Local Gambling Preference and Mortgage Misrepresentation

Jiawei Hu *

July 2024

Abstract

This paper investigates whether gambling preference can serve as an explanation of why borrowers commit mortgage misrepresentation. Using a large sample of mortgages originated from 2005 to 2007, we study the effect of local gambling preference on variations in second-lien misrepresentation and owner-occupancy misreporting and its impact on loan performance by OLS, probit, and causal forest approach. We find that both types of mortgage misrepresentation are more likely to occur in areas with higher levels of local gambling preference. In addition, loans with second-lien misrepresentation, on average, have better performance than loans truthfully reported as having second liens while loans with owner-occupancy misreporting have worse performance than loans truthfully reported as not owner occupied. Such effects are greater in areas with higher levels of local gambling preference. Overall, we find gambling preference is a determinant of mortgage misrepresentation commission.

Keywords: Mortgage fraud, Local gambling preference, Second-lien misrepresentation, Owner occupancy misreporting

JEL Codes: G4, G5, R2, R3

*Jiawei.Hu@utdallas.edu, Naveen Jindal School of Management, University of Texas at Dallas, 800 West Campbell Road, Richardson, Texas, 75080.

1 Introduction

As the US economy step into a new cycle of housing boom in 2020s, the memory of last housing market cycle is recalled. The dramatic boom and bust reveal the potential large quality issues of mortgage. As a result, the fraudulent behavior[1] in mortgage market becomes an area of particular interest in academic. A large body of literature documents the widespread fraudulent activity, such as owner-occupancy status misreporting, simultaneous second-lien misrepresentation, income documentation falsification, home appraisal inflation, and so on[2], and their economic impact. While the motivation of lenders and the incentive of borrowers have been well investigated, little is known about the mechanism of how such motivations turn to actual fraud activity. We fill this gap in the literature by investigating whether borrower's gambling preference help in explaining the commission of mortgage fraud.

Mortgage fraud, as a type of criminal, attracts great attention from the government and judicial organ, especially as it was prevalent before the 2008 financial crisis. To prevent the fraud, the punishment is severe: it may include convictions and prison time, restitution payments, state fines, and/or probation, and the penalty could be up to 30 years imprisonment and a fine of no more than $1 million (Henning, 2009; Federal Housing Finance Agency, 2023). The average sentence length for mortgage fraud offenders was around two years in 2010s (United States Sentencing Commission, 2015)[3]. Though mortgage frauds are not caught all the time, the detected cases are not rare: just in year 2006, financial institutions filed 37,313 Suspicious Activity Report citing suspected mortgage loan fraud.

When borrowers decide to commit mortgage fraud, which is probably premeditated, they would balance the benefits and costs of such action. On one hand, they could get a larger amount of the loan (higher leverage) or a lower financial cost (lower interest rate).

[1]We will use fraud and misrepresentation interchangeably throughout the paper

[2]See, for example, Griffin and Maturana (2016) and Kruger and Maturana (2021) for appraisal inflation, Piskorski et al. (2015) and Griffin and Maturana (2016) for second-lien misrepresentation and owner-occupancy misreporting, Elul et al. (2010), Ambrose et al. (2016), and Mian and Suf (2017) for income overstatement.

[3]The Mortgage Offenses Report by United Stated Sentencing Commission is available from 2015, and the average sentence length from 2015 to 2020 is range from 18 to 28 months.

On the other hand, they could be caught and punished. From the perspective of legislation, the expected punishment should be set greater than the expected benefit from crimes to prevent them. Therefore, we expect that a rational household investor would rarely commit mortgage fraud. To help in explaining such contradiction, we propose a preference-based conjecture: people with high gambling preference tend to commit mortgage fraud.

A similar irrational action which people take is gambling. It is also attractive to many people even if people know the rational expected payoff is negative. The preference of gambling affects people widely in decision-making process, not only letting individuals play casinos or buy lotteries but also affecting them when they invest in financial markets (Kumar, 2009). One prevalent theory in explaining gambling preference is the prospect theory proposed by Kahneman and Tversky (1979) and Tversky and Kahneman (1992). If households also overweight the tail events, as described by the probability weighting component in the prospect theory, when they make decisions on mortgage, then it is easy to understand why they take fraudulent action even if the rational expected payoff is negative. High gambling preference people would probably transform the objective probability so that the expectation based on the values become positive. Therefore, we conjecture that people with high gambling preference are more likely to misrepresent in mortgage application.

One major challenge of the study is to find a proper measure for gambling preference. Since the potential distribution of losses and gains for individual loans is not observable, as well as hard to infer, as in financial markets, we cannot directly apply the prospect theory in housing market[4]. In addition, since properties are illiquid assets and attached to specific locations, using a household representative in aggregate market is hard. However, due to the feature of attachment to specific locations, we can instead study how the geographic difference of gambling preference affect the propensity of mortgage fraud. It is plausible also because people who live in different areas may have different cultures and thus have distinct attitude towards gambling. Therefore, we follow Kumar (2009) and Kumar et al. (2011)

[4]See, for example, Benartzi and Thaler (1995) (bond market), Barberis et al. (2016) (stock market), and Baele et al. (2019) (option market) for the application of prospect theory in financial markets.

to measure local gambling preference. We use the ratio of Catholic residents to Protestant residents (CPRATIO) at county level as the measure of local gambling preference.

To investigate our conjecture, we study two types of mortgage misrepresentation: second-lien misrepresentation and owner-occupancy misreporting. Second-lien misrepresentation is that when the borrower applies for a first lien on the property, she intentionally hides her simultaneous second lien. By doing so, the loan is easier approved so that she can use higher leverage. Owner-occupancy misreporting is that when the borrower applies for a loan, he reports the occupancy status of the property as fully owner-occupied though the true status may be second-home or investment. By doing so, he can get lower interest rate, decreasing the financial cost of the loan. We follow Piskorski et al. (2015) to construct the measures but compare in different samples. To study the effect of gambling preference on mortgage fraud, we require that the loans to have a basis for fraud. For example, in order to have second-lien misrepresentation, the loan should have a simultaneous second lien, otherwise the borrower has no option for misrepresentation at all. Therefore, instead of comparing loans with misrepresented second liens to loans truthfully have no second lien, we compare loans with misrepresented second liens to loans have simultaneous second liens but honestly reported.

With the gambling preference measure and mortgage fraud measures, we first investigate whether borrower's gambling preference help in explaining the commission of mortgage fraud. By comparing the proportion of mortgage misrepresentation in different areas, we can link the local gambling preference to the propensity of mortgage fraud after controlling gambling preference related factors and other key determinants of mortgage fraud. Indeed, we find that both types of mortgage misrepresentation are more likely to occur in counties with higher levels of local gambling preference, indicating that gambling preference is an important element that drives borrowers to decide commission of mortgage fraud. We further investigate the effect in subsamples. We find that the effect of gambling preference on second-lien misrepresentation tends to be greater when the borrowers have low credit scores, the occupancy status is owner-occupied, or the loan purpose is to purchase the property. On the contrary, the effect of gambling preference on owner-occupancy misreporting is larger when the borrowers have high credit scores or the loan purpose is to refinance.

Second, we study the economic impact of mortgage fraud driven by gambling preference. Since gambling preference is one determinant of mortgage misrepresentation, we want to know whether it leads to worse loan performance through mortgage fraud as mortgage fraud is found to cause bad loan performance (Piskorski et al., 2015; Griffin and Maturana, 2016). We find that in our samples, the loans with second-lien misrepresentation, on average, have better loan performance than loans truthfully reported as having second liens. In addition, such effect on loan performance is greater in areas with higher levels of gambling preference. In contrast, similar to samples in other literature (e.g., Piskorski et al. (2015)), we find that the the loans with owner-occupancy misreporting have worse loan performance than loans truthfully reported as not owner occupied. Similarly, such effect on loan performance is greater in areas with higher levels of gambling preference, though the economic magnitude is smaller. In subsample analysis, we find that the effect of gambling preference on loan performance through second-lien misrepresentation holds in almost all subsamples, but the effect through owner-occupancy misreporting is significant only when the loan purpose is to purchase, when the borrowers have low credit scores, or when the income documentation requirement is low.

Finally, we perform multiple robustness checks. First, to make sure our results are not driven occasionally by a single default measure, we test other default measures that are either more or less restricted, and the results are not materially affected. Second, to ensure that our findings are not driven by the use of linear regression model (OLS), we also estimate probit regression models and find similar results. Third, to enhance the causal inference and get a more accurate estimation of treatment effects, we apply causal forest approach proposed by Wager and Athey (2018), which gives similar findings.

Our paper contributes to two strands of literature. First, our paper complement the literature of mortgage fraud in proposing preference-based explanation. While the prevalence and the economic impact of mortgage fraud are well investigated (Elul et al., 2010; Piskorski et al., 2015; Ambrose et al., 2016; Griffin and Maturana, 2016; Mian and Suf, 2017; Kruger and Maturana, 2021) and the incentives of fraud are clear in practice, the mechanism of how such motivations turn to actual fraud activity is uncovered yet, especially when legislative and judicial organs have clear penalties to prevent mortgage fraud. Conklin et al. (2022) shows empirical evidence that religiosity has significant effect on mortgage fraud, which potentially helps link the motivations to actual fraud activity, but it is from the direction of why borrowers do not commit fraud rather than how they turn the motivations to actual activity.

Second, our paper builds on the empirical literature linking gambling preferences with investment decisions. Prior literature studies the effect of gambling preference on stock market (Barberis and Huang, 2008; Kumar, 2009; Kumar et al., 2011; Barberis et al., 2016, 2021), bond market (Benartzi and Thaler, 1995), and option market (Baele et al., 2019). Literature also explores the effect in decision-making of corporate policy (Kumar et al., 2011; Chen et al., 2014) and fund strategy (Shu et al., 2012). To our best knowledge, we are the first to apply gambling preference on mortgage market for household investment decision-making process.

The rest of the paper is organized as follows. Section 2 describes the data and variables. In Section 3, we study the effect of gambling preference on mortgage misrepresentation. In Section 4, we investigate the economic impact of gambling preference associated mortgage misrepresentation. We do robustness check in Section 5 and conclude in Section 6.

2 Data and Summary Statistics

Our sample mainly contains three groups of data, including loans-related records, religiosity data, and demographics information, from 2005 to 2007. We draw them from numerous datasets. Loans-related records include loan-level mortgage data from BlackBox Logic (since acquired by Moody's) and borrower level credit report information collected by Equifax. Religiosity data, which is county-level information on prevalent religious adherence, is collected from the American Religion Data Archive (ARDA). Demographics information in county and zipcode level, such as age and income, is from the U.S. Census Bureau. In addition, we also collect house price data from Federal Housing Finance Agency.

2.1 Mortgage Misrepresentation Measures

Following Piskorski et al. (2015) and Zhang et al. (2024), we generate two measures for mortgage misrepresentations: second-lien misrepresentation and owner-occupancy misreporting[5].

Second-lien misrepresentation happens when a first-lien loan backed by property is reported as no associated higher lien but is actually financed with a simultaneously originated second mortgage identified by the credit bureau data. Such misrepresentation allows the borrower to take additional debts, which give borrower less incentive to repay the loans. As a result, second-lien misrepresentation makes the initial debt riskier. We identify second-lien misrepresentation by comparing loan-level mortgage data from BlackBox Logic and borrower-level credit report information from Equifax. Following the procedure in Piskorski et al. (2015), we first focus on the first-lien loans that have merge confidence interval greater than or equal to 0.89. Second, we pick up the loans that have new second-lien with origination date within one-month before or after of the first lien[6]. This filter using credit information keeps the first-lien loans that truly have a simultaneous second lien. Third, to find out the misrepresented second-lien loan, we require the loan to have nonmissing reported cumulative loan-to-value (CLTV) ratio within 1% of its loan-to-value (LTV) ratio. The small difference between reported CLTV and LTV shows that the borrower reports no

[5]See Griffin and Maturana (2016) for alternative construction methods of the two measures using different datasets.

[6]Piskorski et al. (2015) uses 45-days range while Zhang et al. (2024) uses one-month range. Since the data are more accurate at monthly-level time point than daily-level time point, we conservatively use one-month level. We also check 45-day range and two-month range, and the results have minor difference.

simultaneous second lien.

Owner-occupancy misreporting is the case that the borrowers report the occupancy status as owner occupied while the true occupancy type is not[7]. Since borrowers who own and fully occupy the property are less likely to default on this mortgage than borrowers who do not, originators usually offer more favorable loan terms, such as low down payments and low interest rates. Therefore, borrowers may have incentive to misreport occupancy status in order to take the advantage. We construct this measure using procedure described in Piskorski et al. (2015) Internet Appendix. First, we pick up the loans that are reported as primary in BlackBox Logic (acquired by Moody's). The borrower of these loans report the occupancy status as fully owner-occupied. Second, we compare the property zip code reported to BlackBox Logic at origination to the zip code reported to Equifax each month over the first year of the loan's life. The occupancy status is classified as misreporting if none of the twelve zip code reported to Equifax match with the zip code reported to BlackBox Logic. This mismatch means that the zip code used by the borrower is not the zip code she receive her credit report, possibly implying that the property is not fully occupied[8].

2.2 Gambling Preference and Religiosity

Following Kumar et al. (2011), we construct the county level gambling preference based on religion data, which is the ratio of the county's Catholic residents to Protestant residents (CPRATIO). Although direct measures of local gambling preference are not available, we can measure the propensity by looking at the proportion of different religious population who have distinct gambling attitude according to the religious views. Two widespread religions in US that have different gambling views are Catholicism and Protestantism. While Protestant

[7]Typical examples of other occupancy tpye are second home and non-owner occupied/investment. The BlackBox Logic record for occupancy type also includes other, unknown, and vacant. We restrict our sample to owner occupied, second home, and non-owner occupied.

