Local Gambling Preference and Mortgage Misrepresentation

Jiawei Hu *

August 2025

Abstract

This paper examines the role of behavioral traits in borrowers' decisions during mortgage applications. Analyzing a large sample of mortgages originated between 2005 and 2007, we investigate the impact of local gambling preferences on second-lien misrepresentation and its subsequent effect on loan performance using OLS, probit, and causal forest approaches. Our findings indicate that second-lien misrepresentation is more prevalent in areas with higher local gambling preferences. Furthermore, loans with second-lien misrepresentation in high gambling preference areas exhibit poorer performance compared to those in low gambling preference areas. Utilizing RDD and difference-in-discontinuities approaches, we compare the number of loans and default rates around a FICO score of 620 between high and low gambling preference areas. Complementing previous literature that studies second-lien misrepresentation from the perspective of intermediation, our results suggest that borrowers also play an important role in the misrepresentation. The influence of gambling preferences on misrepresentation is more likely attributable to borrower behavior rather than lender practices.

Keywords: Mortgage misrepresentation, Local gambling preference, Second-lien misrepresentation

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

Housing, encompassing both consumption and investment properties, is the most important asset for most households (Campbell, 2006; Badarinza et al., 2016; Gomes et al., 2021). Housing decisions significantly affect both macroeconomic stability and households' portfolio choices (Cocco, 2005; Yao and Zhang, 2005; Piazzesi and Schneider, 2016). As the primary means for households to finance real estate purchases, mortgages play a crucial role in housing decisions.

The origins of the 2008 financial crisis, driven by the boom and bust in mortgage lending, remain the subject of debate. Within an environment of optimistic house-price beliefs, two dominant views attribute the crisis to different borrower groups. The credit-expansion view emphasizes the subprime entry, arguing that high-risk households gained mortgage access through loosened lending standards and abundant credit (Mian and Sufi, 2009; Justiniano et al., 2015), adding risk beyond what price expectations alone would suggest. The expectations view models belief distortions and highlights middle-class borrowers, suggesting that typically safe households took on excessive risk due to overly optimistic price expectations (Burnside et al., 2016; Adelino et al., 2018). Much of the literature thus explains risky borrowing through belief distortions, emphasizing income class. This paper operates within that belief-distortion environment and adds a complementary perspective by introducing borrower behavioral traits into the analysis of household financial decision-making. We shift the focus from borrower class to borrower traits, arguing that behavioral predispositions, specifically gambling preference, shape the willingness to turn optimistic beliefs and credit access into extreme risk-taking across income groups. In our framework, credit expansion provides the means, expectations provide the rationale, and behavioral traits determine the threshold for undertaking high-stakes actions.

Our study investigates this trait-based channel from the borrower's perspective. In particular, we focus on second-lien misrepresentation, a high-stakes behavior during mortgage application where the existence of a simultaneous second lien on a property is not disclosed when applying for the first lien. While previous literature discuss this issue within the scope of intermediation (Piskorski et al., 2015; Griffin and Maturana, 2016; Yavas and Zhu, 2024), we shed light on the misrepresentation from the perspective of borrower behavior. Our findings suggest that, beyond intermediation dynamics, borrower traits play a significant role in driving second-lien misrepresentation.

In the context of second-lien misrepresentation, we study a specific behavioral trait: gambling preference. This trait reflects a behavioral tendency to engage in gambling activities, commonly defined as the intentional wagering of something of value on events with uncertain outcomes in pursuit of potential rewards (Rose and Loeb, 1999). Studies have shown that individual gambling propensity develops through cognitive biases (e.g., illusion of control), is associated with various psychological traits (e.g., impulsivity, risk-seeking, deceitfulness, and low levels of remorse), and is shaped by local social norms that reflect cultural values, religious teachings, and peer influences (Armstrong et al., 2020; Mishra et al., 2010; Blaszczynski and Nower, 2002; Raylu and Oei, 2004; Parrado-González et al., 2023). Therefore, we conjecture that gambling preference is associated with second-lien misrepresentation for two reasons.

First, second-lien misrepresentation resembles a gamble: borrowers wager their future financial stability by concealing simultaneous second liens to improve approval chances or secure better terms. The primary prize is first-lien approval and homeownership, offering substantial emotional benefits and significant financial returns[1]. Other rewards may include more favorable loan terms, such as lower interest rates. Conversely, the downsides include loan denial, correction requests, or legal consequences[2]. For those with stronger gambling preferences, cognitive biases like the illusion of control inflate perceived odds of success, while traits such as impulsivity and low remorse make concealment easier to justify and

[1]Emotionally, homeownership is associated with the American Dream in the U.S. (Clinton, 1995; Bush, 2003). Financially, with appropriate mortgage terms, homeownership brings large financial returns and helps households accumulate wealth (Bostic and Lee, 2008; Goodman and Mayer, 2018)

[2]Denial and corrections are common but switching lenders is relatively low cost; legal penalties (Henning, 2009; Federal Housing Finance Agency, 2023) were rare pre-crisis when screening was loose and enforcement focused on fraud for profit rather than fraud for housing (FBI, 2006, 2007).

more appealing.

Second, local gambling preference may influence misrepresentation through social influence. Following Christensen et al. (2018), who link local gambling attitudes to higher corporate misreporting, we posit that similar mechanisms apply at the household level. Similar to the fraud triangle theory (Cressey, 1953), misrepresentation meets the elements of rationalization, incentive, and opportunity, with gambling preference driving rationalization. Local gambling attitudes promote speculative, high-stakes decision-making and increase risk tolerance (Chen et al., 2014), helping borrowers justify the gamble. Incentives can take many forms, such as pursuing homeownership in socially supportive contexts despite financial constraints. Opportunity exists because borrowers self-report second liens, which can be concealed if lenders fail to detect them or choose to ignore them in the pre-crisis environment[3]. Therefore, given both the incentive and opportunity to conceal leverage, we argue that borrowers in areas with high gambling preference are more likely to engage in second-lien misrepresentation. However, as Christensen et al. (2018) note, the spillover from gambling attitudes to misreporting remains empirical, since gambling is generally legal while concealing leverage may violate the law.

We measure second-lien misrepresentation and gambling preference following previous literature. We construct the second-lien misrepresentation measure following Piskorski et al. (2015), which compares loan-level mortgage data with borrowers' credit report information. We follow Kumar (2009) and Kumar et al. (2011) to measure local gambling preference, using the ratio of Catholic residents to Protestant residents (CPRATIO) at the county level. We choose this measure for two reasons. First, since the potential distribution of losses and gains for individual loans is not observable, we cannot directly apply existing theoretical frameworks in the housing market[4]. Second, geographic measures are appropriate as people in the same area share social norms, while different areas often exhibit distinct cultural

[3]Second liens may be missed if not closed during title searches, and even when detected, lenders often approved loans in the pre-crisis period.

[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.

attitudes toward gambling. We interpret CPRATIO strictly as a regional cultural indicator of gambling preference that reflects social norms around speculative risk-taking in financial decisions, without inferring individual morality, personality traits, or religious affiliation[5].

With these measures in place, we first investigate whether gambling preference helps explain the occurrence of mortgage misrepresentation. Comparing areas with different levels of local gambling preference, we find that second-lien misrepresentation is more likely to occur in counties with higher levels of local gambling preference, indicating it is an important driver. We then conduct subsample analyses. First, we assess whether borrower traits matter across income classes and find consistent effects for both lower- and middle-income groups. Second, to test our hypothesis that misrepresentation, when it resembles a gamble, is more attractive to borrowers with high gambling preferences, we split the sample by characteristics that reflect different potential benefits. The effect is stronger when the property is owner-occupied, the loan is for purchase, or the borrower has a low credit score, consistent with these cases offering greater potential gains to borrowers.

Second, we examine the economic impact of mortgage misrepresentation associated with gambling preference. Given the strong correlation between gambling preference and misrepresentation, and the established link between misrepresentation and poor loan performance (Piskorski et al., 2015; Griffin and Maturana, 2016), we test whether gambling preference is related to worse outcomes through this channel. We find that loans with second-lien misrepresentation perform worse on average, with the effect more pronounced in high gambling preference areas. Subsample analyses show that this pattern holds in nearly all groups, indicating that the behavioral bias linked to gambling preference is consistently associated with poorer loan performance.

