Y		Y	
Observations	502918	502918	146827	356091

Note: Marginal effects from probit models for the probability of a borrower having bankcard utilization greater than or equal to 80 percent two years following origination or at the time of first default (if earlier). The models also include an interaction term between borrower type and mortgage default (reported in Table 8b). All models include origination half-year and state fixed effects. Other covariates in online appendix. Standard errors are in parentheses (clustered at county level); *** p<0.01, ** p<0.05, * p<0.1.
Source: Authors' calculations of McDash, CCP, CRISM, CoreLogic, and BLS data.

Table 8b: High Bankcard Utilization, Default and Fraud (Interactions)

	(1)	(2)	(3)	(4)
	2005–17 Orig.		2005–07 Orig.	2008–17 Orig.
Honest Homeowner				
Mortgage not in Default	0.006**	-0.011***	0.001	0.012***
	(0.003)	(0.004)	(0.046)	(0.004)
Mortgage in Default	0.178***	0.157***	0.136***	0.185***
	(0.010)	(0.011)	(0.011)	(0.024)
χ^2	258.33***	230.41***	127.5***	46.28***
Declared Investor				
Mortgage not in Default	-0.004	-0.014***	-0.000*	-0.002
	(0.004)	(0.004)	(0.005)	(0.005)
Mortgage in Default	0.079***	0.070***	0.083***	0.107**
	(0.016)	(0.017)	(0.018)	(0.0376)
χ^2	25.99***	25.75***	20.20***	9.05***
Second Homeowner				
Mortgage not in Default	0.032***	0.022***	0.028***	0.036***
	(0.004)	(0.005)	(0.006)	(0.005)
Mortgage in Default	0.092***	0.083***	0.110***	0.053
	(0.026)	(0.026)	(0.030)	(0.049)
χ^2	5.26**	5.18**	6.63***	0.11
Single 1st & Moved Controls	N	Y	N	N
State and Vintage F.E.	Y	Y	Y	Y
Additional Covariates	Y	Y	Y	Y
Observations	502918	502918	146827	356091

Note: The marginal effects of changing borrower type relative to the baseline type of fraudulent investor, interacted with whether or not the borrower's mortgage was in default (60+ DPD) within two years of origination, for the probit models of high utilization from Table 8a. Columns correspond to those in Table 8a. The chi-squared statistic is from a test of equality of the marginal effects across default status. Standard errors are in parentheses (clustered at county level); *** p<0.01, ** p<0.05, * p<0.1.
Source: Authors' calculations of McDash, CCP, CRISM, CoreLogic, and BLS data.

Table 9a: Alternative Hypothesis: "Accidental Fraud" — Determinants of Multiple Liens and Fraud

	(1) Multiple 1sts	(2) Multiple 1sts	(3) No Move	(4) No Move	(5) Fraud	(6) Fraud
DOM in [60,120) days		0.010***		0.001		0.001
		(0.002)		(0.002)		(0.002)
DOM in [120,180) days		0.016***		0.003		0.003
		(0.003)		(0.002)		(0.002)
DOM in [180,365)		0.023***		0.007***		0.007***
		(0.004)		(0.003)		(0.003)
DOM ≥365 days		0.034***		0.012**		0.012**
		(0.008)		(0.005)		(0.005)
HPI Chg: 1 year ahead		-0.194***		-0.100***		-0.100***
		(0.018)		(0.009)		(0.009)
Additional Covariates	Y	Y	Y	Y	Y	Y
State and Vintage F.E.	Y	Y	Y	Y	Y	Y
Observations	468572	468572	468572	468572	468572	468572
R^2	0.111	0.112	0.042	0.043	0.042	0.043

Note: Linear probability models for the likelihood of a borrower having one or more characteristics used to define fraud. Sample restricted to loans that report being owner-occupied in CRISM. All specifications include constant, and origination half-year and state fixed effects. Standard covariates included but not reported. Standard errors (clustered by pre-origination CCP county) are in parentheses; *** p<0.01, ** p<0.05, * p<0.1.
Source: Authors' calculations of McDash, CCP, CRISM, CoreLogic, and BLS data.

Table 9b: Alternative Hypothesis: "Accidental Fraud" — Mortgage Default

	2005–17 Origs.	2005–17 Origs.	2005–17 Origs.
Honest Homeowner	-0.034***	-0.033***	-0.026***
	(0.002)	(0.002)	(0.003)
Declared Investor	-0.042***	-0.041***	-0.036***
	(0.002)	(0.002)	(0.003)
Second Home	-0.042***	-0.041***	-0.035***
	(0.003)	(0.003)	(0.003)
DOM [60,120)		0.005***	0.005***
		(0.001)	(0.001)
DOM [120,180)		0.007***	0.007***
		(0.001)	(0.001)
DOM [180,365)		0.010***	0.010***
		(0.002)	(0.002)
DOM ≥365		0.011***	0.011***
		(0.003)	(0.003)
HPI apprec: 1 yr. since orig.		-0.098***	-0.096***
		(0.007)	(0.007)
Controls for Single First Lien and Moving	N	N	Y
Additional Covariates	Y	Y	Y
State and Vintage F.E.	Y	Y	Y
Observations	531511	531511	531511

Note: Marginal effects from probit models for the probability of a borrower being 60+ DPD after two years. Other covariates (not reported) as in Table 7a. All models include origination half-year and state fixed effects. Standard errors are in parentheses (clustered at county level); *** p<0.01, ** p<0.05, * p<0.1.
Source: Authors' calculations of McDash, CCP, CRISM, CoreLogic, and BLS data.