[8]As noted in Piskorski et al. (2015) paper, owner-occupancy misreporting measure is different from second-lien misrepresentation measure. While second-lien misrepresentation can be directly observed, owner-occupancy misreporting in given databases could only be inferred. Thus, the owner-occupancy measure could be more noisy.

churches generally oppose gambling, Catholic churches remain a tolerant attitude towards moderate levels of gambling. Such different views between two religions are empirically supported to extend to financial market (Kumar, 2009; Kumar et al., 2011; Han and Kumar, 2013; Chen et al., 2014). Therefore, regions with higher Catholic–Protestant ratios have stronger gambling propensity.

In order to use CPRATIO with potential establishment of causality, we need to control for the religiosity in the county. According to Kumar et al. (2011) and Chen et al. (2014), religiosity should be considered when using CPRATIO in financial market studies because risk aversion increases with religiosity, irrespective the type of religion (Hilary and Hui, 2009). Thus, including the overall level of religiosity of the county as a control help us make sure that our local gambling preference proxy is independent of religion-induced risk aversion. In addition, religiosity also plays a significant role in deterring many types of mortgage misrepresentation (Conklin et al., 2022) from the perspective of social norms and ethical behavior. Therefore, we follow Kumar et al. (2011) to construct the religiosity measure.

We use dataset "Longitudinal Religious Congregations and Membership File, 1980-2010 (County Level)" from ARDA to capture the county level geographical variation in religious composition. We use the sum of the number of adherents in religious traditional category Evangelical Protestant, Mainline Protestant, and Black Protestant for Protestant population and the number of adherents in the category Catholic for Catholic population. We use the sum of the number of adherents in all the category for the total religious adherents in the county. Since the data is available for each decade, from 1980 to 2010, we follow previous studies to linearly interpolate the data to obtain the values for missing years (Alesina and La Ferrara, 2000; Hilary and Hui, 2009; Kumar et al., 2011; Chen et al., 2014), and then calculate the CPRATIO (i.e., Catholic population to Protestant population) and REL (i.e., total religious adherents to total population) in each year.

2.3 Geographic Controls

Except for religiosity, variation in religion-induced gambling preference may also correlated with other geographic characteristics (Kumar et al., 2011; Chen et al., 2014; Conklin et al., 2022). To help in establishing causality, we control for the following factors (Kumar, 2009; Kumar et al., 2011; Chen et al., 2014). U.S. Census Bureau contains a rich set of demographic data. We employ county level information about education, marriage, living area, population, age, male-female ratio, and minority proportion. Since income plays an important role not only in potential correlation with religion distribution but also in household decision in mortgage, we use an even smaller cluster level, zip code level, for the control variables. Moreover, we also collect annual house price indices (HPI) at zip code level to calculate the house price appreciation, which is also an crucial determinants in household decision in mortgage.

2.4 Mortgage Microdata

In addition to the two mortgage misrepresentation measures, we also obtain other loan-level mortgage data from BlackBox Logic. The database includes a rich set of loan characteristics at origination, including loan interest rate at origination, borrower's FICO credit score at loan origination, initial loan balance, loan-to-value ratio at origination, amortization type (e.g., full amortization, interest only, negative amortization), income documentation type (e.g., full, low-doc, or no-doc), interest rate type (e.g., fixed or adjustable), prepayment penalty, loan purpose (e.g., purchase or refinance), reported occupancy status (e.g., owner-occupied, investment, or second-home), delinquency method (i.e., MBA or OTS), and delinquency status (e.g., current, 30, 60, 90, etc.). We show the definition of variables constructed by the information above in Table 1.

[insert Table 1 here]

2.5 Descriptive Statistics

We study the effect of gambling preference on mortgage fraud in two samples, one for second-lien misrepresentation and one for owner-occupancy misreporting. We select the sample period of 2005 to 2007 that mortgage misrepresentation measures can be calculated using the databases and the frauds are not rare. Table 2 shows the summary statistics for the two samples separately in panel A (second-lien misrepresentation) and panel B (owner-occupancy misreporting). For each continuous variable, we report the number of observations, mean, standard deviation, 25th percentile, 50th percentile, and 75th percentile, and for dummy variables, we report the number of observation and mean. The variables are winsorized at 1 percent level so that our results are not driven by extreme values[9].

For second-lien misrepresentation, we pick up the loans that have second lien, either honestly reported or fraudulently hidden. By comparing cases that borrowers honestly report their simultaneous second lien and cases that borrowers fraudulently hide, we can clearly study whether borrowers in high gambling preference areas are more likely to cheat in mortgage market. In our sample, we have about 633,904 loans that have second lien and the borrower's area CPRATIO is available[10]. Among these cases, about 34 percent of borrowers chose to misrepresent the second lien. This percentage is higher than the one shown in Piskorski et al. (2015) (7.13%) whose sample focuses on loans that are reported as having no simultaneous second lien (fraudulently hidden and truly don't have) and the one shown in Griffin and Maturana (2016) (10.2%) who use different database and include all loans (honestly report, fraudulently hide, and truly don't have).

To investigate whether borrowers in high gambling preference areas are also more likely to misreport occupancy status, we choose the loans whose status are investment or second-home, either honestly reported or fraudulently hidden. The sample size is similar to the

[9]Winsorize or not has minor influence on our regression results.

[10]We also drop the cases that LTV is greater than 99% since our measure is calculated using LTV. A LTV that is greater than 99% is probably a wrong data or highly unlikely to have a simultaneously second lien.

size of second-lien misrepresentation: we have about 646,457 loans that are not fully owner-occupied and the borrower's area CPRATIO is available. In addition, about 35 percent of borrowers in this sample chose to misreport the occupancy status. The higher proportion of mortgage fraud than previous literature is also caused by the selection of sample components (6.42% in Piskorski et al. (2015) and 6.7% in Griffin and Maturana (2016)).

Our primary independent variable of interest is local gambling preference (CPRATIO). The means of the two samples are 1.31 and 1.51, and the medians are 0.94 and 1.32. They are higher than the mean and the median across all the 3,090 data available U.S. counties from 2005 to 2007, which is only about 0.56 and 0.19, respectively[11]. The large differences mean that the loans with simultaneous second lien or not owner occupied status in our sample concentrated more in higher gambling preference areas. Figure 1 plots the country-wide distribution of the data. The top panel shows the CPRATIO, indicating that higher gambling preference areas concentrated in west coast, southwest, and east coast. The middle and the bottom panels plots the proportion of second-lien misrepresentation and the proportion of owner-occupancy misreporting, respectively, in the samples. Except for the south east of the country, the proportion of the two types of mortgage misrepresentations in the samples seems to increase with CPRATIO.

[insert Figure 1 here]

The loan characteristics in Table 2 reflect the features of loans with second lien and the features of loans that are not fully owner-occupied. The average natural logarithm of FICO score are about 6.53 and 6.56, which translate to 686 and 705 FICO score. They are slightly higher than the lower bound of Good level. A large part of LTV ratio for second-lien misrepresentation sample is 80, which indicates that the borrowers do have enough incentive to have a simultaneous second lien. More than half of loans in both samples are low or no documentation loans (68% and 76%), implying a relatively loose income documentation

[11] The mean and the median of CPRATIO from 1980 to 2005 are 0.60 and 0.23, respectively, in Kumar et al. (2011), which is quite close to our results.

requirement at origination. The total proportions of cash-out and no-cash-out refinance in the two samples are only 22% and 36%, implying a even lower proportion of misrepresentation in refinancing cases. The proportion of purpose of investment and second-home are quite different in the two samples (12% v.s. 65%). This difference reflects the potential difference in purpose of the fraud, as the natures of housing include both consumption and investment (Campbell, 2006).

[insert Table 2 here]

3 Gambling Preference and Mortgage Misrepresentation

In this section, we investigate the relationship between gambling preference and mortgage fraud. Since the expected penalty for fraud should be greater than the expected gain from fraud by legislative design, a rational borrower should not take fraudulent action. However, as explained by the prospect theory (Tversky and Kahneman, 1992; Barberis, 2013), high gambling preference people would probably transform the objective probability, overweighting the probability of tail event (i.e., large loss from being punished and large gain from successful fraud), so that the expectation of transformed values become positive. The greater the distortion on probability weighting function is, the larger probability of a positive payoff will appear. Thus, we expect that borrowers in high local gambling preference areas are more likely to misrepresent material information in mortgage. We formalize this thought in Hypothesis 1 and 2:

Hypothesis 1 *Second-lien misrepresentation is more likely to occur in counties with higher levels of local gambling preference.*

Hypothesis 2 *Owner-occupancy misreporting is more likely to occur in counties with higher levels of local gambling preference.*

To test our hypotheses, we estimate loan-level linear regressions of the following form:

$$Y_{it} = \alpha + \beta CPRATIO_{ct} + \gamma X_{it} + \delta_s + \eta_t + \lambda_o + \epsilon \tag{1}$$

where Y_i is an indicator for mortgage fraud on loan i originated at time t; $CPRATIO_{ct}$ is the county level measure of gambling preference at time t; X_{it} includes loan i's religiosity, geographic controls, and loan characteristics at time t; δ_s is state fix effects; η_t is origination time fixed effects, which can be either year (Conklin et al., 2022) or half year (Piskorski et al., 2015); λ_o is originator fixed effects[12]; ϵ is an error term. Moreover, since our primary independent variable of interest (CPRATIO) is measured at the county level, we also cluster heteroskedasticity robust standard errors by county. We include the control variables, fixed effect, and standard errors clustering gradually in the specifications. Besides winsorizating at 1 percent level, we also standardize all the continuous variables.

3.1 Does Borrowers in High Gambling Preference Areas Are More Likely to Misrepresent Second Lien?

To test hypothesis 1, we use second-lien misrepresentation dummy variable as the dependent variable in equation 1. We represent the results in Table 3. Column (1) presents the basic model without any controls variables, fixed effects, or clustering method. Column (2) adds geographic controls only. Column (3) adds loan characteristics controls only. Column (4) adds all control variables. Column (5) adds state fixed effects and year fixed effects. Column (6) adds originator fixed effects. Column (7) further cluster standard errors at county level in column (6). Finally, column (8) changes the year fixed effects to half-year fixed effects in column (7).

Our results reveal that the counties with higher levels of gambling preference indeed tend

[12]Although both originators and underwriters play important roles in mortgage roles (Griffin and Maturana, 2016) and the related fixed effects are used in different literature, we control for originator fixed effects because originators are the agency that interact with the borrowers so that they may affect the decision of borrowers while underwriters do not have such influence in the borrowers decision-making process.

to have more second-lien misrepresentation. One standard deviation increase of CPRATIO in whole sample leads to a 1.6 percent increase of the probability of second-lien misrepresentation without fixed effects and a 0.4 percent increase with state, half-year, and originator fixed effects[13]. The significance and economic magnitude also show that CPRATIO is a leading variable among all county level variables[14]. It is also comparable to loan-level control variables, such as interest rate (-0.4 percent with all fixed effects) and FICO (-2.9 percent with all fixed effects).

Turning to geographic controls, we see that religiosity and income are consistently negatively related to second-lien misrepresentation. In contrast, total population, proportion of old people population, and minority share of population are consistently positively related to this type of mortgage fraud. The relationship of other geographic characteristics varies in different situations. Moreover, except for our variable of interest (CPRATIO), only total population remain robust significance. The loan level characteristics show that high interest rate, low credit score, high initial balance, and high LTV ratio lead to greater probability of misreporting second lien. In addition, second-lien misrepresentation is also more likely to happen when the loan has no convertibility clause, is no or low documented, has no prepayment penalty, is refinancing, or is not fully owner-occupied. Fixed on originator, the sign would change for ARM and negative amortization.

[insert Table 3 here]

As borrowers in areas with higher levels of gambling preference are more likely to misreport second lien in general, we further ask whether it holds in different subsamples. We divide our samples in several ways: primary or non-primary, purchase or refinance, full or low/no documentation, and high or low credit score. The high or low credit score is divided by FICO score of 670, which is the boundary of fair level and good level evaluated by the

[13]Moving from 25th to 75th percentile, CPRATIO increases about 1.27 standard deviation.

[14]The coefficient of CPRATIO is larger than the coefficient of REL, which is proved to be an important factor in mortgage fraud by Conklin et al. (2022), and even makes REL insignificant in column (6) to (8).

institution[15]. We present the results in Table 4. For all specifications, we include state, half-year, and originator fixed effects, and the standard errors are all clustered at the county level.

Column (1) and (2) report the results for primary and non-primary (i.e., fully owner-occupied vs investment plus second-home) subsamples. Most observations of the whole sample belong to the subsample of primary, and only the coefficient of CPRATIO in subsample of primary is significant. In unshown results, we also explore investment subsample and second-home subsample separately, and none of them is significant. This outcome implies that high gambling preference borrowers tend to choose second-lien misrepresentation only when their purpose is to fully owner-occupy the house. Column (3) and (4) present the results for purchase and refinance subsamples. About 77 percent of observations is from the subsample of purchase, and only the coefficent of CPRATIO in subsample of purchase is significant. In unshown results, neither cash-out refinance nor no-cash-out refinance has significant coefficients. It implies that high gambling preference borrowers tend to cheat on second lien representation only when they purchase the house. Column (5) and (6) show the results for full and low or no documentation subsamples. About 64 percent of observations go to the subsample of low or no documentation subsample, but both subsamples have significant results. The economic magnitude in full documentation subsample is greater but not too high compare to the magnitude of low or no documentation subsample. Finally, column (7) and (8) show the results for high and low credit score subsamples. The observations are roughly evenly spread into the two subsamples. While low FICO score subsample has a significant and larger coefficient for CPRATIO, high FICO score subsample's result is insignificant. However, the economic magnitude of the coefficient in high FICO score subsample is close to the one in the whole sample. It implies a positive but noisy propensity of cheating when borrowers has high credit score while the propensity is much more clear among low FICO score borrowers. In general, the subsample analysis may indicate that

[15]We also tried the median of the sample, 686, which gives similar results.

when the benefit of cheating is to achieve greater leverage (low FICO score borrowers are harder to borrow), the borrowers are more likely to show the positive correlation between gambling preference and mortgage fraud when the purpose concentrates on consumption (to occupy by own) or new property (to purchase).