Third, we study the extent to which lenders facilitate second-lien misrepresentation and examine whether the positive correlation between gambling preference and second-lien misrepresentation reflects borrower or lender issue. Using the Keys et al. (2010) setting, where

[5]We also employ alternative measures that do not rely on religiously constructed information in robustness check to mitigate the concern on broader religious difference.

ease of securitization for low-documentation loans with FICO scores equal to or greater than 620 provides a shock to lenders but not borrowers, we apply a regression discontinuity design (RDD) to calculate jumps in number of loans and default rates, and compare these across high and low gambling preference areas using difference-in-discontinuities (diff-in-disc). Across all areas, the jump ratio in loans with misrepresented simultaneous seconds exceeds that of loans without seconds but is smaller than that of loans with correctly reported seconds, indicating lenders facilitate loans with simultaneous seconds but not specifically misrepresentation. The jump in default rates for loans with simultaneous seconds is small and insignificant, suggesting screening effort is not a main driver. Comparing high and low gambling preference areas, jumps in both number of loans and default rates for misrepresented seconds are smaller in high gambling preference areas, implying lenders there do not play a larger role in facilitating misrepresentation. Therefore, the positive correlation we found is more likely a borrower issue than a lender issue.

Finally, we conduct several robustness checks to reinforce our findings. First, to address concerns that CPRATIO may capture broader religious traits rather than gambling preference, we use alternative measures based on observed gambling activity, and the results remain consistent. Second, to test whether CPRATIO reflects tolerance for law-breaking, we examine its association with crime rates and find no positive correlation with crimes other than illegal gambling. Third, to check that results are not driven by a specific default definition, we use alternative measures that are more or less restrictive, obtaining similar outcomes. Fourth, to verify that findings are not model-specific, we re-estimate using probit models and again find consistent results. Fifth, to strengthen causal inference, we apply the causal forest approach of Wager and Athey (2018), which yields similar conclusions. Last, to confirm the robustness of our diff-in-disc results, we conduct multiple specification tests and find the patterns in high and low gambling-preference areas persist.

Our paper contributes to four strands of literature. First, our paper introduces a trait-based behavioral channel into the debate on the 2008 financial crisis. Existing views emphasize either subprime entry via abundant credit (Mian and Sufi, 2009; Justiniano et al., 2015) or excessive risk-taking by middle-class borrowers with optimistic price beliefs (Burnside et al., 2016; Adelino et al., 2018), both focusing on borrower class within a belief-distortion framework. Our paper shifts the focus to borrower traits, showing that behavioral predispositions, particularly gambling preference, shape willingness to convert beliefs and credit access into high-stakes mortgage choices across income groups.

Second, our paper complements the literature by discussing where in the credit supply chain (from borrowers to MBS underwriters) second-lien misrepresentations took place. By studying privately securitized loans, Griffin and Maturana (2016) point out that lenders were likely aware of the misrepresentation, attributing such misreporting more to lenders than to underwriters. Conversely, Piskorski et al. (2015) show that underwriters could easily uncover the misrepresentation and also play an important role in this issue. Using both portfolio loans and securitized loans, Yavas and Zhu (2024) support the argument that second-lien misrepresentation occurs in the early stages of intermediation by lenders, while underwriters have limited but significant effects on reducing the occurrence of misrepresentation through screening. Our paper examines this issue from the borrower's perspective, showing that borrowers significantly contribute to second-lien misrepresentation.

Third, our paper contributes to the literature on mortgage fraud by proposing a preference-based explanation. A substantial body of literature documents the prevalence of mortgage fraud and its economic impact (Elul et al., 2010; Piskorski et al., 2015; Ambrose et al., 2016; Griffin and Maturana, 2016; Mian and Suf, 2017; Kruger and Maturana, 2021). While previous studies have considered associated loan characteristics and household features, the role of preference in the decision-making process remains unclear. From a cultural perspective, Conklin et al. (2022) provide empirical evidence that religiosity helps constrain fraudulent activity but do not distinguish between ethics and risk channels. Bypassing the issue of differentiating ethics and risk preference, we show that gambling preference[6] influences mortgage

[6] See Kumar (2009), Kumar et al. (2011), Kumar et al. (2016) for how to differentiate gambling preference from ethics.

applications.

Fourth, our paper builds on the empirical literature linking gambling preferences with investment decisions. Prior research has examined the effect of gambling preference on the 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). Additionally, literature explores its impact on corporate policy decisions (Kumar et al., 2011; Chen et al., 2014) and fund strategies (Shu et al., 2012). To the best of our knowledge, we are the first to apply the concept of gambling preference to the mortgage market in the context of household investment decision-making. This is important because existing literature on behavioral biases in the housing market predominantly centers on pricing dynamics and speculative bubbles from the investment perspective (Lux, 1995; Genesove and Mayer, 2001; Farlow, 2004; Shiller, 2005; Farlow, 2013), whereas relatively little attention has been paid to behavioral influences on mortgage choices, which is the financing side of housing decisions.

The rest of the paper is organized as follows. Section 2 describes the data and variables. In Section 3, we examine the effect of gambling preference on second-lien misrepresentation. Section 4 investigates the economic impact of mortgage misrepresentation associated with gambling preference. In Section 5, we assess whether the effect of gambling preference is a borrower issue or a lender issue. Robustness checks are conducted in Section 6, and conclusions are presented in Section 7.

2 Data and Summary Statistics

Our sample comprises three main groups of data: loan-related records, religiosity data, and demographic information, covering the period from 2005 to 2007. These data are sourced from various datasets. Loan-related records include loan-level mortgage data from BlackBox Logic (now part of Moody's) and borrower-level credit report information from Equifax. Religiosity data, which provides county-level information on prevalent religious adherence, is obtained from the American Religion Data Archive (ARDA). Demographic information at the county and zipcode levels, such as age and income, is sourced from the U.S. Census Bureau. Additionally, we collect house price data from the Federal Housing Finance Agency (FHFA) and unemployment data from U.S. Bureau of Labor Statistics.

2.1 Second-lien Misrepresentation Measure

Second-lien misrepresentation occurs when a first-lien loan backed by property is reported as having no associated higher lien but is actually financed with a simultaneously originated second mortgage identified by credit bureau data. This misrepresentation allows the borrower to take on additional debt, reducing their incentive to repay the loans and making the initial debt riskier. We identify second-lien misrepresentation by comparing loan-level mortgage data from BlackBox Logic with borrower-level credit report information from Equifax. Following the procedure in Piskorski et al. (2015), we first focus on first-lien loans with a merge confidence interval greater than or equal to 0.89. Second, we select loans with a new second-lien originating within one month before or after the first lien[7]. This filter, using credit information, retains first-lien loans that truly have a simultaneous second lien. Third, to identify misrepresented second-lien loans, we require the loan to have a non-missing reported cumulative loan-to-value (CLTV) ratio within 1% of its loan-to-value (LTV) ratio. The small difference between the reported CLTV and LTV indicates that the borrower reports no simultaneous second lien.

2.2 Gambling Preference and Religiosity

Following Kumar et al. (2011), we construct county-level gambling preference based on religious data, specifically the ratio of Catholic residents to Protestant residents (CPRATIO).

[7] Piskorski et al. (2015) uses a 45-day range, while Zhang et al. (2024) uses a one-month range. Since Equifax data identifies the close date of the second lien at the month level, we conservatively use a one-month range. A 45-day range can be achieved by assuming the close date is in the middle of the month. We also checked the 45-day range and found minor differences in the results.

Although direct measures of local gambling preference are unavailable, we can infer the propensity by examining the proportion of different religious populations, which have distinct attitudes towards gambling according to their religious views. In the U.S., two widespread religions with differing views on gambling are Catholicism and Protestantism. While Protestant churches generally oppose gambling, Catholic churches maintain a tolerant attitude towards moderate levels of gambling. These differing views between the two religions are empirically supported to extend to financial markets (Kumar, 2009; Kumar et al., 2011; Han and Kumar, 2013; Chen et al., 2014). Therefore, regions with higher Catholic–Protestant ratios exhibit stronger gambling propensities.

To use CPRATIO for potential causal inference, we need to control for religiosity in the county. According to Kumar et al. (2011) and Chen et al. (2014), religiosity should be considered because risk aversion increases with religiosity, regardless of the type of religion (Hilary and Hui, 2009). Including the overall level of religiosity in the county as a control ensures that our local gambling preference proxy is independent of religion-induced risk aversion. Additionally, religiosity plays a significant role in deterring various 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 the dataset "Longitudinal Religious Congregations and Membership File, 1980-2010 (County Level)" from ARDA to capture county-level geographical variation in religious composition. We sum the number of adherents in the religious traditional categories Evangelical Protestant, Mainline Protestant, and Black Protestant for the Protestant population and the number of adherents in the category Catholic for the Catholic population. We use the sum of adherents in all categories 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 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) for each year.