Table 9c: Alternative Hypothesis — "Default and Moving"

Models of Mortgage Default: Drop Defaults in 1st Year

	(1) 2005-17 Origs.	(2) 2005–07 Origs	(3) 2008-17 Origs.
Owner-occupant	-0.018***	-0.036***	-0.001
	(0.002)	(0.003)	(0.002)
Investor	-0.025***	-0.036***	-0.013***
	(0.002)	(0.003)	(0.002)
Second home	-0.024***	-0.042***	-0.008**
	(0.002)	(0.004)	(0.003)
Additional Covariates	Y	Y	Y
State and Vintage F.E.	Y	Y	Y
Observations	568932	170371	398561

Note: Marginal effects from probit models for the probability of a borrower being 60+ DPD after two years. Borrowers who defaulted in the first year after origination are dropped from the sample. Other covariates (not reported) as in Table 7a. All models include origination half-year and state fixed effects. Standard errors are in parentheses (clustered at county level); *** p<0.01, ** p<0.05, * p<0.1.
Source: Authors' calculations of McDash, CCP, CRISM, CoreLogic, and BLS data.

Two-Year Default Rates for Declared Investors and Second Homeowners

	2005–2007 Orig.	**2008–2017 Orig.**
Movers	0.074	0.01
Non-movers	0.075	0.023

Source: Authors' calculations of McDash, CCP, and CRISM data.

Table 10: Summary Statistics Including Single-Lien Non-Movers

	Honest	Fraud	Investor	2nd Home	Single-Lien Non-mover
Count	491822	22431	47300	22946	75332
Share	0.75	0.03	0.07	0.03	0.11
Delinq.	0.06	0.14	0.05	0.02	0.08
Moved	1.00	0.00	0.44	0.38	0.00
Zip Distance ≤75	0.96	0.69	0.71	0.32	0.75
Multi	0.13	1.00	0.47	0.42	0.00
Bubble State	0.20	0.33	0.26	0.33	0.23
FICO Orig. (mean)	718	725	740	757	718
LTV Ratio at Orig. (mean)	87	80	76	74	84
Orig. Amt. (mean)	225698	287788	173493	267992	218967
HPI Chg: 2 yrs ahead	0.03	-0.02	0.01	0.02	0.03
HPI Chg.: 2 yr lag	0.05	0.10	0.07	0.10	0.06
Has Second	0.14	0.39	0.32	0.30	0.13
Orig. Int. Rate	5.03	5.60	5.77	5.13	4.96
Brokered	0.27	0.24	0.16	0.15	0.26
FRM	0.92	0.78	0.91	0.83	0.92
ARM; 1 year intro rate	0.00	0.01	0.01	0.02	0.00
ARM; 2 year intro rate	0.02	0.07	0.02	0.01	0.03
ARM; 3 year intro rate	0.01	0.02	0.01	0.01	0.01
ARM; 5 year intro rate	0.03	0.08	0.04	0.07	0.03
ARM; 7 year intro rate	0.01	0.02	0.01	0.03	0.01
ARM; 10 year intro rate	0.01	0.02	0.01	0.02	0.01
IO	0.04	0.13	0.04	0.09	0.03
OptionARM	0.01	0.04	0.02	0.00	0.01
Lowdoc	0.13	0.18	0.15	0.15	0.12
Unknown Doc	0.16	0.25	0.30	0.26	0.18
FHA	0.41	0.14	0.14	0.00	0.36
GSE	0.44	0.55	0.72	0.78	0.49
Portfolio	0.07	0.11	0.05	0.12	0.07
Private Securitized	0.08	0.19	0.10	0.10	0.08
Bankcard Util > 80 (in default)	0.63	0.34	0.46	0.37	0.50
Bankcard Util >80 (no default)	0.22	0.14	0.16	0.13	0.22

Table 11: Default Regressions — Robustness (2005–17 Origs.)

	(1) Include Single Lien No-Mover	(2) Include Single Lien No-Mover	(3) Drop Fraud >75 mi. Zip	(4) Multiple Firsts Only
Owner-occupant	-0.035***	-0.028***	-0.031***	-0.020***
	(0.002)	(0.003)	(0.002)	(0.002)
Investor	-0.042***	-0.037***	-0.039***	-0.033***
	(0.002)	(0.003)	(0.003)	(0.003)
Second home	-0.044***	-0.038***	-0.040***	-0.039***
	(0.003)	(0.003)	(0.003)	(0.003)
Single First Lien Non-mover	-0.023***	-0.010***		
	(0.002)	(0.002)		
Single First Lien		-0.012***		
		(0.001)		
Moved		0.004*		
		(0.002)		
Moving & Single First Controls	N	Y	N	N
Additional Covariates	Y	Y	Y	Y
State and Vintage F.E.	Y	Y	Y	Y
Observations	659831	659831	577594	118065

Note: Marginal effects from probit models for the probability of a borrower being 60+ DPD after two years. Other covariates (not reported) as in Table 7a. All models include origination half-year and state fixed effects. Standard errors are in parentheses (clustered at county level); *** p<0.01, ** p<0.05, * p<0.1.
Source: Authors' calculations of McDash, CCP, CRISM, CoreLogic, and BLS data.