3.2 Does Borrowers in High Gambling Preference Areas Are More Likely to Misreport Owner-occupancy Status?

We now turn to hypothesis 2 to see whether borrowers in high gambling preference counties are also more likely to misreport owner-occupancy status. We use owner-occupancy misreporting dummy variable as the dependent variable in equation 1 and report the results in Table 5. We gradually add control variables, fixed effects, and clustering method in the same order as for second-lien misrepresentation in Table 3.

The results support our hypothesis that owner-occupancy misreporting is more likely to occur in areas where local gambling preference level are higher. Moving from 25th to 75th percentile of CPRATIO, the likelihood of owner-occupancy misreporting increases about 1.39 percent (1.39×0.01) in the case of considering all fixed effects. Though not as primary as in second-lien misrepresentation, gambling preference is still quite important as the economic magnitude is slightly larger than the magnitude of religiosity. It is also comparable to other geographic and loan-level control variables.

For the geographic controls, we see much more consistent sign of the variables with statistical significance, potentially caused by the smaller influence from gambling preference. Specifically, religiosity, income, proportion of population in urban, and proportion of old people population are consistently negative and statistically significant. Education, proportion of married population, total population, and minority share of population are positive and statistically significant. House price appreciation remains positive but become insignificant if standard errors are clustered at the county level. Male-female ratio has little influence in this setting. Except for interest rate at origination, which has nagative sign, other loan-level characteristics have similar effect on owner-occupancy misreporting as the second-lien misrepresentation. In addition, loans that are ARM, allow negative amortization, require low or no documentation, have prepayment penalty, or are not refinancing are more likely to be misreported in owner-occupancy status.

[insert Table 5 here]

Although borrowers who have high gambling preference also more likely to cheat on reporting owner-occupancy status in general, the results in subsample analysis may be different from the ones for second-lien misrepresentation due to potential distinct purpose of mortgage fraud. Thus, we explore the effect of gambling preference separately in the following subsamples: purchase or refinance, full or low/no documentation, and high or low credit score[16]. We report the results in Table 6. For all specifications, we include state, half-year, and originator fixed effects, and the standard errors are all clustered at the county level.

Column (1) and (2) present the results for purchase and refinance subsamples. About 62 percent of observations belong to subsample of purchase, but the coefficients of CPRATIO is significant only in subsample of refinance. In unshown results, we find that both cash-out refinance and no-cash-out refinance have significant coefficients for CPRATIO, and the economic magnitudes are similar. In contrast to the findings in second-lien misrepresentation case, this outcome implies that high gambling preference borrowers tend to select misreport owner-occupancy when their purpose is to invest or only partially occupy the property. Column (3) and (4) report the results for full documentation and low or no documentation subsamples. About 75 percent of observations are from low or no documentation subsamples. Unlike second-lien misrepresentation, only low or no documentation subsample has significant results, even though the economic magtitude of the subsamples are quite close. It indicates that though positive, the propensity of cheating is quite noisy if we divide the whole sample by requirement of income documentation. Finally, column (5) and (6) show the results for high and low FICO score subsamples. The median of FICO score in this

[16]The boundary is also 670.

sample is a little bit far away from the boundary of fair and good level so that the number of observations in high FICO score subsample is greater[17]. Distinct from second-lien misrepresentation, the effect of gambling preference is obvious in high FICO group but is unclear in low FICO group. In general, the implication for owner-occupancy misreporting is different from the one for second-lien misrepresentation: when the benefit of fraud is to decrease financing cost for investment (high FICO score borrowers can borrow more, leading to greater monthly payment), the borrowers are more likely to show the positive correlation between gambling preference and mortgage fraud when the target concentrates on current investment (to refinance).

4 Economic Impact: Gambling Preference Associated Mortgage Misrepresentation and Loan Performance

In the previous section, we find significant explanation power of gambling preference on mortgage fraud. Our next step is to investigate whether gambling preference associated mortgage misrepresentation have meaningful influence on loan's performance. While mortgage fraud usually leads to worse loan performance (Piskorski et al., 2015; Griffin and Maturana, 2016), it is unclear whether gambling preference associated mortgage misrepresentation leads to higher or lower default rate. We conjecture that if the purpose of fraud is to increase the leverage (pay back interest is not a big problem), then gambling to fraud will not lead to worse loan performance. On the contrary, if interest payment is an issue to some extent, which indicates potential financial constraint in cash flow, we conjecture that gambling to fraud will not have better loan performance. Thus, we infer that on one hand, gambling preference associated second-lien will not lead to worse loan performance, and on the other hand, gambling preference associated owner-occupancy misreporting will not lead to better loan performance. We formalize this inference in Hypothesis 3 and 4:

[17] We also tried the median of the sample, 705, which gives similar results.

Hypothesis 3 *Second-lien misrepresented loans in higher levels of local gambling preference counties are equally or less likely to default.*

Hypothesis 4 *Owner-occupancy misreported loans in higher levels of local gambling preference counties are equally or more likely to default.*

To test the hypotheses, we include an interaction term between mortgage fraud measure and gambling preference measure in the form of equation 1:

$$\begin{aligned} Y_{it} = \quad & \alpha + \beta Misrepresentation_{it} + \gamma CPRATIO_{ct} \\ & + \kappa Misrepresentation_{it} \times CPRATIO_{ct} + \phi X_{it} + \delta_s + \eta_t + \lambda_o + \epsilon \end{aligned} \tag{2}$$

where Y_i is the default variable on loan i originated at time t, which equals one if the loan becomes 90 days or more delinquent using MBA method and zero otherwise; $Misrepresentation_{it}$ is the mortgage misrepresentation indicator (either for second-lien misrepresentation or for owner-occupancy misreporting) on loan i originated at time t; $CPRATIO_{ct}$ is the county level measure of gambling preference at time t; $Misrepresentation_{it} \times CPRATIO_{ct}$ is the interaction term between mortgage fraud measure and gambling preference measure; X_{it} includes loan i's religiosity, geographic controls, and loan characteristics at time t; δ_s is state fix effects; η_t is origination time fixed effects for half year; λ_o is originator fixed effects; ϵ is an error term. Moreover, since our primary independent variable of interest (CPRATIO) is measured at the county level, we also cluster heteroskedasticity robust standard errors by county. All continuous variables are winsorizated at 1 percent level and then standardized. We include the control variables, fixed effect, and standard errors clustering in all specifications but gradually include mortgage misrepresentation term, gambling preference term, and the interaction term.

4.1 Does Gambling Preference Associated Second-lien Misrepresentation Affect Loan Performance?

To test hypothesis 3, we use second-lien misrepresentation dummy variable for the mortgage fraud indicator in equation 2. The results are reported in Table 7. In column (1), we include the mortgage fraud measure only with other control variables to see if second-lien misrepresentation has effect on loan performance in our sample. Column (2) includes the gambling preference measure only with other control variables to see if areas with high level of gambling preference also have high rate of default. Column (3) includes both mortgage fraud measure and gambling preference measure. Finally, we include the interaction term between mortgage fraud measure and gambling preference measure in column (4) to test whether gambling preference associated second-lien misrepresentation affect the default rate.

The coefficient of second-lien misrepresentation, which is significantly negative, show that mortgage defaults on loans with misreported second liens are lower than on similar loans that are truthfully disclosed the second liens. In contrast to the findings in Piskorski et al. (2015) who compare loans with misreported second liens with loans that are truthfully disclosed as not having such liens, this outcome implies that conditional on having second liens, loans that are misrepresented the second-lien do not necessarily riskier. The coefficient of CPRATIO show that areas with higher level of gambling preference do not necessarily have greater default rate. The effects of local gambling preference on default show up through its influence on second-lien misrepresentation. The coefficient of the interaction term is negative and statistically significant, indicating that second-lien misrepresented loans in higher levels of local gambling preference counties are actually less likely to delinquent. In addition, the economic magnitude of the coefficient is large (about 60% of second-lien misrepresentation coefficient magnitude), implying a major explanation channel about why second-lien misrepresentation leads to lower rate of delinquency conditional on the existence of second liens.

Since we find that the effect of gambling preference on second-lien misrepresentation may vary in subsamples in previous section, we further ask whether this variation also exists for loan performance. The results presented in Table 8 show that the interaction term is negative and statistical significant in all subsamples except for full documentation. Even if in the subsample of full documentation, the sign is negative and the economic magnitude is large. In general, these results indicate that as long as borrowers who have simultaneous second liens misreport the second liens due to their high level of gambling preference, the loans are less likely to default no matter what purpose they have for the higher leverage.

4.2 Does Gambling Preference Associated Owner-occupancy Misreporting Affect Loan Performance?

Turning to hypothesis 4, we use owner-occupancy misreporting dummy variable for the mortgage fraud indicator in equation 2. We report the results in Table 9 and the column arrangement is same to the arrangement in Table 7 except changing the mortgage fraud measure to owner-occupancy misreporting measure.

Unlike second-lien misrepresentation, the coefficient of owner-occupancy misreporting is significantly positive, which is the same sign as in Piskorski et al. (2015) who compare loans with misrepresented borrower occupancy status to loans with similar characteristics that are indeed owner-occupied. It means that no matter whichever control group to compare, owner-occupancy misreporting loans are more likely to default. The coefficient of CPRATIO is insignificant again, indicating that counties with higher levels of gambling preference do not necessarily have greater default rate. Our main variable of interest in the setting, the interaction term, has a positive and statistical significant coefficient. However, the economic magnitude is not large compared to the magnitude of owner-occupancy misreporting coefficient (about 8%). This outcome reveals that owner-occupancy misreported loans in higher levels of local gambling preference counties are slightly more likely to delinquent.

We also investigate the effect in subsamples and show the results in Table 10. Different from the cases for second-lien misrepresentation, the coefficients of the interaction term are significant only in subsample of purchase, low or no documentation, and low FICO score, while the magnitudes of the coefficient in all subsamples are comparable to the one in whole sample. Note that in Table 5, we find that the effect of gambling preference on owner-occupancy misreporting concentrates on the subsample of refinance, low or no documentation, and high FICO score. It is surprising to see that the subsamples for default with significance are different from the ones for fraud. Thus, it indicates a mixed route from gambling preference to default through owner-occupancy misreporting: borrowers with high gambling preference are more likely to cheat through misreporting owner-occupancy status when their target is current investment (refinance), but the default rate is higher among borrowers whose target is additional investment (i.e., purchase). It is possible that on one hand, when high gambling preference borrowers cheat and successfully reduce the financial cost of current investment, the less financial restriction make default risk lower; on the other hand, if they instead get new property, the financial condition is more constrained even if the additional financial cost is lowered by cheating, increasing their risk to fail to pay. In general, borrowers in high level of gambling preference are more likely to misreport owner-occupancy status, but such influence on loan performance is complex, depending on what purpose the borrowers want to achieve.

5 Robustness

5.1 Alternative Measures of Default

Borrowers could default their mortgage for many reasons in different time horizons. Our measure of default uses MBA method of delinquency for 90 days or more and is restricted to the range of the first three years after origination. By changing the definition more restricted (less default cases) or less restricted (more default cases), we can investigate whether the effect found in section 4 is generally held or not. Therefore, we try the following default measures: 90 days or more delinquency using MBA method in the first two years after origination (more restricted), 60 days or more delinquency using MBA method in the first three years after origination (less restricted), and bankruptcy/foreclosure/REO in the first three years after origination (more restricted).

Table 11 presents the results of different default measures. We report the results for second-lien misrepresentation in panel A and the results for owner-occupancy misreporting in panel B. We also include the results of the original measure in column (1). The results show that the effect of mortgage fraud (i.e., coefficient of mortgage fraud) is greater when the default measure is less restricted. The effect of gambling preference associated mortgage misrepresentation (i.e., coefficient of the interaction term) is also greater when the default measure is less restricted. While such effect (i.e., coefficient of the interaction term) holds widely for second-lien misrepresentation, it is not significant for owner-occupancy misreporting when the default measure is too much restricted (i.e., consider only bankruptcy/foreclosure/REO).

5.2 Nonlinear Model

The results in previous sections are estimated using linear probability model (OLS). To ensure our results are not driven by this modeling choice, we also use nonlinear specification (probit) for the inference. We present the results for the relationship between gambling preference and mortgage misrepresentation in Table 12. Column (1) reports the results for second-lien misrepresentation and column (2) presents the results for owner-occupancy misreporting. The marginal effects of CPRATIO remain significant and the economic magnitudes are even larger.

Since it could be problematic to interpret the marginal effect of an interaction term in a nonlinear model (Ai and Norton, 2003; Buis, 2010; Williams, 2012), we estimate the effect of mortgage misrepresentation on default at different levels of CPRATIO following Williams (2012). Specifically, we divide the sample into quarters by CPRATIO and estimate probit specification with all variables used in Table 7 or Table 9 column (3). The top panel of Figure 2 shows the results for second-lien misrepresentation. The marginal effect of second-lien misrepresentation on default is more negative at higher levels of CPRATIO, which is consistent with our linear model results. The bottom panel presents the results for owner-occupancy misreporting. The effect is amplified at areas with higher levels of gambling preference, which is also consistent with our linear model results.