2.3 Geographic Controls

In addition to religiosity, variations in religion-induced gambling preferences may also correlate with other geographic characteristics (Kumar et al., 2011; Chen et al., 2014; Conklin et al., 2022). To help establish causality, we control for the following factors (Kumar, 2009; Kumar et al., 2011; Chen et al., 2014). The U.S. Census Bureau provides a rich set of demographic data. We employ county-level information in 2000 on education, marriage, living area, population, age, male-female ratio, and minority proportion. Since income plays an important role not only in potential correlation with religious distribution but also in household mortgage decisions, we use an even smaller cluster level, the zip code level, for control variables. We utilize county-level monthly data of unemployment rate from U.S. Bureau of Labor Statistics. Additionally, we collect annual house price indices (HPI) at the zip code level to calculate house price appreciation, which is also a crucial determinant in household mortgage decisions.

2.4 Mortgage Microdata

In addition to the second-lien misrepresentation measure, we also obtain other loan-level mortgage data from BlackBox Logic. The database includes a comprehensive set of loan characteristics at origination, such as the loan interest rate, borrower's FICO credit score, initial loan balance, loan-to-value ratio, 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 days, etc.). The definitions of variables constructed from this information are presented in Table 1.

2.5 Descriptive Statistics

We study the effect of gambling preference on second-lien misrepresentation in a sample that includes loans without simultaneous second liens, loans with correctly reported simultaneous second liens, and loans with misrepresented simultaneous second liens. We select the sample period of 2005 to 2007, during which mortgage misrepresentation measures can be calculated using the databases and misrepresentations are not rare. Table 2 shows the summary statistics. For each continuous variable, we report the number of observations, mean, standard deviation, 25th percentile, 50th percentile, and 75th percentile. For dummy variables, we report the number of observations and mean. The continuous variables are winsorized at the 0.5 percent level to ensure our results are not driven by extreme values.

In our sample, we have 3,031,921 loans for which the borrower's area CPRATIO is available[8]. Among these loans, about 7.15 percent have unreported second liens, while 13.91 percent have reported second liens, resulting in a total of 21.06 percent. This percentage is comparable to the one reported by Griffin and Maturana (2016) (10.2%), who used a different database and included all loans (honestly reported, fraudulently hidden, and truly without second liens). Figure 1 plots the nationwide distribution of the county-level proportion of second-lien misrepresentation in all loans, showing variations both across and within states.

Our primary independent variable of interest is local gambling preference (CPRATIO). While there are 3,090 available U.S. counties from 2005 to 2007, we only retain county-year level data where loan-level data is also available. For the 8,716 data points, the mean CPRATIO is 0.51 and the median is 0.19, indicating a large positive skewness[9]. Figure 2 plots the nationwide distribution of CPRATIO. Although high gambling preference areas are mostly concentrated on the west coast, southwest, and east coast, within-state variations exist in many states, such as Texas, Wisconsin, and Indiana.

[8]We also exclude cases where the LTV is greater than 99% since our measure is calculated using LTV. An LTV greater than 99% is likely erroneous or highly unlikely to have a simultaneous second lien.

[9]The mean and median CPRATIO from 1980 to 2005 are 0.60 and 0.23, respectively, in Kumar et al. (2011), which is quite close to our results if we include all counties.

One concern with using local gambling preference is its potential correlation with other geographic variables. Therefore, we examine the correlation between CPRATIO and other geographic variables. Table 3 shows the correlation between geographic variables at the county-year level. Income and HPA at the county level are obtained from the same data sources as the zip code level data, which are the U.S. Census Bureau and the Federal Housing Finance Agency, respectively. Unemployment is the average of the monthly unemployment rate within each year. None of the correlations in the first column exceed 0.32, indicating that CPRATIO contains information not solely covered by other common geographic characteristics. Therefore, the results from categorizing counties by CPRATIO are less likely driven by any other single geographic characteristic.

3 Gambling Preference and Second-lien Misrepresentation

In this section, we investigate the relationship between gambling preference and second-lien misrepresentation. When making decisions, individuals often exhibit various behavioral biases, with gambling preference being prevalent in financial markets. Misreporting a second lien can be seen as a gamble during our sample period, as borrowers could achieve significant gains by betting on the success of such misrepresentation. On one hand, housing prices increased rapidly in the early 2000s, making the financial return of owning a house very attractive (Bostic and Lee, 2008; Goodman and Mayer, 2018). On the other hand, homeownership became an integral part of the American dream (Clinton, 1995; Bush, 2003), making the purchase of a house more than just an investment. Consequently, for borrowers with high gambling preferences who need simultaneous second liens to finance housing, misreporting a second lien may seem like a favorable choice. Thus, we expect that borrowers in areas with high local gambling preferences are more likely to misrepresent material information in mortgage applications. We formalize this idea in Hypothesis 1:

Hypothesis 1 *Second-lien misrepresentation 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 second-lien misrepresentation 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 half-year fixed effect (Piskorski et al., 2015); λ_o is originator fixed effects[10]; ϵ is an error term. Moreover, since our primary independent variable of interest (CPRATIO) is measured at the county level, we cluster heteroskedasticity-robust standard errors by county. We gradually include control variables and fixed effects in the specifications, with standard errors in all specifications clustered at the county level. Additionally, we winsorize all continuous variables at the 0.5 percent level and then standardize them.

Table 4 presents the results of regressions. Column (1) presents the basic model with only control of simultaneous second lien. Column (2) adds geographic controls only. Column (3) adds loan characteristics controls only. Column (4) adds all control variables. Column (5) to (7) adds originator fixed effects, state fixed effects, and half-year fixed effects gradually.

Our results reveal that counties with higher levels of gambling preference indeed tend to have more second-lien misrepresentation. A one standard deviation increase in CPRATIO in the whole sample leads to a 0.12 percent increase in the probability of second-lien misrepresentation without fixed effects and a 0.09 percent increase with state, half-year, and originator fixed effects[11]. The significance and economic magnitude also show that

[10]Although both originators and underwriters play important 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.

[11]The magnitude of the coefficient is small because the total misrepresentation rate is low (7.15 percent). The increase corresponds to a 1.67 percent rise compared to the total misrepresentation rate.

CPRATIO is a leading variable among all county-level variables[12]. It is also comparable to loan-level control variables, such as interest rate (-0.73 percent with all fixed effects) and FICO score (-0.62 percent with all fixed effects).

Simultaneous second lien is a key control in all specifications because, to misrepresent, one must first have a simultaneous second lien. Indeed, the coefficients of simultaneous second lien are significant in all columns, showing that about one-third of simultaneous second liens are unreported. Turning to geographic controls, we see that religiosity, income, and education are consistently negatively, though not always significantly, related to second-lien misrepresentation. In contrast, total population and the proportion of the elderly population are consistently positively, though not always significantly, related to this type of mortgage misrepresentation. The relationship of other geographic characteristics varies in different situations. Moreover, except for our variable of interest (CPRATIO), only total population remains robustly significant. The loan-level characteristics show that second-lien misrepresentation is associated with lower interest rates after controlling for simultaneous second liens, indicating potential gains from misrepresenting second liens. Low credit scores, low initial balances, and high LTV ratios are associated with unreported second liens, indicating that second-lien misrepresentation occurs more frequently in loans with these characteristics.

As borrowers in areas with higher levels of gambling preference are more likely to misreport second liens in general, we further investigate whether this holds in different subsamples. If behavioral traits played a role in the financial crisis irrespective of income class, we expect similar effects among borrowers in both lower-income and middle-income groups[13]. If a subsample contains loans that are more likely to be gambles, we expect the effect to be stronger in that subsample. We divide our samples in several ways: lower-class or middle-class, primary or non-primary, purchase or refinance, and high or low credit score. The

[12]The coefficient of CPRATIO is larger than that of REL, which has been proven to be an important factor in mortgage misrepresentation by Conklin et al. (2022), and even renders REL insignificant in columns (6) and (7).

[13]We do not examine high-income borrowers, as they fall outside the scope of the existing literature arguments and, from a theoretical standpoint, have limited incentive to engage in second-lien misrepresentation.

income class thresholds are calculated at the state level following the idea of the Pew Research Center[14]. We classify loans based on the income characteristics of their corresponding ZIP code areas: loans in areas with median household income below the lower bound of the middle-class threshold are designated as belonging to the lower-income class, while those within the middle-class range are categorized as middle-income class. The high or low credit score is divided by a FICO score of 670, which is the boundary between fair and good levels as evaluated by the institution[15]. We present the results in Table 5. For all specifications, we include all control variables and state, half-year, and originator fixed effects.

Columns (1) and (2) report the results for the lower-income and middle-income classes, respectively. A substantial proportion of loans fall within the middle-income category. As anticipated, the estimated effects are statistically significant in both income classes, with coefficients of similar magnitude. This finding suggests that behavioral traits play a role irrespective of income class.

Columns (3) and (4) report the results for primary and non-primary (i.e., fully owner-occupied vs. investment plus second-home) subsamples. Most observations in the whole sample belong to the primary subsample, and only the coefficient of CPRATIO in the primary subsample is significant[16]. This outcome implies that high gambling preference borrowers tend to choose second-lien misrepresentation mainly when their purpose is to fully owner-occupy the house. Indeed, if borrowers bet on the success of a misrepresentation for housing, the attractiveness of interest rate reduction for people with multiple houses should be much smaller than the attractiveness of homeownership for those trying to buy a primary home for living. Columns (5) and (6) present the results for purchase and refinance subsamples. About 43 percent of observations come from the purchase subsample, and only the coefficient of CPRATIO in the purchase subsample is significant. This implies that high gambling

[14]According to Pew Research Center, middle-class households are defined as those earning between two-thirds and twice the median household income for a given area.

[15]We also tried the median of the sample, 682, which yielded similar results.

[16]In unshown results, we also explore the investment subsample and second-home subsample separately, and neither is significant.

preference borrowers tend to misrepresent second liens mainly when purchasing a house. Similarly, owning a house (purchase) offers larger payoffs than refinancing the current home (refinance)[17]. Finally, columns (7) and (8) show the results for high and low credit score subsamples. About 44 percent of observations come from the low FICO score subsample. While the low FICO score subsample has a significant and larger coefficient for CPRATIO, the high FICO score subsample's result is insignificant. However, the economic magnitude of the coefficient in the high FICO score subsample is close to that in the whole sample. This implies a positive but noisy propensity for misrepresentation among borrowers with high credit scores, while the propensity is much clearer among low FICO score borrowers. Again, low FICO score borrowers find it harder to get first lien approval with a simultaneous second lien, so gambling for the loan is more necessary. In general, the subsample analysis indicates that when second-lien misrepresentation is more likely a high-payoff gamble, its rate increases more with gambling preference, supporting our conjecture about the correlation between misrepresentation and gambling preference.

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

In the previous section, we found a significant correlation between gambling preference and mortgage misrepresentation. Our next step is to investigate whether mortgage misrepresentation associated with gambling preference has a meaningful influence on loan performance. While mortgage misrepresentation usually leads to worse loan performance (Piskorski et al., 2015; Griffin and Maturana, 2016), it is unclear whether mortgage misrepresentation associated with gambling preference leads to higher or lower default rates. Since gambling, in general, is associated with negative outcomes, we conjecture that gambling-related misrep-

[17]In unshown results, the coefficient is insignificant in the cash-out refinance subsample but significant in the no-cash-out refinance subsample. However, most observations are in the cash-out refinance subsample.

resentation will be associated with worse loan performance. We formalize this inference in Hypothesis 2:

Hypothesis 2 *Second-lien misrepresented loans in higher levels of local gambling preference counties are more likely to default.*

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

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

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. For independent variables, $Misrepresentation_{it}$ is the second-lien misrepresentation indicator 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 misrepresentation measure and gambling preference measure; X_{it} includes loan i's correctly reported simultaneous second, 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 winsorized at the 0.5 percent level and then standardized. We include control variables, fixed effects, and standard error clustering in all specifications, but we gradually include the mortgage misrepresentation term, gambling preference term, and the interaction term.

The results are reported in Table 6. In column (1), we include the mortgage misrepresentation measure only to see if second-lien misrepresentation affects loan performance in our sample. Column (2) includes the gambling preference measure only to see if areas with high levels of gambling preference also have high default rates. Column (3) includes both the mortgage misrepresentation measure and the gambling preference measure. Finally, we include the interaction term between the mortgage misrepresentation measure and the gambling preference measure in column (4) to test whether second-lien misrepresentation associated with gambling preference affects the default rate.

The coefficient of second-lien misrepresentation, which is significantly positive in all columns, shows that mortgage loans with misreported second liens are more likely to default than loans without simultaneous second liens. The coefficient of correctly reported second liens is also significantly positive, consistent with the findings in Griffin and Maturana (2016). The coefficient of CPRATIO shows that areas with higher levels of gambling preference do not necessarily have higher default rates. The effects of local gambling preference on default manifest through its influence on second-lien misrepresentation. The coefficient of the interaction term is positive and statistically significant, indicating that second-lien misrepresented loans in counties with higher levels of local gambling preference are more likely to become delinquent. Thus, gambling in mortgage applications likely leads to worse outcomes, similar to other financial settings.

Since we found that the effect of gambling preference on second-lien misrepresentation may vary in subsamples in the previous section, we further investigate whether this variation also exists for loan performance. The results presented in Table 7 show that the interaction term is positive and statistically significant in most subsamples. As anticipated, the interaction term is significant for both the lower-income and middle-income classes, indicating that the influence of behavioral traits on default behavior persists irrespective of income class. The insignificance in the non-primary subsample is plausible because second-lien misrepresentation is not strongly associated with gambling preference in this subsample. It also indicates that gambling to misrepresent in this small subsample does not significantly impact loan performance. In contrast, although second-lien misrepresentation is not significantly associated with gambling preference in the refinance and high FICO subsamples, the default rate is significantly associated with the interaction term between misrepresentation and gambling preference, implying that gambling to misrepresent is indeed associated with worse loan performance in these two subsamples. In general, these results indicate that gambling is usually associated with worse outcomes.

5 Is this a borrower issue or lender issue?

While we find a positive correlation between gambling preference and second-lien misrepresentation, an important question is whether this is primarily a borrower issue. Previous studies examining second-lien misreporting in privately securitized loans show that both originators and underwriters play significant roles in mortgage misrepresentation (Griffin and Maturana, 2016; Piskorski et al., 2015). They are likely aware of hidden second liens but still misreport them. Additionally, by comparing lenders' portfolio loans and privately securitized mortgages, Yavas and Zhu (2024) provide strong evidence that second-lien misrepresentation occurs in the early stages of intermediation by lenders rather than underwriters. As a result, the relationship we find could be driven by differences among lenders rather than borrowers.

To address this issue, we employ the setting of ease of securitization around a FICO score of 620, as described by Keys et al. (2010). Due to the underwriting guidelines established by government-sponsored enterprises, Fannie Mae and Freddie Mac, low documentation loans made to borrowers with FICO scores of 620 or higher are much easier to securitize. Consequently, lenders have lax screening standards for such loans, and a regression discontinuity design could reveal an upward jump in the default rate for low documentation loans. Additionally, Griffin and Maturana (2016) find that the second-lien misrepresentation rate also jumps at this cutoff, implying that this type of misrepresentation is associated with lenders' incentives to securitize the loan.

Utilizing this shock to lenders, we first study the extent to which lenders facilitate second-lien misrepresentation by comparing loans with misrepresented simultaneous seconds to other types of loans. Specifically, we examine the jumps in the number of loans and the jumps in the default rate for all loans, loans without simultaneous seconds, loans with correctly reported simultaneous seconds, and loans with misrepresented simultaneous seconds. If lenders specifically facilitate second-lien misrepresentation, then the jump ratio in the number of loans with misrepresented simultaneous seconds should be greater than for other types of loans. Additionally, if such misrepresentation is caused by lenders' screening efforts, the jumps in the default rate for loans with misrepresented simultaneous seconds should be significantly positive. In contrast, if the jump ratio in the number of loans with misrepresented simultaneous seconds is similar to or smaller than for other types of loans, lenders do not specifically facilitate second-lien misrepresentation but increase it with the increase of other loans. If the jumps in the default rate for loans with misrepresented simultaneous seconds are insignificant from zero, such misrepresentation is not mainly driven by lenders' screening efforts.

Second, we explore whether the effect of gambling preference on second-lien misrepresentation is mainly driven by lenders' differences or borrowers' preferences. Since the variation in second-lien misrepresentation related to gambling preference could come from either lenders or borrowers, and the shock we employed is only for lenders, if the increase in second-lien misrepresentation in high gambling preference areas is smaller, it indicates that lenders play a smaller role in increasing misrepresentation in high gambling preference areas. This implies that the positive relationship between gambling preference and second-lien misrepresentation is mainly attributed to borrowers. Specifically, we examine the jumps in the number of loans and the jumps in the default rate to understand how lender's roles differ across areas. If the effect of local gambling preference on second-lien misrepresentation is mainly driven by lenders' differences (i.e., lenders make second-lien misrepresentation more prevalent in high gambling preference areas), this shock to lenders would lead to different outcomes between high and low gambling preference areas, favoring high gambling preference areas (i.e., a greater increase in the number of loans with misrepresentation and a larger increase in the default rate). In contrast, if the effect is mainly driven by borrower's preferences, this shock to lenders would lead to less favorable results for high gambling preference areas. In other words, a smaller increase in the number of misrepresented loans and the default rates of misrepresented loans in high gambling preference areas indicates that lenders did not facilitate more misrepresentation in such areas.