5.3 Causal Forest and Causality

To enhance the causal inference between gambling preference and mortgage misrepresentation and to explore the heterogeneous treatment effects of mortgage misrepresentation on mortgage default conditioned on gambling preference, we employ causal forest approach proposed by Wager and Athey (2018). In general, it uses augmented inverse propensity-weighted estimator by random forest method in machine learning with honesty condition, providing double robustness property (compared to propensity score matching) with high efficiency for high dimensional models (compared to nearest-neighbor matching) in observational settings[18]. It is also used in finance area, such as Rampini and Viswanathan (2022) for secured debt and Gulen et al. (2021) for corporate finance, and is shown to have better performance than traditional causal inference design (Gulen et al., 2021) in estimation from the perspective of accuracy.

We first use this approach to investigate the causality between gambling preference and mortgage misrepresentation. To use causal forest approach, we set mortgage misrepresentation (i.e., second-lien misrepresentation or owner-occupancy misreporting) as the outcome and CPRATIO as the treatment. We use all control variables in Table 12 as matching variables to help grow trees and forests[19]. All continuous variables, including CPRATIO, are winsorized at 1 percent level and then standardized. Since fixed effects are not applicable in this approach, we create a variable for half-year, starting from the first half of 2005 as 1, to account for the potential effect from time. Similar to Athey and Wager (2019) who clus-

[18]See Wager and Athey (2018) for statistical illustration. See Athey and Wager (2019) for example application.

[19]We require observations to be nonmissing for all variables.

ter observations by school ID, we cluster observations by state but give each unit the same weight so that bigger clusters get more weight[20]. By doing such cross-fitting, our results does not solely come from any single state but the states with greater amount of observations do have greater weights. To balance the calculation capacity and the accuracy of confidence interval, we grow 2000 trees for the forest, which is also the default setting as in Athey and Wager (2019). For honesty property, we set the splitting fraction as default option: for each sample, we use 50 percent of data for splitting and the left data for estimation. Table 13 shows the results using causal forest. The coefficients in regressions of both second-lien misrepresentation and owner-occupancy misreporting remain significant and the magnitudes are similar to the ones estimated by OLS.

Second, we study the heterogeneous treatment effects of mortgage misrepresentation on default conditioned on different levels of gambling preference. We set default as the outcome and mortgage misrepresentation as treatment. Except for adding CPRATIO into matching variable matrix, the other matching variables and clustering variable are the same as the ones in above causal forest regressions. For same consideration, we also grow 2000 trees for the forest and use 50 percent of data in each sample for splits. After the estimation of causal forest, the average treatment effects of mortgage misrepresentation on default are calculated for each quarter of sample sorted by CPRATIO. The top panel of Figure 3 shows the results for second-lien misrepresentation. The treatment effects are generally more negative when CPRATIO is greater, which is consistent with the results from OLS and probit model. However, the difference of results between causal forest and probit is that the treatment effects are all negative when estimated by causal forest even when CPRATIO is low. The bottom panel represents the results for owner-occupancy misreporting. The treatment effects are all positive and generally increase with CPRATIO, which are also consistent with the findings from OLS and probit model.

[20]Due to the limitation of grf package in R, which requires the matching variables to be numerical, we do not create a numerical variable for originators so that the process would not falsely treat closer values as shorter distance. We also do not cluster observations by originator because the large number of originator would excessively boost the calculation burden of the process.

6 Conclusion

Mortgage fraud is a severe issue in economics and contribute a lot on the 2008 financial crisis. While the motivation, prevalence, and economic impact are well investigated, the mechanism of turning motivation to actual fraudulent behavior has not been explored. We fill this gap by empirically examining a preference-based explanation to help explain why borrowers decide to commit mortgage fraud even if the penalty for such fraud is large.

Using a religion-based proxy of local gambling preferences, we find that mortgage misrepresentation is more likely to occur in areas with higher levels of local gambling preference. In addition, the effect of mortgage fraud on loan performance is amplified at areas with higher levels of gambling preference. Therefore, gambling preference is an important factor that making borrowers commit mortgage fraud. Our findings encourage further work on theoretical model in explaining the trade-off between benefits and costs when households making decision in mortgage market.

References

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

Alesina, A., La Ferrara, E., 2000. Local gambling preferences and corporate innovative success. Quarterly Journal of Economics 115, 847–904.

Ambrose, B.W., Conklin, J., Yoshida, J., 2016. Credit rationing, income exaggeration, and adverse selection in the mortgage market. Journal of Finance 71, 2637–2686.

Athey, S., Wager, S., 2019. Estimating treatment effects with causal forests: An application. Observational Studies 5, 37–51.

Baele, L., Driessen, J., Ebert, S., Londono, J.M., Spalt, O.G., 2019. Cumulative prospect theory, option returns, and the variance premium. Review of Financial Studies 32, 3667–3723.

Barberis, N., 2013. The psychology of tail events: Progress and challenges. American Economic Review 103, 611–616.

Barberis, N., Huang, M., 2008. Stocks as lotteries: The implications of probability weighting for security prices. American Economic Review 98, 2066–2100.

Barberis, N., Jin, L.J., Wang, B., 2021. Prospect theory and stock market anomalies. Journal of Finance 76, 2639–2687.

Barberis, N., Mukherjee, A., Wang, B., 2016. Prospect theory and stock returns: An empirical test. Review of Financial Studies 29, 3068–3107.

Benartzi, S., Thaler, R.H., 1995. Myopic loss aversion and the equity premium puzzle. Quarterly Journal of Economics 110, 73–92.

Buis, M.L., 2010. Stata tip 87: Interpretation of interactions in nonlinear models. The Stata Journal 10, 305–308.

Campbell, J.Y., 2006. Household finance. Journal of Finance 61, 1553–1604.

Chen, Y., Podolski, E.J., Rhee, S.G., Veeraraghavan, M., 2014. Local gambling preferences and corporate innovative success. Journal of Financial and Quantitative Analysis 49, 77–106.

Conklin, J., Diop, M., Qiu, M., 2022. Religion and mortgage misrepresentation. Journal of Business Ethics 179, 273–295.

Elul, R., Souleles, N.S., Chomsisengphet, S., Glennon, D., Hunt, R., 2010. What "triggers" mortgage default? American Economic Review 100, 490–494.

Federal Housing Finance Agency, 2023. Fraud prevention. URL: `https://www.fhfa.gov/PolicyProgramsResearch/Programs/Pages/Fraud-Prevention.asp`.

Griffin, J.M., Maturana, G., 2016. Who facilitated misreporting in securitized loans? Review of Financial Studies 29, 384–419.

Gulen, H., Jens, C.E., Page, T.B., 2021. The heterogeneous effects of default on investment: An application of causal forest in corporate finance. Working paper.

Han, B., Kumar, A., 2013. Speculative retail trading and asset prices. Journal of Financial and Quantitative Analysis 48, 377–404.

Henning, A.C., 2009. Mortgage Fraud: Federal Criminal Provisions. Technical Report. Congressional Research Service.

Hilary, G., Hui, K.W., 2009. Does religion matter in corporate decision making in america? Journal of Financial Economics 93, 455–473.

Kahneman, D., Tversky, A., 1979. Prospect theory: An analysis of decision under risk. Econometrica 47, 263–291.

Kruger, S., Maturana, G., 2021. Collateral misreporting in the residential mortgage-backed security market. Management Science 67, 2729–2750.

Kumar, A., 2009. Who gambles in the stock market? Journal of Finance 64, 1889–1933.

Kumar, A., Page, J.K., Spalt, O.G., 2011. Religious beliefs, gambling attitudes, and financial market outcomes. Journal of Financial Economics 102, 671–708.

Mian, A., Suf, A., 2017. Fraudulent income overstatement on mortgage applications during the credit expansion of 2002 to 2005. Review of Financial Studies 30, 1832–1864.

Piskorski, T., Seru, A., Witkin, J., 2015. Asset quality misrepresentation by financial intermediaries: Evidence from the rmbs market. Journal of Finance 70, 2635–2678.

Rampini, A.A., Viswanathan, S., 2022. Collateral and secured debt. Working paper.

Shu, T., Sulaeman, J., Yeung, P.E., 2012. Local religious beliefs and mutual fund risk-taking behaviors. Management Science 58, 1779–1796.

Tversky, A., Kahneman, D., 1992. Advances in prospect theory: Cumulative representation of uncertainty. Journal of Risk and Uncertainty 5, 297–323.

United States Sentencing Commission, 2015. Quick Facts: Mortgage Fraud Offenses. Technical Report. United States Sentencing Commission.

Wager, S., Athey, S., 2018. Estimation and inference of heterogeneous treatment effects using random forests. Journal of the American Statistical Association 113, 1228–1242.

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

Zhang, H.H., Zhao, F., Zhao, X., 2024. Asset quality misrepresentation by financial intermediaries: Evidence from the rmbs market. Journal of Finance 79, 129–172.

Table 1
Variable Definitions

Variable name	Definition
Mortgage misrepresentation	
Second-lien misrepresentation	Indicator that equals one if the borrower misrepresented second-lien on the loan application
Owner-occupancy misreporting	Indicator that equals one if the borrower misreported occupancy status on the loan application
Gambling preference and religiosity	
CPRATIO	Ratio of the county's Catholic residents to Protestant residents
REL	Proportion of the county population that are religious adherents
Geographic controls	
HPA	Zip code house price appreciation in the two years prior to loan origination year. County level data used if zip code index is not available
Education	Proportion of the county population over age 25 that has completed a bachelor's degree or higher
Married	Proportion of the county population over age 15 that is married
Income	Natural logarithm of the zip code level median household income
Urban	Proportion of the county population that lives in urban area
Total population	Natural logarithm of the county total population
Over65	Proportion of the county population over age 65
Male-female ratio	Ratio of the county's male residents to female residents
Minority	Proportion of the county non-white residents
Loan characteristics	
Interest rate	Loan interest rate at origination
FICO	Natural logarithm of borrower's FICO credit score at loan origination
Balance	Natural logarithm of the initial loan balance
LTV	Loan-to-value ratio at loan origination
ARM	Indicator that equals one if the loan is an adjustable rate mortgage
Option ARM	Indicator that equals one if the loan has an ARM convertibility clause
Negative amortization	Indicator that equals one if the loan allows negative amortization
Low or no doc.	Indicator that equals one if the loan is originated with no or limited documentation
Prepayment penalty	Indicator that equals one if the loan would be assessed a penalty on any early voluntary prepayment
Cash-out	Indicator that equals one if the loan purpose is cash out refinancing
No-cash-out	Indicator that equals one if the loan purpose is no cash out refinancing
Investment	Indicator that equals one if the loan occupancy status is investment
Second-home	Indicator that equals one if the loan occupancy status is second-home
Default	Indicator that equals one if the loan becomes 90 days or more delinquent using MBA method in the first three years after origination

The table reports the variable definitions used in the empirical analysis part.

Table 2
Descriptive Statistics

	Panel A. Sample of Loans for Second-lien Misrepresentation						Panel B. Sample of Loans for Owner-occupancy Misreporting					
variable	N	mean	sd	p25	p50	p75	N	mean	sd	p25	p50	p75
Misrepresentation	633904	0.34					646457	0.35				
CPRATIO	633904	1.31	1.12	0.44	0.94	1.86	646457	1.51	1.16	0.52	1.32	2.14
REL	633904	0.51	0.12	0.42	0.49	0.59	646457	0.51	0.11	0.42	0.47	0.58
HPA	633397	1.26	0.18	1.10	1.25	1.41	645851	1.28	0.17	1.13	1.27	1.40
Education (%)	633904	27.10	9.20	21.30	25.90	31.20	646457	27.49	8.94	21.70	25.90	31.20
Married (%)	633904	56.43	5.20	53.10	56.80	59.90	646457	55.80	5.23	52.70	56.20	59.00
Income (ln)	606094	10.79	0.33	10.56	10.79	11.02	602073	10.82	0.36	10.57	10.83	11.07
Urban	633904	0.89	0.16	0.88	0.96	0.99	646457	0.91	0.15	0.90	0.97	0.99
Total population (ln)	633904	13.45	1.37	12.57	13.56	14.34	646457	13.58	1.37	12.75	13.68	14.37
Over65 (%)	633904	11.23	3.37	9.50	10.90	12.60	646457	11.86	3.65	9.70	11.20	13.20
Male-female ratio	633904	0.97	0.04	0.94	0.97	1.00	646457	0.97	0.04	0.94	0.97	0.99
Minority	633904	0.29	0.15	0.17	0.28	0.41	646457	0.30	0.15	0.18	0.29	0.41
Interest rate (%)	633822	6.84	1.58	6.25	6.88	7.69	645868	6.28	2.28	5.88	6.63	7.63
FICO (ln)	622411	6.53	0.08	6.47	6.53	6.59	634881	6.56	0.09	6.51	6.56	6.62
Balance (ln)	633904	12.26	0.67	11.77	12.24	12.78	646457	12.28	0.77	11.74	12.26	12.86
LTV (%)	633904	78.07	8.43	80.00	80.00	80.00	646457	75.69	12.61	70.10	80.00	80.00
ARM	633904	0.64					646457	0.62				
Option ARM	633904	0.00					646457	0.00				
Negative amortization	633904	0.07					646457	0.19				
Low or no doc.	633904	0.68					646457	0.76				
Prepayment penalty	633904	0.45					646457	0.40				
Cash-out	633904	0.12					646457	0.25				
No-cash-out	633904	0.10					646457	0.11				
Investment	633904	0.09					646457	0.49				
Second-home	633904	0.03					646457	0.17				
Default	632534	0.37					644070	0.26				

The table shows descriptive statistics for the variables used in our study. Panel A is for second-lien misrepresentation and Panel B is for owner-occupancy misreporting. Both sample periods are from 2005 to 2007. We report the number of observations, mean, standard deviation, 25th percentile, 50th percentile, and 75th percentile for all continuous variables and the number of observation and mean for all dummy variables. The variables are winsorized at 1 percent level. (ln) indicates the value of the variable is the natural logarithm of the original value. (%) indicates the value of the variable is in percentage.