We take RDD approach using local linear regressions. The credit scores are normalized as follows:

$$C = Credit\ Score - Threshold \tag{3}$$

where $Threshold$ is 620 for low documentation loans. Then, we define D to distinguish credit scores that are over or below the threshold as follows:

$$D = \begin{cases} 1, & if\ C \geq 0 \\ 0, & otherwise \end{cases} \tag{4}$$

To differentiate the effect between high and low local gambling preference areas, we define high local gambling preference ares:

$$G = \begin{cases} 1, & if\ CPRATIO \geq 1.2921 \\ 0, & otherwise \end{cases} \tag{5}$$

We use 1.2921 as the cutoff for two reasons[18]. First, only 10 percent of county-years in our sample have a CPRATIO greater than or equal to 1.2921, indicating that areas with such a CPRATIO indeed have a high local gambling preference. Second, about half of all loans in our sample are originated in these areas, making our findings about high and low gambling preference areas comparable.

To apply the RDD approach using local linear regressions, we choose a bandwidth of 10, which corresponds to a FICO score range of 610 to 630. We selected this window because

[18]We also tried other reasonable cutoffs, such as 1, and the results are robust.

it ensures that no other jumps are found in the literature[19] and it is close to the optimal bandwidth generated by a data-driven bandwidth selection method[20] (Calonico et al., 2020). We use the fixed bandwidth rather than the optimal bandwidth so that the RDD results in different cases are comparable[21]. For the default rate and number of loans in high and low gambling preference areas, we estimate the four cases separately using the following specification:

$$Y = \alpha + \beta D + \gamma C + \delta D \times C + \epsilon \tag{6}$$

where Y is the default dummy variable for loan i originated at time t or the number of loans at each FICO score. In our baseline model, we use uniform kernel for estimation.

Finally, to calculate the difference in the discontinuities (diff-in-disc) between high and low gambling preference areas, we follow Dickert-Conlin and Elder (2010) and Grembi et al. (2016) to estimate the following model:

$$Y = \beta_0 + \beta_1 D + \beta_2 C + \beta_3 D \times C + \beta_4 G + \beta_5 G \times C + \beta_6 G \times D + \beta_7 G \times C \times D + \epsilon \tag{7}$$

where β_6 is the parameter that measures the difference in the discontinuity in Y_{it} at credit score cutoff for high versus low gambling preference areas. When we apply this equation to calculate the difference in jump ratios between high and low gambling preference areas, we first rescale the data to obtain the t-statistics. We divide the data in high gambling preference areas by the estimated value at FICO 620^{-} in those areas. Similarly, we divide the data in low gambling preference areas by their corresponding estimated value at FICO 620^{-}. After this rescaling, the discontinuity estimated in high or low gambling preference areas can be directly interpreted as a multiple of its corresponding estimated value at FICO

[19]For example, FICO scores of 600 and 660 could be other potential cutoffs for jumps.
[20]The optimal bandwidth for the number of all loans is 10.50.
[21]Using the optimal bandwidth does not affect our findings.

620^-. The difference between the two discontinuities represents the difference in jump ratios between high and low gambling preference areas.

We first look at the jumps in the number of loans and the jumps in the default rate in all areas, examining the pattern of second-lien misrepresentation given a shock to lenders. Panel A of Table 8 presents the results for jumps in the total number of loans, loans without simultaneous seconds, loans with correctly reported simultaneous seconds, and loans with misrepresented simultaneous seconds. Overall, the total number of low-documentation loans doubles from 620^- to 620^+, consistent with the findings in Keys et al. (2010). Loans with simultaneous seconds increase more than those without, whether correctly reported or misrepresented. This shows that the ease of securitization is indeed applicable to such loans. Moreover, the jump ratio of misrepresented simultaneous seconds is less than the jump ratio of correctly reported simultaneous seconds, indicating that lenders do not facilitate second-lien misrepresentation more than they do correctly reported seconds. Panel B further looks at the jumps in the default rate, which reflects the laxity of lender screening. Consistent with previous literature, there is a jump of 0.037 in the default rate for all loans, indicating overall reduced screening for 620^+ loans. This lax screening is primarily observed in loans with correctly reported simultaneous seconds, with a jump of 0.108. For the loans with misrepresented simultaneous seconds, the jump in the default rate is small and insignificant (0.014), indicating a minor reduction in screening. These findings illustrate that lenders facilitate loans with simultaneous second liens, not limited to misrepresenting seconds, and that lenders' screening effort is not the main factor driving second-lien misrepresentation.

After looking at the jumps in all areas, we further explore the patterns between high and low gambling preference areas to determine whether it is the effect of lenders' differences or borrowers' preferences. Table 9 shows the results for the number of loans in Panel A and the results for default rates in Panel B. For the number of loans, the magnitude of increases is similar between high and low gambling preference areas for all loans, loans without simultaneous seconds, and loans with correctly reported simultaneous seconds. This indicates no systematic difference between high and low gambling preference areas in ease of securitization. In contrast, the magnitude of the increase in the number of loans with misrepresented seconds in high gambling preference areas is much smaller than in low gambling preference areas. Given that the ease of securitization is a shock to lenders (Keys et al., 2010) and lenders facilitate misrepresentation with the objective of securitization (Griffin and Maturana, 2016), this result suggests that lenders do not play a larger role in increasing second-lien misrepresentation in high gambling preference areas[22]. For default rates, the jumps between high and low gambling preference in all types of loans do not show significant differences. Additionally, for loans with misrepresented simultaneous seconds, the jump in high gambling preference areas (0.001) is smaller than in low gambling preference areas (0.052), although not significant. This suggests that lenders in high gambling preference areas are not more likely to facilitate misrepresentation by having laxer screening in such areas. Therefore, combined with the evidence from the number of loans, these findings suggest that lenders in high gambling preference areas are not more inclined to increase such misrepresentation than lenders in low gambling preference areas. The findings support the notion that the higher likelihood of second-lien misrepresentation in high gambling preference areas is more likely due to borrower preferences rather than lender differences.

6 Robustness

6.1 Alternative Gambling Preference Measures

While CPRATIO serves as our primary measure of gambling preference, we also employ alternative measures to address concerns that CPRATIO may reflect broader religious differences rather than gambling-specific tendencies. Following studies such as Markham et al. (2023),

[22]Since borrowers play a larger role in increasing second-lien misrepresentation in high gambling preference areas, lenders' effect in increasing the number is more limited, making the magnitude of the jump smaller. In contrast, since borrowers play a smaller role in increasing second-lien misrepresentation in low gambling preference areas, lenders' effect in increasing the number is less limited, making the magnitude of the jump bigger

which use the density of gambling outlets as a measure of gambling exposure, we incorporate employment and the number of gambling establishments as alternative measures that reflect gambling preference through local supply. Specifically, we utilize per capita employment in the gambling industry and per capita number of gambling establishments at the core-based statistical area (CBSA) level. These alternatives help isolate local gambling behavior more directly. We regress the second-lien misrepresentation dummy variable on these alternative measures, alongside geographic controls at the CBSA level, loan-level controls, and fixed effects (i.e., state, lender, and half-year). Additionally, we include controls for employment and establishment counts in casino hotels to account for tourist-driven gambling demand[23]. We also include nonemployer gambling establishments to capture complementary gambling demand that may be excluded by the primary measures due to database limitations. See Appendix A for detailed descriptions of the data and empirical tests.

Table 10 presents the results of the primary regressions. Columns (1) to (5) report specifications using per capita employment in the gambling industry, per capita number of gambling establishments, per capita employment in casino hotels, per capita number of casino hotel establishments, and the number of nonemployer gambling establishments, respectively. Columns (6) and (7) include both employment and establishment metrics for the gambling industry and casino hotels, as well as nonemployer gambling establishments. The regression results indicate that both employment and establishment metrics for the gambling industry are significantly positive, suggesting a strong link between local gambling preference and the likelihood of second-lien misrepresentation. By contrast, casino hotel variables, which reflect a mix of local and tourist demand, yield coefficients that are generally insignificant. Although nonemployer gambling businesses primarily reflect local demand, they tend to offer limited gambling access, and their associated coefficients are likewise insignificant. Overall, we find that per capita employment and establishment measures in the gambling industry, which

[23]In the CBP database, casino hotels are categorized separately from the gambling industry. Because casino hotel employees also provide hotel-related services beyond gambling, it is necessary to control for casino hotels separately.

are used as alternative proxies for local gambling preference, support the hypothesis that second-lien misrepresentation is more prevalent in regions with elevated gambling preference.

6.2 Local Tolerance for Law-breaking Behaviors

Since second-lien misrepresentation is a type of misrepresentation that violates the law, our findings could potentially be driven by the possibility that our gambling measure inadvertently captures variation in the general local tolerance for law-breaking behavior. To rule out this channel, we test whether CPRATIO is associated with local crime rates for several types of criminal activity, including violent crime, property crime, arson, fraud, and illegal gambling, within our sample period. See Appendix B for a detailed description of the data.

We regress crime rates for these various types of criminal activity on CPRATIO, controlling for other geographic characteristics. Table 11 reports the regression results. We find that CPRATIO is either negatively significant or statistically insignificant in all regressions except for illegal gambling. This result is consistent with expectations, as illegal gambling is closely aligned with gambling preference, while other types of crime are not expected to be attractive to individuals with high gambling preference. These findings suggest that our results are not primarily driven by variation in general local tolerance for law-breaking behavior captured by CPRATIO.

6.3 Alternative Measures of Default

Borrowers can default on their mortgages for various reasons over different time horizons. Our measure of default uses the MBA method of delinquency for 90 days or more and is restricted to the first three years after origination. By adjusting the definition to be more restricted (fewer default cases) or less restricted (more default cases), we can investigate whether the effect found in section 4 is generally applicable. Therefore, we consider the following default measures: 90 days or more delinquency using the MBA method in the first two years after origination (more restricted), 60 days or more delinquency using the 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 12 presents the results of different default measures. We also include the results of the original measure in column (1). The significance of second-lien misrepresentation and its interaction with CPRATIO are robust. Additionally, when the default measure is less restricted, the effect of mortgage misrepresentation is greater. The effect of gambling preference-associated mortgage misrepresentation is also greater when the default measure is less restricted.

6.4 Nonlinear Model

The results in the previous sections are estimated using a linear probability model (OLS). To ensure our findings are not driven by this modeling choice, we also use a nonlinear specification (probit) for inference. We present the marginal effects of CPRATIO in Table 13. Since we control for simultaneous second liens in the regression, the probit model drops all observations without simultaneous seconds due to perfect failure prediction of misrepresentation. The marginal effects of CPRATIO remain significant.

6.5 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 the causal forest approach proposed by Wager and Athey (2018). This method uses an augmented inverse propensity-weighted estimator with the random forest method in machine learning, incorporating an honesty condition. It provides double robustness (compared to propensity score matching) and high efficiency for high-dimensional models (compared to nearest-neighbor matching) in observational settings[24]. This approach has also been used in finance, such as in Rampini and Viswanathan (2022) for secured debt and Gulen et al. (2021) for corporate finance, and has shown better performance than traditional causal inference designs (Gulen et al., 2021) in terms of accuracy.

We first use this approach to investigate the causality between gambling preference and mortgage misrepresentation. In the causal forest approach, we set mortgage misrepresentation as the outcome and CPRATIO as the treatment. We use all control variables from Table 4 as matching variables to help grow trees and forests[25]. All continuous variables, including CPRATIO, are winsorized at the 0.5 percent level and then standardized. Since fixed effects are not applicable in this approach, we create a variable for half-year periods, starting from the first half of 2005 as 1, to account for potential time effects. Similar to Athey and Wager (2019), who cluster observations by school ID, we cluster observations by state but give each unit the same weight so that larger clusters receive more weight[26]. By using such cross-fitting, our results do not solely come from any single state, but states with a greater number of observations do have greater weights. To balance computational capacity and the accuracy of confidence intervals, we grow 2000 trees for the forest, which is also the default setting in Athey and Wager (2019). For the honesty property, we use the default splitting fraction: for each sample, we use 50 percent of the data for splitting and the remaining data for estimation. Table 14 shows the results using the causal forest. The coefficients in the regressions of second-lien misrepresentation remain significant, and the magnitude is similar to those 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

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

[25]We require observations to be non-missing for all variables.

[26]Due to the limitation of the grf package in R, which requires the matching variables to be numerical, we do not create a numerical variable for originators to avoid falsely treating closer values as shorter distances. We also do not cluster observations by originator because the large number of originators would excessively increase the computational burden.

outcome and mortgage misrepresentation as the treatment. Except for adding CPRATIO to the matching variable matrix, the other matching variables and clustering variables are the same as those in the above causal forest regressions. For the same considerations, we also grow 2000 trees for the forest and use 50 percent of the data in each sample for splits. After estimating the causal forest, the average treatment effects of mortgage misrepresentation on default are calculated for each quarter of the sample sorted by CPRATIO. Figure 3 shows the results for second-lien misrepresentation. The treatment effects are generally larger when CPRATIO is greater, which is consistent with the results from the OLS model.

6.6 Diff-in-disc concerns

To validate our application of the diff-in-disc approach in comparing loan volumes and default rates between high and low gambling preference areas, we address several methodological concerns. First, we examine and test the three core assumptions outlined by Grembi et al. (2016) regarding the implementation of the diff-in-disc framework. Second, we assess the sensitivity of our results to key modeling choices, including the inclusion of control variables, bandwidth selection, and estimation kernel specification. Our findings remain robust across these variations. Third, we perform a placebo test to confirm that the observed effects are not driven by random variation. For full details of these robustness checks and supporting evidence, see Appendix C.

7 Conclusion

Many years after the 2008 financial crisis, the housing market continues to exhibit cyclical dynamics and remains a significant driver of global economic conditions. Although considerable progress has been made in understanding the origins of the crisis, debate persists over which segments of society played the central role. By shifting the focus from income classes to behavioral traits, this paper highlights a trait-based behavioral channel that complements the credit expansion and expectations views of the crisis.

Using a religion-based proxy for local gambling preferences, we show that gambling preference, a persistent behavioral trait, is associated with higher rates of second-lien misrepresentation across borrower classes, ultimately leading to poorer loan performance. Our findings suggest that behavioral predispositions shape borrower actions in ways that contribute to systemic risk. Recognizing these traits is essential for designing policies that address the behavioral foundations of financial instability.

References

Adelino, M., Schoar, A., Severino, F., 2018. The role of housing and mortgage markets in the financial crisis. Annual Review of Financial Economics 10, 25–41.

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.

Armstrong, T., Rockloff, M., Browne, M., 2020. Gamble with your head and not your heart: A conceptual model for how thinking-style promotes irrational gambling beliefs. Journal of Gambling Studies 36, 183–206.

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

Badarinza, C., Campbell, J.Y., Ramadorai, T., 2016. International comparative household finance. Annual Review of Economics 8, 111–144.

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., 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. The Quarterly Journal of Economics 110, 73–92.

Blaszczynski, A., Nower, L., 2002. A pathways model of problem and pathological gambling. Addiction 97, 487–499.

Bostic, R.W., Lee, K.O., 2008. Mortgages, risk, and homeownership among low- and moderate-income families. American Economic Review 98, 310–314.

Burnside, C., Eichenbaum, M., Rebelo, S., 2016. Understanding booms and busts in housing markets. Journal of Political Economy 124, 1088–1147.

Bush, G.W., 2003. Remarks on signing the american dream down-payment act. URL: `https://www.presidency.ucsb.edu/documents/remarks-signing-the-american-dream-downpayment-act`.

Calonico, S., Cattaneo, M.D., Farrell, M.H., 2020. Optimal bandwidth choice for robust bias corrected inference in regression discontinuity designs. Econometrics Journal 23, 192–210.

Calonico, S., Cattaneo, M.D., Titiunik, R., 2014. Robust nonparametric confidence intervals for regression-discontinuity designs. Econometrica 82, 2295–2326.

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.

Christensen, D.M., Jones, K.L., Kenchington, D.G., 2018. Gambling attitudes and financial misreporting. Contemporary Accounting Research 35, 1229–1261.

Clinton, W.J., 1995. Remarks on the national homeownership strategy. URL: https://www.presidency.ucsb.edu/documents/remarks-the-national-homeownership-strategy.

Cocco, J.F., 2005. Portfolio choice in the presence of housing. Review of Financial Studies 18, 535–567.

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

Cressey, D.R., 1953. Other People's Money: A Study in the Social Psychology of Embezzlement. Free Press, Glencoe, IL.

Dickert-Conlin, S., Elder, T., 2010. Suburban legend: School cutoff dates and the timing of births. Economics of Education Review 29, 826–841.

Eckert, F., Fort, T.C., Schott, P.K., Yang, N.J., 2021. Imputing missing values in the us census bureau's county business patterns. Working paper.

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

Farlow, A., 2004. The U.K. Housing Market: Bubbles and Buyers. Technical Report. Credit Suisse First Boston Housing Market Conference. Paper presented at Credit Suisse First Boston Housing Market Conference, January 2004.