Table 3
Gambling Preference and Second-lien Misrepresentation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
CPRATIO	0.006***	0.019***	0.014***	0.016***	0.004***	0.005***	0.005***	0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
REL		-0.011***		-0.009***	-0.002*	-0.001	-0.001	-0.001
		(0.001)		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
HPA		-0.043***		-0.026***	-0.007***	0.004***	0.004*	0.002
		(0.001)		(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
Education		-0.004***		0.000	0.004***	0.001*	0.001	0.001
		(0.001)		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Married		0.002***		0.003***	-0.004***	0.001	0.001	0.001
		(0.001)		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Income		-0.000		-0.003***	-0.001	-0.001	-0.001	-0.002**
		(0.001)		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Urban		0.001		-0.001	-0.004***	-0.001	-0.001	-0.001
		(0.001)		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Total population		0.007***		0.011***	0.006***	0.004***	0.004***	0.004**
		(0.001)		(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
Over65		0.016***		0.016***	0.006***	0.002*	0.002	0.001
		(0.001)		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Male-female ratio		0.002**		0.004***	-0.000	-0.000	-0.000	-0.000
		(0.001)		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Minority		0.005***		0.007***	0.002	0.003***	0.003**	0.003**
		(0.001)		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Interest rate			-0.036***	-0.033***	-0.015***	-0.011***	-0.011***	-0.004***
			(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
FICO			-0.052***	-0.051***	-0.047***	-0.035***	-0.035***	-0.029***
			(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Balance			-0.007***	-0.000	0.003***	0.005***	0.005***	0.012***
			(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
LTV			0.013***	0.012***	0.016***	0.061***	0.061***	0.062***
			(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
ARM			-0.001	0.000	-0.016***	0.050***	0.050***	0.040***
			(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
Option ARM			-0.248***	-0.256***	-0.285***	-0.267***	-0.267***	-0.105***
			(0.012)	(0.013)	(0.013)	(0.012)	(0.013)	(0.013)
Negative amortization			-0.142***	-0.134***	-0.083***	0.062***	0.062***	0.080***
			(0.003)	(0.003)	(0.003)	(0.003)	(0.005)	(0.005)
Low or no doc.			-0.003*	-0.002	-0.011***	-0.030***	-0.030***	-0.028***
			(0.001)	(0.001)	(0.001)	(0.001)	(0.003)	(0.003)
Prepayment penalty			-0.043***	-0.044***	-0.054***	-0.041***	-0.041***	-0.036***
			(0.001)	(0.001)	(0.001)	(0.001)	(0.003)	(0.003)
Cash-out			0.133***	0.133***	0.137***	0.133***	0.133***	0.133***
			(0.002)	(0.002)	(0.002)	(0.002)	(0.004)	(0.004)
No-cash-out			0.060***	0.059***	0.075***	0.062***	0.062***	0.064***
			(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
Investment			0.121***	0.122***	0.114***	0.152***	0.152***	0.158***
			(0.002)	(0.002)	(0.002)	(0.003)	(0.005)	(0.005)
Second-home			0.048***	0.047***	0.046***	0.056***	0.056***	0.064***
			(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)
State FE	N	N	N	N	Y	Y	Y	Y
Year FE	N	N	N	N	Y	Y	Y	N
Half-year FE	N	N	N	N	N	N	N	Y
Originator FE	N	N	N	N	N	Y	Y	Y
SEs Clustered by County	N	N	N	N	N	N	Y	Y
Observations	633,904	605,604	622,373	594,582	594,582	479,775	479,775	479,775
Adj. R^2	0.000	0.005	0.030	0.031	0.043	0.266	0.266	0.301

The table shows OLS estimates from regressions in which the dependent variable takes a value of one if the loan is recognized as second-lien misrepresented and zero otherwise. Specific fixed effects and clustering method are used if Y and N otherwise. Variables are defined in Table 1. All continuous variables are winsorized at 1 percent level and then standardized. Standard errors are in parentheses $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

Table 4
Gambling Preference and Second-lien Misrepresentation - Subsamples

	(1) Primary	(2) Non-primary	(3) Purchase	(4) Refinance	(5) Full Doc	(6) Low/No Doc	(7) High FICO	(8) Low FICO
CPRATIO	0.004***	0.001	0.004***	0.002	0.007***	0.004**	0.004	0.006***
	(0.002)	(0.005)	(0.001)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
REL	-0.001	-0.002	-0.002	0.001	0.003*	-0.004***	-0.002	-0.001
	(0.001)	(0.004)	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)	(0.001)
HPA	0.001	0.003	-0.001	0.013***	0.009***	0.001	0.000	0.006***
	(0.002)	(0.005)	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
Education	0.001	-0.009***	0.001	-0.000	-0.004**	0.003**	0.000	0.001
	(0.001)	(0.003)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
Married	0.001	-0.003	0.001	-0.000	0.001	0.000	0.000	0.001
	(0.001)	(0.004)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
Income	-0.002***	0.006**	-0.000	-0.005**	-0.001	-0.002*	0.001	-0.003**
	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Urban	-0.002	0.007	-0.001	-0.003	-0.004**	0.000	0.001	-0.003*
	(0.001)	(0.004)	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)
Total population	0.004***	-0.001	0.003**	0.005	0.006***	0.004**	0.004**	0.004**
	(0.001)	(0.005)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
Over65	0.001	0.003	-0.000	0.006**	-0.001	0.001	0.000	0.001
	(0.001)	(0.003)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
Male-female ratio	-0.001	0.006	-0.001	0.002	-0.002	0.001	0.001	-0.001
	(0.001)	(0.003)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
Minority	0.004**	-0.004	0.003**	0.005	0.003	0.003	0.002	0.004**
	(0.002)	(0.005)	(0.001)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
Interest rate	-0.006***	0.003	0.002**	-0.012***	-0.002	-0.006***	-0.004***	-0.002
	(0.001)	(0.003)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
FICO	-0.028***	-0.038***	-0.029***	-0.020***	-0.042***	-0.019***	-0.006***	-0.051***
	(0.001)	(0.003)	(0.001)	(0.003)	(0.002)	(0.001)	(0.002)	(0.002)
Balance	0.014***	-0.010***	0.006***	0.022***	0.001	0.017***	0.018***	-0.003
	(0.001)	(0.003)	(0.001)	(0.003)	(0.002)	(0.001)	(0.001)	(0.002)
LTV	0.057***	0.094***	0.077***	0.035***	0.059***	0.069***	0.043***	0.078***
	(0.001)	(0.005)	(0.002)	(0.004)	(0.002)	(0.002)	(0.003)	(0.001)
ARM	0.051***	-0.058***	0.043***	0.009**	0.049***	0.029***	0.052***	0.008***
	(0.002)	(0.007)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)
Option ARM	-0.099***	-0.106**	-0.103***	-0.106***	-0.610***	-0.143***	-0.151***	-0.053***
	(0.013)	(0.047)	(0.010)	(0.026)	(0.146)	(0.013)	(0.016)	(0.014)
Negative amortization	0.055***	0.182***	0.055***	0.123***	0.003	0.080***	0.091***	0.085***
	(0.005)	(0.012)	(0.006)	(0.006)	(0.011)	(0.006)	(0.006)	(0.008)
Low or no doc.	-0.029***	-0.036***	-0.032***	-0.024***			-0.042***	0.004
	(0.003)	(0.006)	(0.003)	(0.004)			(0.003)	(0.003)
Prepayment penalty	-0.036***	-0.024***	-0.026***	-0.058***	-0.072***	-0.018***	-0.024***	-0.054***
	(0.003)	(0.004)	(0.003)	(0.004)	(0.004)	(0.004)	(0.003)	(0.004)
Cash-out	0.118***	0.317***			0.118***	0.149***	0.184***	0.090***
	(0.004)	(0.009)			(0.004)	(0.004)	(0.004)	(0.003)
No-cash-out	0.053***	0.209***			0.061***	0.070***	0.083***	0.035***
	(0.003)	(0.010)			(0.004)	(0.003)	(0.003)	(0.004)
Investment			0.104***	0.288***	0.194***	0.154***	0.137***	0.233***
			(0.005)	(0.007)	(0.008)	(0.005)	(0.005)	(0.008)
Second-home			0.055***	0.130***	0.056***	0.062***	0.060***	0.063***
			(0.005)	(0.012)	(0.012)	(0.004)	(0.005)	(0.012)
State FE	Y	Y	Y	Y	Y	Y	Y	Y
Half-year FE	Y	Y	Y	Y	Y	Y	Y	Y
Originator FE	Y	Y	Y	Y	Y	Y	Y	Y
Observations	440,112	39,468	370,576	108,899	171,967	307,496	276,121	203,472
Adj. R^2	0.302	0.362	0.310	0.292	0.327	0.317	0.280	0.358

The table shows OLS estimates from regressions in which the dependent variable takes a value of one if the loan is recognized as second-lien misrepresented and zero otherwise. We divide the whole samples in several ways: primary or non-primary (column (1) and (2)), purchase or refinance (column (3) and (4)), full or low/no documentation (column (5) and (6)), and high or low credit score (column (7) and (8)). The high or low credit score subsamples are divided by FICO score of 670. Variables are defined in Table 1. All continuous variables are winsorized at 1 percent level and then standardized. All specifications include state, half-year, and originator fixed effects, and standard errors are clustered at the county level are reported in the parentheses. $^*p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

Table 5
Gambling Preference and Owner-occupancy Misreporting

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
CPRATIO	0.037***	0.026***	0.008***	0.006***	0.013***	0.010***	0.010*	0.010*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.006)	(0.005)
REL		-0.007***		-0.003***	-0.009***	-0.009***	-0.009**	-0.009**
		(0.001)		(0.001)	(0.001)	(0.001)	(0.003)	(0.003)
HPA		0.021***		0.001	0.007***	0.002**	0.002	0.002
		(0.001)		(0.001)	(0.001)	(0.001)	(0.004)	(0.004)
Education		0.024***		0.007***	0.012***	0.010***	0.010***	0.010***
		(0.001)		(0.001)	(0.001)	(0.001)	(0.003)	(0.003)
Married		0.014***		0.013***	0.014***	0.011***	0.011**	0.012**
		(0.001)		(0.001)	(0.001)	(0.001)	(0.005)	(0.005)
Income		-0.063***		-0.068***	-0.068***	-0.063***	-0.063***	-0.063***
		(0.001)		(0.001)	(0.001)	(0.001)	(0.003)	(0.003)
Urban		-0.019***		-0.014***	-0.018***	-0.017***	-0.017***	-0.017***
		(0.001)		(0.001)	(0.001)	(0.001)	(0.003)	(0.003)
Total population		0.034***		0.024***	0.022***	0.021***	0.021***	0.021***
		(0.001)		(0.001)	(0.001)	(0.001)	(0.006)	(0.006)
Over65		-0.015***		-0.017***	-0.014***	-0.012***	-0.012***	-0.012***
		(0.001)		(0.001)	(0.001)	(0.001)	(0.004)	(0.004)
Male-female ratio		0.001		0.000	-0.002	-0.001	-0.001	-0.001
		(0.001)		(0.001)	(0.001)	(0.001)	(0.005)	(0.005)
Minority		0.023***		0.012***	0.022***	0.019***	0.019***	0.019***
		(0.001)		(0.001)	(0.001)	(0.001)	(0.006)	(0.006)
Interest rate			-0.027***	-0.027***	-0.030***	-0.041***	-0.041***	-0.040***
			(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
FICO			-0.115***	-0.106***	-0.105***	-0.091***	-0.091***	-0.090***
			(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
Balance			0.128***	0.131***	0.133***	0.129***	0.129***	0.130***
			(0.001)	(0.001)	(0.001)	(0.001)	(0.006)	(0.006)
LTV			0.028***	0.026***	0.025***	0.025***	0.025***	0.025***
			(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
ARM			0.010***	0.018***	0.020***	0.025***	0.025***	0.025***
			(0.001)	(0.001)	(0.001)	(0.002)	(0.003)	(0.003)
Option ARM			0.006	0.004	0.002	-0.029**	-0.029*	-0.012
			(0.011)	(0.011)	(0.011)	(0.012)	(0.015)	(0.016)
Negative amortization			-0.153***	-0.142***	-0.146***	-0.247***	-0.247***	-0.245***
			(0.002)	(0.002)	(0.002)	(0.003)	(0.006)	(0.006)
Low or no doc.			0.028***	0.027***	0.029***	0.042***	0.042***	0.043***
			(0.001)	(0.001)	(0.001)	(0.002)	(0.004)	(0.004)
Prepayment penalty			0.039***	0.027***	0.026***	0.011***	0.011**	0.013***
			(0.001)	(0.001)	(0.001)	(0.002)	(0.004)	(0.004)
Cash-out			-0.144***	-0.159***	-0.158***	-0.174***	-0.174***	-0.175***
			(0.001)	(0.001)	(0.001)	(0.002)	(0.005)	(0.005)
No-cash-out			-0.107***	-0.116***	-0.117***	-0.120***	-0.120***	-0.122***
			(0.002)	(0.002)	(0.002)	(0.002)	(0.005)	(0.005)
State FE	N	N	N	N	Y	Y	Y	Y
Year FE	N	N	N	N	Y	Y	Y	N
Half-year FE	N	N	N	N	N	N	N	Y
Originator FE	N	N	N	N	N	Y	Y	Y
SEs Clustered by County	N	N	N	N	N	N	Y	Y
Observations	646,457	601,513	634,633	590,467	590,467	466,916	466,916	466,916
Adj. R^2	0.006	0.031	0.131	0.154	0.156	0.196	0.196	0.197