Farlow, A., 2013. Crash and Beyond: Causes and Consequences of the Global Financial Crisis. Oxford University Press, Oxford, UK.

FBI, 2006. Mortgage Fraud Report 2006. Technical Report. Federal Bureau of Investigation.

FBI, 2007. Mortgage Fraud Report 2007. Technical Report. Federal Bureau of Investigation.

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

Genesove, D., Mayer, C.J., 2001. Loss aversion and seller behavior: Evidence from the housing market. The Quarterly Journal of Economics 116, 1233–1260.

Gomes, F., Haliassos, M., Ramadorai, T., 2021. Household finance. Journal of Economic Literature 59, 919–1000.

Goodman, L.S., Mayer, C., 2018. Homeownership and the american dream. Journal of Economic Perspectives 32, 31–58.

Grembi, V., Nannicini, T., Troiano, U., 2016. Do fiscal rules matter? American Economic Journal: Applied Economics 8, 1–30.

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.

Justiniano, A., Primiceri, G.E., Tambalotti, A., 2015. Credit supply and the housing boom. Journal of Political Economy 123, 379–413.

Keys, B.J., Mukherjee, T., Seru, A., Vig, V., 2010. Did securitization lead to lax screening? evidence from subprime loans. The Quarterly Journal of Economics 125, 307–362.

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.

Kumar, A., Page, J.K., Spalt, O.G., 2016. Gambling and comovement. Journal of Financial and Quantitative Analysis 51, 85–111.

Lux, T., 1995. Herd behaviour, bubbles and crashes. The Economic Journal 105, 881–896.

Markham, F., Gobaud, A.N., Mehranbod, C.A., Morrison, C.N., 2023. Casino accessibility and suicide: A county-level study of 50 us states, 2000 to 2016. Addiction 118.

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.

Mian, A., Sufi, A., 2009. The consequences of mortgage credit expansion: Evidence from the u.s. mortgage default crisis. The Quarterly Journal of Economics 124, 1449–1496.

Mishra, S., Lalumière, M.L., Williams, R.J., 2010. Gambling as a form of risk-taking: Individual differences in personality, risk-accepting attitudes, and behavioral preferences for risk. Personality and Individual Differences 49, 616–621.

Parrado-González, A., Fernández-Calderón, F., Newall, P.W.S., León-Jariego, J.C., 2023. Peer and parental social norms as determinants of gambling initiation: A prospective study. Journal of Adolescent Health 73, 296–301.

Piazzesi, M., Schneider, M., 2016. Housing and macroeconomics. National Bureau of Economic Research, Cambridge, Mass.

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.

Raylu, N., Oei, T.P., 2004. Role of culture in gambling and problem gambling. Clinical Psychology Review 23, 1087–1114.

Rose, I.N., Loeb, R.A., 1999. Blackjack and the Law. Oakland, CA : RGE Pub, Las Vegas: Huntington Press.

Shiller, R.J., 2005. Irrational Exuberance. 2 ed., Princeton University Press, Princeton, NJ.

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

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

Yao, R., Zhang, H.H., 2005. Optimal consumption and portfolio choices with risky housing and borrowing constraints. Review of Finanical Studies 18, 197–239.

Yavas, A., Zhu, S., 2024. Misreporting of second liens in portfolio mortgages and privately securitized mortgages. Real Estate Economics 52, 73–109.

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
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

	N	Mean	SD	P10	P25	P50	P75	P90
Simultaneous second liens (loan level)								
Misrepresented (%)	3031921	7.15	25.76					
Correct presented (%)	3031921	13.91	34.61					
All (%)	3031921	21.06	40.77					
Loan characteristis (loan level)								
Interest rate (%)	3026958	6.41	2.12	2.75	5.88	6.50	7.55	8.75
FICO	2993630	681.78	72.60	578.00	632.00	688.00	739.00	776.00
Balance (ln)	3031921	12.40	0.73	11.44	11.87	12.40	12.98	13.31
LTV (%)	3031921	74.82	13.34	56.65	70.00	80.00	80.00	90.00
ARM (%)	3031921	58.87	49.21					
Option ARM (%)	3031921	0.24	4.84					
Negative amortization (%)	3031921	13.58	34.26					
Low or no doc. (%)	3031921	65.42	47.56					
Prepayment penalty (%)	3031921	40.99	49.18					
Cash-out (%)	3031921	40.66	49.12					
No-cash-out (%)	3031921	14.73	35.44					
Investment (%)	3031921	10.27	30.36					
Second-home (%)	3031921	3.55	18.49					
Default (%)	3014278	23.51	42.40					
Geographic characteristics (county-year level)								
CPRATIO	8716	0.51	0.88	0.01	0.04	0.19	0.54	1.29
REL (%)	8716	59.59	17.64	36.62	45.90	58.51	72.42	89.21
Geographic characteristics (county-month level)								
unemployment (%)	71156	5.14	1.73	3.20	3.90	4.90	6.00	7.50
Geographic characteristics (county level)								
Education (%)	3035	16.64	7.51	9.70	11.20	14.40	19.30	26.60
Married (%)	3035	60.39	5.10	53.90	57.90	61.30	63.90	66.00
Urban (%)	3035	40.62	30.54	0.00	12.93	40.32	64.65	84.67
Total population (ln)	3035	10.49	1.11	9.52	9.52	10.15	11.05	12.09
Over65 (%)	3035	14.71	4.05	9.90	12.10	14.40	17.00	20.00
Male-female ratio	3035	0.98	0.06	0.92	0.94	0.97	1.00	1.05
Minority (%)	3035	15.32	15.80	2.02	3.34	8.71	22.96	37.90
Geographic characteristics (zipcode-year level)								
HPA	76535	18.45	14.28	4.23	7.55	13.93	26.56	39.99
Geographic characteristics (zipcode level)								
Income (ln)	26517	10.57	0.34	10.17	10.34	10.53	10.77	11.02

The table presents descriptive statistics for the variables used in our study, covering the sample period from 2005 to 2007. For all continuous variables, we report the number of observations, mean, standard deviation, 25th percentile, 50th percentile, and 75th percentile. For all dummy variables, we report the number of observations, mean, and standard deviation. The variables are winsorized at the 0.5 percent level. (ln) indicates that the value of the variable is the natural logarithm of the original value. (%) indicates that the value of the variable is expressed as a percentage.

Table 3
Geographic Variables Correlation

	CPRATIO	REL	Income	Education	Married	Urban	Total population	Over65	Male-female ratio	Minority	Unemployment	HPA
CPRATIO	1.00											
REL	-0.01	1.00										
Income	0.26	-0.11	1.00									
Education	0.26	-0.10	0.64	1.00								
Married	-0.19	0.12	0.13	-0.30	1.00							
Urban	0.32	-0.01	0.41	0.52	-0.40	1.00						
Total population	0.32	-0.18	0.50	0.53	-0.37	0.74	1.00					
Over65	-0.15	0.31	-0.42	-0.34	0.23	-0.33	-0.37	1.00				
Male-female ratio	0.05	-0.21	0.12	-0.01	0.18	-0.17	-0.18	-0.21	1.00			
Minority	0.08	-0.09	-0.15	0.02	-0.52	0.20	0.23	-0.30	-0.12	1.00		
Unemployment	-0.06	-0.20	-0.34	-0.40	-0.14	-0.17	-0.08	0.02	-0.07	0.22	1.00	
HPA	0.26	-0.29	0.18	0.23	-0.08	0.15	0.23	-0.01	0.11	0.14	-0.21	1.00

The table shows the correlation between geographic variables at the county-year level. Income and HPA at the county level are obtained from the same data sources as the zipcode level data, which are the U.S. Census Bureau and the Federal Housing Finance Agency, respectively. Unemployment is the average of the monthly unemployment rate within each year.