The table shows OLS estimates from regressions in which the dependent variable takes a value of one if the loan is recognized as owner-occupancy misreported and zero otherwise. Specific fixed effects and clustering method are used if Y and N otherwise. Variables are defined in Table 1. All continuous variables are winsorized at 1 percent level and then standardized. Standard errors are in parentheses $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

Table 6
Gambling Preference and Owner-occupancy Misreporting - Subsamples

	(1) Purchase	(2) Refinance	(3) Full Doc	(4) Low/No Doc	(5) High FICO	(6) Low FICO
CPRATIO	0.006	0.015***	0.009	0.010*	0.013**	0.002
	(0.006)	(0.005)	(0.007)	(0.005)	(0.005)	(0.006)
REL	-0.009**	-0.006**	-0.005	-0.011***	-0.011***	-0.004
	(0.004)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)
HPA	0.004	-0.003	-0.006	0.006	0.007	-0.003
	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Education	0.011***	0.008***	0.005	0.012***	0.013***	0.000
	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)
Married	0.014***	0.007	0.002	0.014***	0.013***	0.006
	(0.005)	(0.004)	(0.005)	(0.005)	(0.005)	(0.004)
Income	-0.073***	-0.037***	-0.055***	-0.064***	-0.065***	-0.054***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)
Urban	-0.006	-0.028***	-0.023***	-0.013***	-0.011***	-0.025***
	(0.005)	(0.003)	(0.004)	(0.004)	(0.003)	(0.004)
Total population	0.020***	0.017***	0.024***	0.019***	0.020***	0.025***
	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Over65	-0.016***	-0.003	-0.013**	-0.012***	-0.014***	-0.012**
	(0.004)	(0.004)	(0.006)	(0.004)	(0.004)	(0.005)
Male-female ratio	-0.000	-0.001	-0.001	-0.002	-0.001	-0.002
	(0.005)	(0.004)	(0.004)	(0.005)	(0.005)	(0.004)
Minority	0.024***	0.010*	0.002	0.024***	0.024***	0.004
	(0.007)	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)
Interest rate	-0.042***	-0.047***	-0.065***	-0.036***	-0.042***	-0.071***
	(0.003)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)
FICO	-0.101***	-0.077***	-0.113***	-0.075***	-0.047***	-0.121***
	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.004)
Balance	0.137***	0.117***	0.105***	0.137***	0.130***	0.126***
	(0.008)	(0.003)	(0.004)	(0.007)	(0.007)	(0.004)
LTV	0.031***	0.029***	0.015***	0.032***	0.032***	0.019***
	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)
ARM	0.040***	-0.019***	0.035***	0.012***	-0.000	0.054***
	(0.003)	(0.003)	(0.005)	(0.003)	(0.002)	(0.004)
Option ARM	0.001	-0.026	0.057	-0.017	0.008	-0.026*
	(0.019)	(0.020)	(0.068)	(0.018)	(0.023)	(0.016)
Negative amortization	-0.304***	-0.140***	-0.324***	-0.229***	-0.219***	-0.321***
	(0.009)	(0.005)	(0.009)	(0.006)	(0.005)	(0.009)
Low or no doc.	0.044***	0.036***			0.073***	0.031***
	(0.005)	(0.003)			(0.004)	(0.005)
Prepayment penalty	0.027***	-0.022***	0.063***	-0.004	-0.005	0.022***
	(0.006)	(0.003)	(0.007)	(0.004)	(0.004)	(0.007)
Cash-out			-0.210***	-0.158***	-0.128***	-0.269***
			(0.008)	(0.004)	(0.004)	(0.008)
No-cash-out			-0.142***	-0.111***	-0.088***	-0.222***
			(0.006)	(0.004)	(0.004)	(0.007)
State FE	Y	Y	Y	Y	Y	Y
Half-year FE	Y	Y	Y	Y	Y	Y
Originator FE	Y	Y	Y	Y	Y	Y
Observations	288,588	177,976	118,465	348,168	338,805	127,911
Adj. R^2	0.211	0.149	0.214	0.201	0.178	0.213

The table shows OLS estimates from regressions in which the dependent variable takes a value of one if the loan is recognized as owner-occupancy misreported and zero otherwise. We divide the whole samples in several ways: purchase or refinance (column (1) and (2)), full or low/no documentation (column (3) and (4)), and high or low credit score (column (5) and (6)). The high or low credit score subsamples are divided by FICO score of 670. Variables are defined in Table 1. All continuous variables are winsorized at 1 percent level and then standardized. All specifications include state, half-year, and originator fixed effects, and standard errors clustered at the county level are reported in the parentheses. $^*p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

Table 7
Effect of Gambling Preference Associated Second-lien Misrepresentation on Delinquency

	(1)	(2)	(3)	(4)
Second-lien misrepresentation	-0.018***		-0.018***	-0.018***
	(0.003)		(0.003)	(0.002)
CPRATIO		-0.005	-0.005	-0.002
		(0.006)	(0.006)	(0.005)
Second-lien misrepresentation × CPRATIO				-0.011***
				(0.003)
REL	-0.012**	-0.011***	-0.011***	-0.011***
	(0.005)	(0.004)	(0.004)	(0.004)
HPA	0.075***	0.075***	0.075***	0.075***
	(0.005)	(0.005)	(0.005)	(0.005)
Education	0.003	0.002	0.002	0.002
	(0.004)	(0.004)	(0.004)	(0.004)
Married	0.020***	0.020***	0.020***	0.020***
	(0.004)	(0.004)	(0.004)	(0.004)
Income	-0.019***	-0.019***	-0.019***	-0.019***
	(0.002)	(0.002)	(0.002)	(0.002)
Urban	0.013**	0.013**	0.013**	0.013**
	(0.006)	(0.005)	(0.005)	(0.005)
Total population	-0.003	-0.002	-0.002	-0.002
	(0.009)	(0.008)	(0.008)	(0.008)
Over65	0.003	0.003	0.003	0.003
	(0.004)	(0.004)	(0.004)	(0.004)
Male-female ratio	0.000	0.000	0.000	0.000
	(0.004)	(0.004)	(0.004)	(0.004)
Minority	0.019***	0.019***	0.019***	0.020***
	(0.005)	(0.005)	(0.005)	(0.005)
Interest rate	0.046***	0.046***	0.046***	0.046***
	(0.002)	(0.002)	(0.002)	(0.002)
FICO	-0.108***	-0.108***	-0.108***	-0.108***
	(0.002)	(0.002)	(0.002)	(0.002)
Balance	0.041***	0.041***	0.042***	0.042***
	(0.002)	(0.002)	(0.002)	(0.002)
LTV	0.009***	0.008***	0.009***	0.009***
	(0.001)	(0.001)	(0.001)	(0.001)
ARM	0.068***	0.068***	0.068***	0.068***
	(0.004)	(0.004)	(0.004)	(0.004)
Option ARM	0.009	0.011	0.009	0.008
	(0.014)	(0.014)	(0.014)	(0.014)
Negative amortization	0.035***	0.033***	0.035***	0.034***
	(0.007)	(0.007)	(0.007)	(0.007)
Low or no doc.	0.074***	0.075***	0.074***	0.074***
	(0.004)	(0.004)	(0.004)	(0.004)
Prepayment penalty	0.068***	0.069***	0.068***	0.068***
	(0.004)	(0.005)	(0.005)	(0.004)
Cash-out	-0.033***	-0.036***	-0.033***	-0.033***
	(0.003)	(0.003)	(0.003)	(0.003)
No-cash-out	-0.010***	-0.011***	-0.010***	-0.010***
	(0.004)	(0.004)	(0.004)	(0.004)
Investment	0.050***	0.048***	0.050***	0.050***
	(0.008)	(0.008)	(0.008)	(0.008)
Second-home	0.012	0.011	0.012	0.012
	(0.008)	(0.008)	(0.008)	(0.008)
State FE	Y	Y	Y	Y
Half-year FE	Y	Y	Y	Y
Originator FE	Y	Y	Y	Y
Observations	479,341	479,341	479,341	479,341
Adj. R^2	0.244	0.244	0.244	0.244

The table shows OLS estimates from regressions in which the dependent variable takes a value of one if the loan becomes 90 days or more delinquent using MBA method in the first three years after origination and zero otherwise. All continuous variables are winsorized at 1 percent level and then standardized. All specifications include state, half-year, and originator fixed effects, and standard errors clustered at the county level are reported in the parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

Table 8
Effect of Gambling Preference Associated Second-lien Misrepresentation on Delinquency - Subsamples

	(1) Primary	(2) Non-primary	(3) Purchase	(4) Refinance	(5) Full Doc	(6) Low/No Doc	(7) High FICO	(8) Low FICO
Second-lien misrepresentation	-0.018***	-0.021***	-0.013***	-0.038***	-0.001	-0.029***	-0.026***	-0.007**
	(0.002)	(0.006)	(0.002)	(0.004)	(0.003)	(0.003)	(0.002)	(0.003)
CPRATIO	-0.001	0.002	-0.004	0.005	-0.001	-0.002	-0.001	-0.001
	(0.005)	(0.007)	(0.005)	(0.006)	(0.005)	(0.005)	(0.006)	(0.005)
Second-lien misrepresentation	-0.010***	-0.016***	-0.011***	-0.014***	-0.007	-0.011***	-0.012***	-0.013***
× CPRATIO	(0.004)	(0.005)	(0.004)	(0.003)	(0.004)	(0.004)	(0.003)	(0.004)
REL	-0.011***	-0.009	-0.011***	-0.009**	-0.009**	-0.011**	-0.011***	-0.009**
	(0.004)	(0.006)	(0.004)	(0.005)	(0.004)	(0.005)	(0.004)	(0.004)
HPA	0.078***	0.040***	0.076***	0.075***	0.071***	0.077***	0.074***	0.074***
	(0.005)	(0.007)	(0.006)	(0.005)	(0.006)	(0.005)	(0.005)	(0.007)
Education	0.003	-0.006	0.002	0.004	-0.000	0.003	-0.001	0.009**
	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)	(0.005)	(0.004)	(0.004)
Married	0.020***	0.013***	0.022***	0.013***	0.012***	0.024***	0.022***	0.015***
	(0.004)	(0.005)	(0.004)	(0.004)	(0.003)	(0.004)	(0.004)	(0.004)
Income	-0.019***	-0.007**	-0.019***	-0.013***	-0.015***	-0.020***	-0.022***	-0.012***
	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.003)
Urban	0.012**	0.016***	0.015***	0.007	0.000	0.021***	0.017***	0.008
	(0.006)	(0.006)	(0.005)	(0.006)	(0.004)	(0.006)	(0.005)	(0.006)
Total population	-0.001	0.002	-0.004	0.003	0.017***	-0.012	-0.005	0.001
	(0.008)	(0.007)	(0.008)	(0.008)	(0.006)	(0.009)	(0.008)	(0.008)
Over65	0.003	0.004	0.002	0.007	0.003	0.003	0.002	0.006
	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)
Male-female ratio	0.001	0.001	-0.000	0.002	0.003	-0.001	-0.001	0.003
	(0.004)	(0.006)	(0.004)	(0.005)	(0.004)	(0.005)	(0.004)	(0.004)
Minority	0.020***	0.013**	0.023***	0.007	0.013***	0.022***	0.018***	0.021***
	(0.005)	(0.007)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Interest rate	0.045***	0.046***	0.057***	0.025***	0.027***	0.052***	0.048***	0.038***
	(0.002)	(0.005)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.002)
FICO	-0.108***	-0.106***	-0.107***	-0.104***	-0.119***	-0.095***	-0.093***	-0.113***
	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.004)
Balance	0.040***	0.044***	0.046***	0.032***	0.027***	0.049***	0.037***	0.058***
	(0.002)	(0.003)	(0.002)	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)
LTV	0.007***	0.037***	0.008***	0.019***	0.001	0.015***	0.020***	0.004***
	(0.001)	(0.003)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
ARM	0.069***	0.067***	0.061***	0.081***	0.049***	0.080***	0.082***	0.027***
	(0.004)	(0.006)	(0.004)	(0.004)	(0.005)	(0.003)	(0.003)	(0.005)
Option ARM	0.004	0.101*	0.005	0.013	0.022	0.022	0.031	0.004
	(0.016)	(0.061)	(0.016)	(0.029)	(0.136)	(0.015)	(0.021)	(0.021)
Negative amortization	0.029***	0.076***	0.028***	0.017**	-0.027***	0.041***	0.043***	0.004
	(0.006)	(0.015)	(0.008)	(0.007)	(0.010)	(0.008)	(0.009)	(0.007)
Low or no doc.	0.071***	0.114***	0.065***	0.093***			0.102***	0.054***
	(0.004)	(0.006)	(0.004)	(0.004)			(0.004)	(0.004)
Prepayment penalty	0.067***	0.062***	0.073***	0.046***	0.050***	0.075***	0.079***	0.038***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.004)	(0.004)	(0.005)
Cash-out			0.042***	0.079***	0.062***	0.050***	0.045***	0.067***
			(0.008)	(0.009)	(0.009)	(0.008)	(0.007)	(0.013)
No-cash-out			0.010	0.035***	-0.005	0.016*	0.018**	0.003
			(0.008)	(0.012)	(0.011)	(0.008)	(0.008)	(0.015)
Investment	-0.036***	0.009			-0.052***	-0.018***	-0.010**	-0.062***
	(0.003)	(0.008)			(0.004)	(0.004)	(0.004)	(0.003)
Second-home	-0.013***	0.042***			-0.033***	0.005	0.021***	-0.055***
	(0.004)	(0.009)			(0.004)	(0.004)	(0.004)	(0.005)
State FE	Y	Y	Y	Y	Y	Y	Y	Y
Half-year FE	Y	Y	Y	Y	Y	Y	Y	Y
Originator FE	Y	Y	Y	Y	Y	Y	Y	Y
Observations	439,694	39,454	370,220	108,821	171,780	307,248	276,009	203,152
Adj. R^2	0.249	0.220	0.258	0.207	0.229	0.256	0.270	0.177

The table shows OLS estimates from regressions in which the dependent variable takes a value of one if the loan becomes 90 days or more delinquent using MBA method in the first three years after origination and zero otherwise. All continuous variables are winsorized at 1 percent level and then standardized. All specifications include state, half-year, and originator fixed effects, and standard errors clustered at the county level are reported in the parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

Table 9
Effect of Gambling Preference Associated Owner-occupancy Misreporting on Delinquency

	(1)	(2)	(3)	(4)
Owner-occupancy misreporting	0.074***		0.074***	0.073***
	(0.004)		(0.004)	(0.004)
CPRATIO		-0.002	-0.003	-0.005
		(0.005)	(0.004)	(0.005)
Owner-occupancy misreporting × CPRATIO				0.006**
				(0.003)
REL	-0.013***	-0.013***	-0.012***	-0.013***
	(0.004)	(0.004)	(0.003)	(0.003)
HPA	0.037***	0.037***	0.037***	0.037***
	(0.004)	(0.004)	(0.004)	(0.004)
Education	0.001	0.002	0.001	0.001
	(0.003)	(0.003)	(0.003)	(0.003)
Married	0.018***	0.020***	0.019***	0.019***
	(0.003)	(0.003)	(0.003)	(0.003)
Income	-0.015***	-0.020***	-0.015***	-0.015***
	(0.002)	(0.002)	(0.002)	(0.002)
Urban	0.014***	0.013***	0.014***	0.014***
	(0.004)	(0.004)	(0.004)	(0.004)
Total population	0.001	0.003	0.002	0.002
	(0.007)	(0.006)	(0.006)	(0.006)
Over65	-0.000	-0.001	0.000	0.000
	(0.003)	(0.003)	(0.003)	(0.003)
Male-female ratio	-0.001	-0.001	-0.001	-0.001
	(0.003)	(0.003)	(0.003)	(0.003)
Minority	0.016***	0.018***	0.016***	0.016***
	(0.004)	(0.004)	(0.004)	(0.004)
Interest rate	0.053***	0.050***	0.053***	0.053***
	(0.002)	(0.002)	(0.002)	(0.002)
FICO	-0.075***	-0.082***	-0.075***	-0.075***
	(0.001)	(0.001)	(0.001)	(0.001)
Balance	0.014***	0.024***	0.014***	0.014***
	(0.002)	(0.002)	(0.002)	(0.002)
LTV	0.052***	0.054***	0.052***	0.052***
	(0.001)	(0.001)	(0.001)	(0.001)
ARM	0.063***	0.065***	0.063***	0.063***
	(0.002)	(0.002)	(0.002)	(0.002)
Option ARM	0.046***	0.045**	0.046***	0.046***
	(0.017)	(0.017)	(0.017)	(0.017)
Negative amortization	0.056***	0.038***	0.056***	0.056***
	(0.005)	(0.005)	(0.005)	(0.005)
Low or no doc.	0.069***	0.072***	0.069***	0.068***
	(0.003)	(0.003)	(0.003)	(0.003)
Prepayment penalty	0.063***	0.064***	0.063***	0.063***
	(0.003)	(0.003)	(0.003)	(0.003)
Cash-out	-0.026***	-0.039***	-0.026***	-0.026***
	(0.003)	(0.004)	(0.003)	(0.003)
No-cash-out	0.019***	0.010***	0.019***	0.019***
	(0.003)	(0.003)	(0.003)	(0.003)
State FE	Y	Y	Y	Y
Half-year FE	Y	Y	Y	Y
Originator FE	Y	Y	Y	Y
Observations	466,660	466,660	466,660	466,660
Adj. R^2	0.247	0.242	0.247	0.247

The table shows OLS estimates from regressions in which the dependent variable takes a value of one if the loan becomes 90 days or more delinquent using MBA method in the first three years after origination and zero otherwise. All continuous variables are winsorized at 1 percent level and then standardized. All specifications include state, half-year, and originator fixed effects, and standard errors clustered at the county level are reported in the parentheses. $*p < 0.10$, $**p < 0.05$, and $***p < 0.01$.

Table 10
Effect of Gambling Preference Associated Owner-occupancy Misreporting on Delinquency - Subsamples

	(1) Purchase	(2) Refinance	(3) Full Doc	(4) Low/No Doc	(5) High FICO	(6) Low FICO
Owner-occupancy misreporting	0.093***	0.012***	0.064***	0.077***	0.079***	0.060***
	(0.004)	(0.004)	(0.006)	(0.003)	(0.004)	(0.005)
CPRATIO	-0.005	-0.005	0.000	-0.007	-0.003	-0.014**
	(0.005)	(0.004)	(0.006)	(0.005)	(0.004)	(0.006)
Owner-occupancy misreporting × CPRATIO	0.007**	0.004	0.006	0.005**	0.004	0.009**
	(0.003)	(0.003)	(0.005)	(0.003)	(0.003)	(0.004)
REL	-0.014***	-0.009***	-0.014***	-0.011***	-0.012***	-0.010**
	(0.004)	(0.003)	(0.004)	(0.004)	(0.003)	(0.004)
HPA	0.037***	0.035***	0.025***	0.039***	0.037***	0.041***
	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)
Education	0.003	-0.001	0.003	0.000	0.000	0.004
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Married	0.023***	0.011***	0.010***	0.021***	0.019***	0.017***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)
Income	-0.020***	-0.004**	-0.013***	-0.015***	-0.015***	-0.012***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Urban	0.016***	0.011***	0.002	0.018***	0.013***	0.013**
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)
Total population	-0.001	0.003	0.017***	-0.003	-0.001	0.007
	(0.006)	(0.006)	(0.006)	(0.007)	(0.006)	(0.008)
Over65	-0.003	0.007**	0.001	-0.000	-0.000	0.001
	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)
Male-female ratio	-0.002	0.002	-0.001	-0.001	-0.001	-0.000
	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)	(0.004)
Minority	0.023***	0.006	0.010**	0.018***	0.016***	0.016***
	(0.005)	(0.004)	(0.005)	(0.004)	(0.004)	(0.005)
Interest rate	0.063***	0.035***	0.038***	0.056***	0.058***	0.045***
	(0.003)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)
FICO	-0.077***	-0.076***	-0.075***	-0.075***	-0.086***	-0.045***
	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.004)
Balance	0.018***	0.010***	0.008***	0.016***	0.011***	0.026***
	(0.003)	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)
LTV	0.038***	0.069***	0.044***	0.055***	0.045***	0.063***
	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)
ARM	0.067***	0.044***	0.068***	0.066***	0.071***	0.043***
	(0.003)	(0.003)	(0.004)	(0.002)	(0.003)	(0.004)
Option ARM	0.052***	0.078***	0.016*	0.063***	0.062***	0.037***
	(0.007)	(0.006)	(0.009)	(0.005)	(0.006)	(0.007)
Negative amortization	0.037*	0.058***	-0.021	0.077***	0.092***	0.000
	(0.020)	(0.022)	(0.083)	(0.016)	(0.018)	(0.024)
Low or no doc.	0.081***	0.025***	0.047***	0.068***	0.069***	0.047***
	(0.003)	(0.003)	(0.006)	(0.003)	(0.002)	(0.005)
Prepayment penalty	0.068***	0.068***			0.078***	0.042***
	(0.003)	(0.003)			(0.003)	(0.005)
Cash-out			-0.035***	-0.025***	-0.018***	-0.047***
			(0.004)	(0.004)	(0.004)	(0.004)
No-cash-out			-0.003	0.024***	0.027***	-0.019***
			(0.005)	(0.004)	(0.003)	(0.006)
State FE	Y	Y	Y	Y	Y	Y
Half-year FE	Y	Y	Y	Y	Y	Y
Originator FE	Y	Y	Y	Y	Y	Y
Observations	288,427	177,883	118,380	347,998	338,713	127,748
Adj. R^2	0.268	0.224	0.266	0.246	0.239	0.188

The table shows OLS estimates from regressions in which the dependent variable takes a value of one if the loan becomes 90 days or more delinquent using MBA method in the first three years after origination and zero otherwise. All continuous variables are winsorized at 1 percent level and then standardized. All specifications include state, half-year, and originator fixed effects, and standard errors clustered at the county level are reported in the parentheses. $^*p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

Table 11
Alternative Default Measures for Loan Performance Tests

Panel A. Second-lien Misrepresentation

	(1) 90+ delinq in 3 year	(2) 90+ delinq in 2 year	(3) 60+ delinq in 3 year	(4) B/F/REO in 3 year
Second-lien misrepresentation	-0.018***	-0.013***	-0.019***	-0.008***
	(0.002)	(0.002)	(0.002)	(0.002)
CPRATIO	-0.002	-0.006	-0.001	-0.005
	(0.005)	(0.004)	(0.005)	(0.006)
Second-lien misrepresentation × CPRATIO	-0.011***	-0.008***	-0.011***	-0.006***
	(0.003)	(0.002)	(0.003)	(0.002)
Other controls	Y	Y	Y	Y
State FE	Y	Y	Y	Y
Half-year FE	Y	Y	Y	Y
Originator FE	Y	Y	Y	Y
Observations	479,341	479,341	479,341	479,341
Adj. R^2	0.244	0.167	0.248	0.179

Panel B. Owner-occupancy Misreporting

	(1) 90+ delinq in 3 year	(2) 90+ delinq in 2 year	(3) 60+ delinq in 3 year	(4) B/F/REO in 3 year
Owner-occupancy misreporting	0.073***	0.063***	0.075***	0.048***
	(0.004)	(0.003)	(0.004)	(0.003)
CPRATIO	-0.005	-0.006**	-0.004	-0.004
	(0.005)	(0.003)	(0.005)	(0.004)
Owner-occupancy misreporting × CPRATIO	0.006**	0.006**	0.005*	0.004
	(0.003)	(0.002)	(0.003)	(0.003)
Other controls	Y	Y	Y	Y
State FE	Y	Y	Y	Y
Half-year FE	Y	Y	Y	Y
Originator FE	Y	Y	Y	Y
Observations	466,660	466,660	466,660	466,660
Adj. R^2	0.247	0.195	0.255	0.204

The table shows OLS estimates from regressions in which the dependent variable uses different measures of default. Column (1) uses 90 days or more delinquency using MBA method in the first three years after origination (baseline). Column (2) uses 90 days or more delinquency using MBA method in the first two years after origination. Column (3) uses 60 days or more delinquency using MBA method in the first two years after origination. Column (4) uses bankruptcy/foreclosure/REO in the first three years after origination. Panel A is for second-lien misrepresentation and panel B is for owner-occupancy misreporting. Other control variables include all control variables used in Table 7 or Table 9. All continuous variables are winsorized at 1 percent level and then standardized. All specifications include state, half-year, and originator fixed effects, and standard errors clustered at the county level are reported in the parentheses. $*p < 0.10$, $**p < 0.05$, and $***p < 0.01$.

Table 12
Gambling Preference and Mortgage Misrepresentation - Probit Model

	(1) Second-lien misrepresentation	(2) Owner-occupancy misreporting
CPRATIO	0.016**	0.035**
	(0.006)	(0.018)
REL	-0.003	-0.033*
	(0.005)	(0.012)
HPA	0.012	0.004
	(0.008)	(0.013)
Education	0.005	0.034*
	(0.004)	(0.010)
Married	0.003	0.036**
	(0.005)	(0.016)
Income	-0.007**	-0.199*
	(0.003)	(0.010)
Urban	-0.005	-0.058*
	(0.005)	(0.011)
Total population	0.016**	0.065*
	(0.006)	(0.019)
Over65	0.003	-0.038*
	(0.004)	(0.014)
Male-female ratio	-0.001	-0.002
	(0.004)	(0.015)
Minority	0.014**	0.060*
	(0.006)	(0.020)
Interest rate	-0.020*	-0.137*
	(0.004)	(0.008)
FICO	-0.109*	-0.295*
	(0.005)	(0.008)
Balance	0.056*	0.438*
	(0.005)	(0.017)
LTV	0.259*	0.097*
	(0.009)	(0.005)
ARM	0.145*	0.076*
	(0.009)	(0.009)
Option ARM	-0.540*	-0.016
	(0.070)	(0.047)
Negative amortization	0.306*	-0.783*
	(0.017)	(0.021)
Low or no doc.	-0.107*	0.134*
	(0.009)	(0.013)
Prepayment penalty	-0.148*	0.032**
	(0.012)	(0.013)
Cash-out	0.508*	-0.578*
	(0.014)	(0.018)
No-cash-out	0.270*	-0.394*
	(0.011)	(0.016)
Investment	0.614*	
	(0.016)	
Second-home	0.267*	
	(0.019)	
State FE	Y	Y
Half-year FE	Y	Y
Originator FE	Y	Y
Observations	443,241	452,695
Pseudo R^2	0.236	0.170

The table shows nonlinear model (probit) estimates from regressions in which the dependent variable takes a value of one if the loan is misrepresented and zero otherwise. Column (1) uses second-lien misrepresentation measure for dependent variable and column (2) uses owner-occupancy misreporting measure for dependent variable. All continuous variables are winsorized at 1 percent level and then standardized. All specifications include state, half-year, and originator fixed effects, and standard errors clustered at the county level are reported in the parentheses. $^*p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

Table 13
Gambling Preference and Mortgage Misrepresentation - Causal Forest

	(1) Second-lien misrepresentation	(2) Owner-occupancy misreporting
CPRATIO	0.006**	0.011***
	(0.003)	(0.003)

The table shows causal forest estimates from regressions in which the dependent variable takes a value of one if the loan is misrepresented and zero otherwise. Column (1) uses second-lien misrepresentation measure for dependent variable and column (2) uses owner-occupancy misreporting measure for dependent variable. All control variables in Table 12 are used for growing trees and forests. Before growing the trees, all continuous variables are winsorized at 1 percent level and then standardized. Half-year variable are created, starting from 2005 first half as 1, and used as a variable in growing trees and forests. Observations are clustered by state, and each unit is given the same weight (so that bigger clusters get more weight). 2000 trees are grown in causal forest. The fraction used for determining splits (honesty) is 50 percent. Standard errors are reported in the parentheses. *$p < 0.10$, **$p < 0.05$, and ***$p < 0.01$.

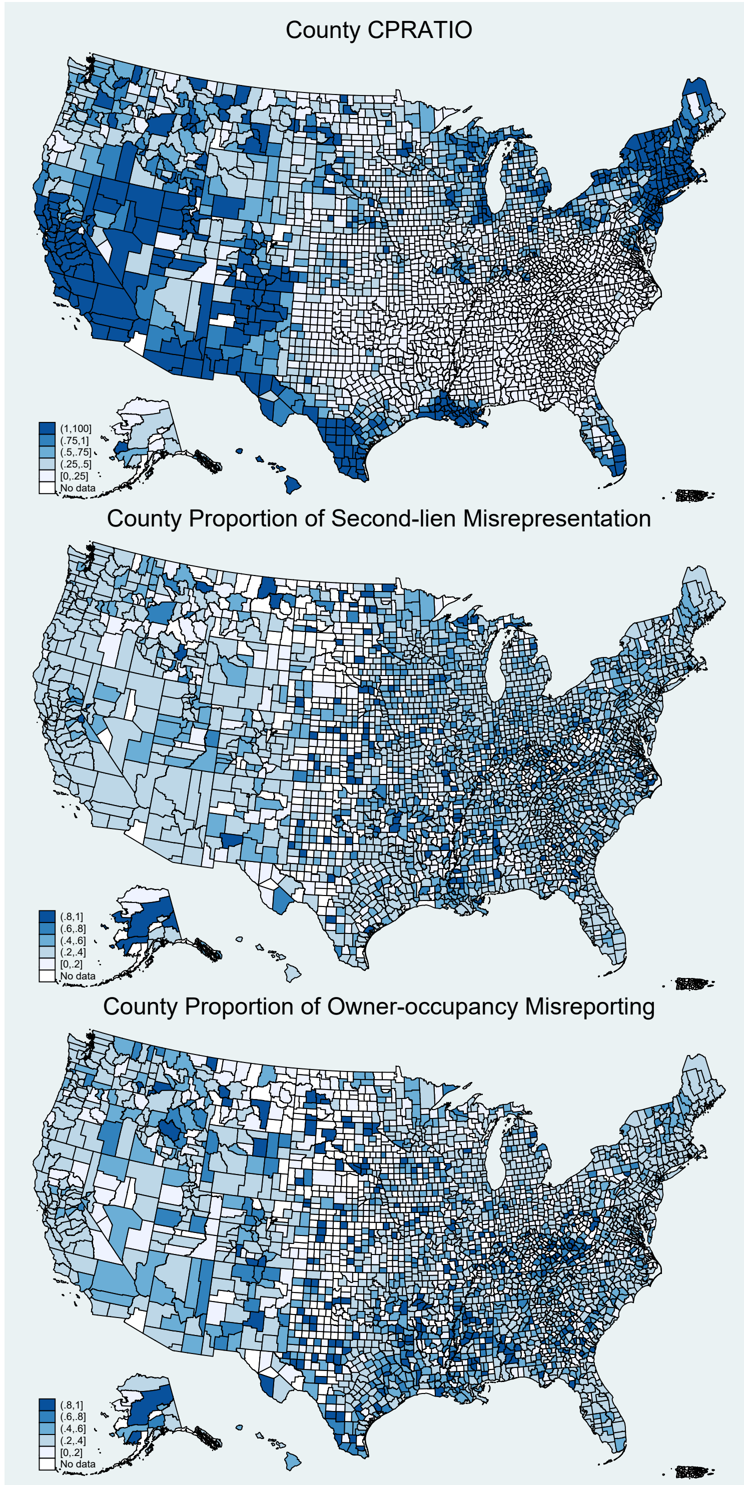

Figure 1
CPRATIO and proportion of mortgage misrepresentations
The top panel plots the county level CPRATIO across the US. The middle panel plots the county level proportion of second-lien misrepresentation in the sample. The bottom panel plots the county level proportion of owner-occupancy misreporting in the sample.

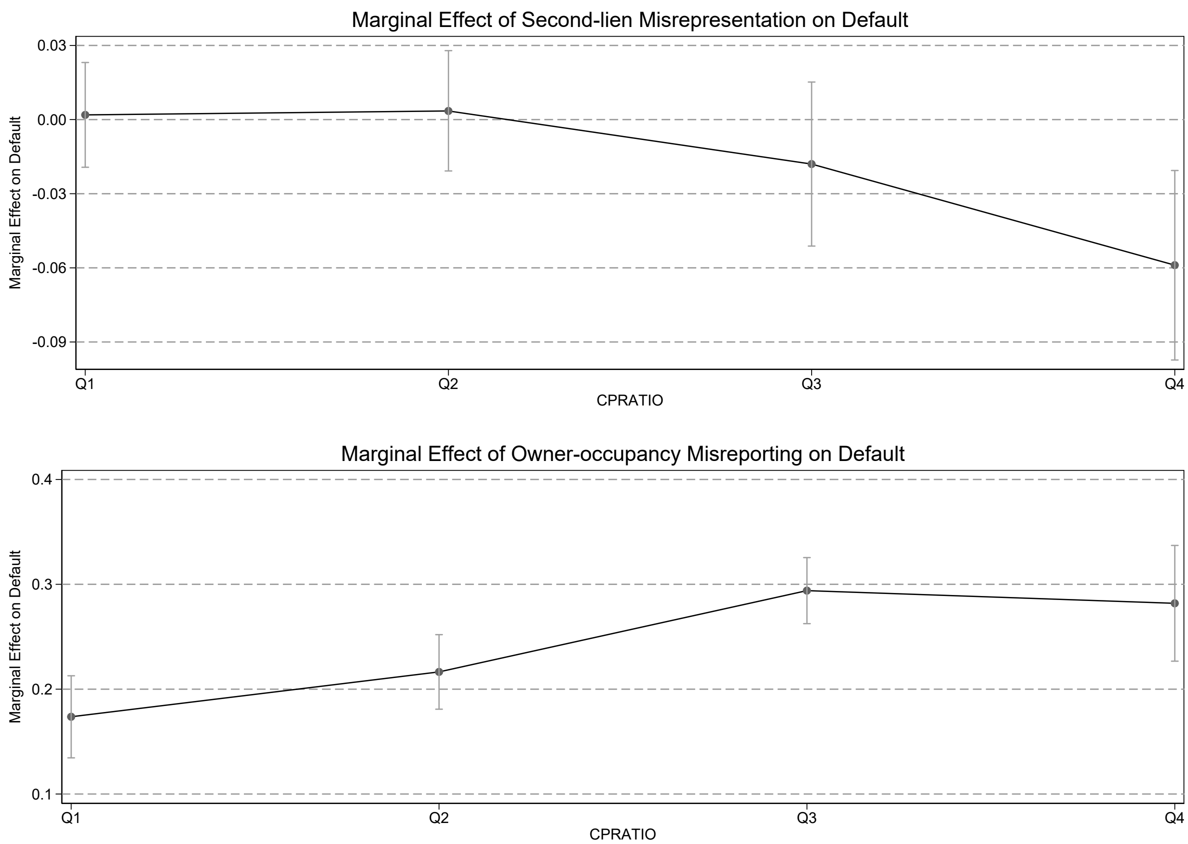

Figure 2
Marginal Effects of Mortgage Misrepresentation on Default at Different Levels of Gambling Preference

The figure plots the marginal effects of mortgage fraud on the probability of default from a probit model. The top panel is for second-lien misrepresentation and the bottom panel is for owner-occupancy misreporting. The dependent variable is an indicator variable that takes a value of one if the loan becomes 90 days or more delinquent using MBA method in the first three years after origination. The marginal effect of mortgage fraud is estimated at different levels of local gambling preference (four quarters divided by CPRATIO). The control varaibles are identical to the linear probability model (OLS) estimated in column (3) of Table 7 or Table 9.

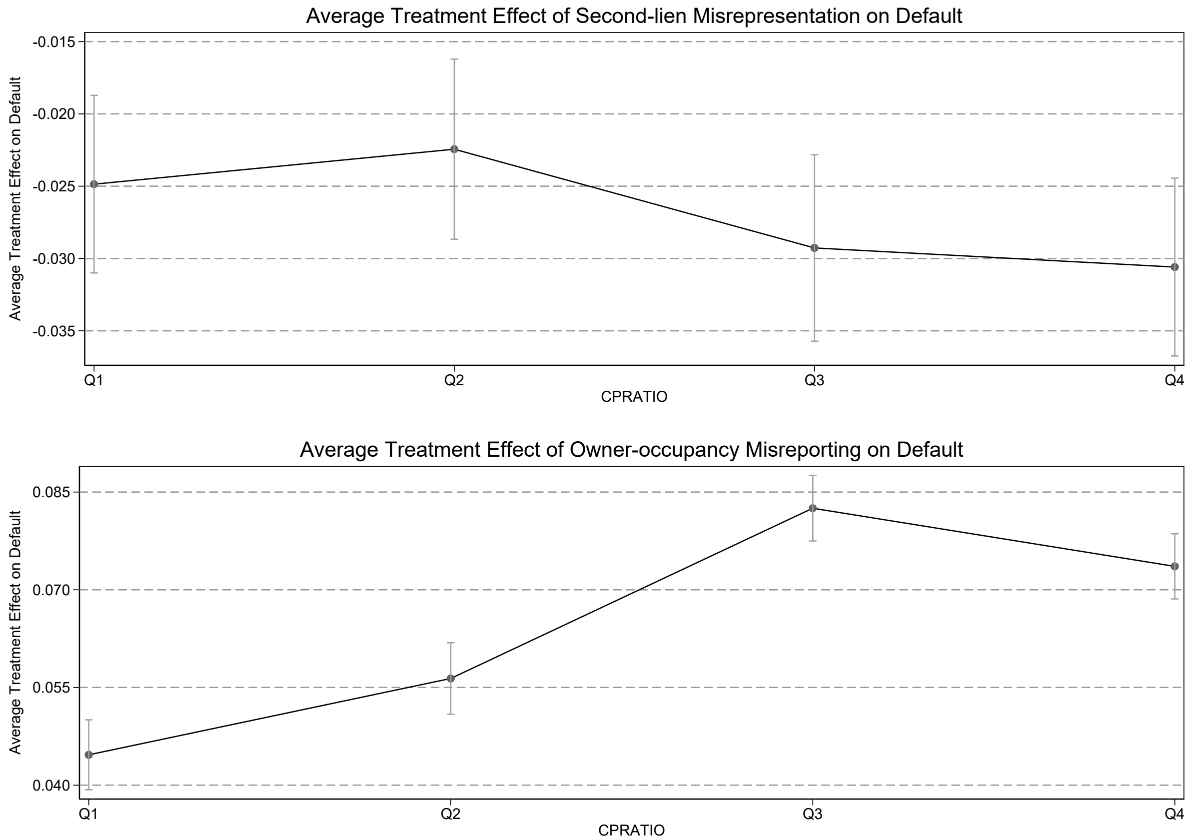

Figure 3
Heterogeneous Treatment Effects of Mortgage Misrepresentation on Default conditioned on Different Levels of Gambling Preference
The figure plots the heterogeneous treatment effects of mortgage fraud on the default conditioned on different levels of gambling preference using causal forest approach. The top panel is for second-lien misrepresentation and the bottom panel is for owner-occupancy misreporting. The dependent variable is an indicator variable that takes a value of one if the loan becomes 90 days or more delinquent using MBA method in the first three years after origination. The average treatment effect of mortgage fraud is estimated at different levels of local gambling preference (four quarters divided by CPRATIO). All control variables in in column (3) of Table 7 or Table 9 are used for growing trees and forests. Before growing the trees, all continuous variables are winsorized at 1 percent level and then standardized. Half-year variable are created, starting from 2005 first half as 1, and used as a variable in growing trees and forests. Observations are clustered by state, and each unit is given the same weight (so that bigger clusters get more weight). 2000 trees are grown in causal forest. The fraction used for determining splits (honesty) is 50 percent.