Table 4
Gambling Preference and Second-lien Misrepresentation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
CPRATIO	0.0012***	0.0023***	0.0023***	0.0022***	0.0016***	0.0012**	0.0009**
	(0.0005)	(0.0005)	(0.0003)	(0.0005)	(0.0003)	(0.0006)	(0.0004)
Simul. Second	0.3396***	0.3408***	0.3370***	0.3388***	0.3098***	0.3105***	0.3131***
	(0.0020)	(0.0019)	(0.0019)	(0.0018)	(0.0019)	(0.0019)	(0.0019)
REL		-0.0019***		-0.0015***	-0.0012***	-0.0004	-0.0001
		(0.0006)		(0.0005)	(0.0004)	(0.0004)	(0.0003)
Unemployment		0.0041***		0.0035***	0.0018***	0.0051***	-0.0006
		(0.0007)		(0.0007)	(0.0005)	(0.0009)	(0.0005)
HPA		-0.0071***		-0.0050***	-0.0008	-0.0034***	0.0017***
		(0.0010)		(0.0008)	(0.0006)	(0.0005)	(0.0005)
Education		0.0004		0.0011**	0.0003	0.0015***	0.0002
		(0.0006)		(0.0005)	(0.0004)	(0.0005)	(0.0003)
Married		0.0006		-0.0000	0.0004	0.0004	-0.0000
		(0.0006)		(0.0006)	(0.0004)	(0.0005)	(0.0003)
Income		-0.0003		-0.0002	-0.0004*	-0.0004	-0.0004*
		(0.0003)		(0.0003)	(0.0002)	(0.0003)	(0.0002)
Urban		0.0005		-0.0000	0.0000	-0.0009**	-0.0003
		(0.0005)		(0.0004)	(0.0003)	(0.0004)	(0.0003)
Total population		0.0026***		0.0028***	0.0025***	0.0039***	0.0009**
		(0.0009)		(0.0008)	(0.0006)	(0.0005)	(0.0004)
Over65		0.0035***		0.0036***	0.0018***	0.0001	0.0005*
		(0.0005)		(0.0004)	(0.0003)	(0.0004)	(0.0003)
Male-female ratio		0.0003		0.0005	0.0011***	-0.0000	-0.0001
		(0.0006)		(0.0005)	(0.0004)	(0.0004)	(0.0003)
Minority		-0.0006		-0.0006	0.0001	-0.0020***	0.0005
		(0.0008)		(0.0007)	(0.0005)	(0.0007)	(0.0004)
Interest rate			-0.0127***	-0.0119***	-0.0121***	-0.0113***	-0.0073***
			(0.0005)	(0.0004)	(0.0006)	(0.0005)	(0.0004)
FICO			-0.0092***	-0.0093***	-0.0071***	-0.0072***	-0.0062***
			(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)
Balance			-0.0005	0.0011***	0.0030***	0.0027***	0.0048***
			(0.0004)	(0.0004)	(0.0002)	(0.0003)	(0.0002)
LTV			0.0040***	0.0036***	0.0056***	0.0058***	0.0056***
			(0.0004)	(0.0004)	(0.0003)	(0.0003)	(0.0003)
ARM			0.0022***	0.0017**	0.0121***	0.0114***	0.0088***
			(0.0008)	(0.0009)	(0.0008)	(0.0008)	(0.0006)
Negative amortization			-0.0301***	-0.0281***	-0.0151***	-0.0130***	-0.0043***
			(0.0013)	(0.0013)	(0.0020)	(0.0018)	(0.0012)
Option ARM			-0.0486***	-0.0492***	-0.0433***	-0.0427***	-0.0227***
			(0.0031)	(0.0032)	(0.0046)	(0.0046)	(0.0049)
Prepayment penalty			-0.0093***	-0.0086***	-0.0096***	-0.0106***	-0.0081***
			(0.0008)	(0.0009)	(0.0008)	(0.0008)	(0.0008)
Low or no doc.			-0.0027***	-0.0023***	-0.0069***	-0.0072***	-0.0080***
			(0.0006)	(0.0005)	(0.0005)	(0.0005)	(0.0004)
Cash-out			0.0178***	0.0182***	0.0182***	0.0180***	0.0182***
			(0.0006)	(0.0006)	(0.0005)	(0.0005)	(0.0006)
No-cash-out			0.0151***	0.0144***	0.0151***	0.0153***	0.0158***
			(0.0010)	(0.0010)	(0.0007)	(0.0007)	(0.0007)
Investment			0.0232***	0.0234***	0.0256***	0.0252***	0.0258***
			(0.0010)	(0.0011)	(0.0009)	(0.0009)	(0.0009)
Second-home			0.0108***	0.0106***	0.0138***	0.0138***	0.0157***
			(0.0012)	(0.0011)	(0.0010)	(0.0010)	(0.0010)
Originator FE	N	N	N	N	Y	Y	Y
State FE	N	N	N	N	N	Y	Y
Half-year FE	N	N	N	N	N	N	Y
Observations	3,031,921	2,907,303	2,991,961	2,869,014	2,384,686	2,384,686	2,384,686
Adj. R^2	0.289	0.291	0.282	0.284	0.313	0.313	0.318

The table shows OLS estimates from regressions where the dependent variable takes a value of one if the loan is recognized as second-lien misrepresented and zero otherwise. Specific fixed effects are used if indicated by Y and not used if indicated by N. Variables are defined in Table 1. All continuous variables are winsorized at the 0.5 percent level and then standardized. Standard errors clustered at the county level are reported in parentheses. $*p < 0.10$, $**p < 0.05$, and $***p < 0.01$.

Table 5
Gambling Preference and Second-lien Misrepresentation - Subsamples

	(1) Lower-class	(2) Middle-class	(3) Primary	(4) Non-primary	(5) Purchase	(6) Refinance	(7) High FICO	(8) Low FICO
CPRATIO	0.0010**	0.0009*	0.0009**	0.0005	0.0014**	0.0004	0.0007	0.0014***
	(0.0005)	(0.0005)	(0.0004)	(0.0007)	(0.0007)	(0.0003)	(0.0006)	(0.0005)
Simul. Second	0.319***	0.312***	0.3032***	0.4117***	0.2775***	0.3831***	0.2823***	0.3581***
	(0.0030)	(0.0019)	(0.0020)	(0.0062)	(0.0018)	(0.0032)	(0.0025)	(0.0019)
REL	-0.0005	0.0001	-0.0001	-0.0003	-0.0004	0.0001	-0.0007*	0.0003
	(0.0004)	(0.0004)	(0.0003)	(0.0006)	(0.0005)	(0.0002)	(0.0004)	(0.0004)
Unemployment	-0.0005	-0.0006	-0.0006	-0.0008	-0.0007	-0.0000	-0.0011*	-0.0001
	(0.0006)	(0.0006)	(0.0005)	(0.0006)	(0.0009)	(0.0003)	(0.0006)	(0.0005)
HPA	0.0007	0.0024***	0.0017***	-0.0000	0.0021**	0.0019***	0.0021***	0.0021***
	(0.0006)	(0.0007)	(0.0006)	(0.0007)	(0.0010)	(0.0004)	(0.0007)	(0.0008)
Education	0.0001	0.0002	0.0003	-0.0013**	0.0002	0.0002	-0.0002	0.0007*
	(0.0007)	(0.0003)	(0.0003)	(0.0005)	(0.0006)	(0.0002)	(0.0004)	(0.0004)
Married	0.0006	-0.0002	0.0001	-0.0007	0.0002	-0.0002	0.0003	-0.0008**
	(0.0005)	(0.0004)	(0.0004)	(0.0005)	(0.0006)	(0.0002)	(0.0005)	(0.0004)
Income	-0.0017***	-0.0000	-0.0010**	0.0017**	-0.0003	-0.0004**	-0.0001	-0.0008**
	(0.0006)	(0.0003)	(0.0003)	(0.0004)	(0.0004)	(0.0002)	(0.0003)	(0.0004)
Urban	-0.0005	-0.0002	-0.0004	0.0013**	-0.0000	-0.0004**	0.0003	-0.0009**
	(0.0004)	(0.0004)	(0.0003)	(0.0006)	(0.0006)	(0.0002)	(0.0005)	(0.0004)
Total population	0.0016**	0.0007	0.0010**	-0.0008	0.0018**	0.0007**	0.0014**	0.0005
	(0.0006)	(0.0005)	(0.0004)	(0.0008)	(0.0007)	(0.0003)	(0.0006)	(0.0007)
Over65	0.0001	0.0004	0.0005	0.0003	-0.0001	0.0004**	0.0002	0.0004
	(0.0005)	(0.0003)	(0.0003)	(0.0005)	(0.0004)	(0.0002)	(0.0003)	(0.0003)
Male-female ratio	-0.0004	0.0002	-0.0001	0.0007	-0.0001	0.0003	0.0005	-0.0004
	(0.0004)	(0.0003)	(0.0003)	(0.0005)	(0.0005)	(0.0002)	(0.0003)	(0.0003)
Minority	-0.0003	0.0009*	0.0006	-0.0006	0.0007	0.0002	0.0007	0.0002
	(0.0006)	(0.0005)	(0.0005)	(0.0007)	(0.0008)	(0.0003)	(0.0006)	(0.0005)
Interest rate	-0.0087***	-0.0070***	-0.0088***	-0.0000	-0.0157***	-0.0015***	-0.0027***	-0.0128***
	(0.0006)	(0.0004)	(0.0004)	(0.0007)	(0.0008)	(0.0003)	(0.0004)	(0.0005)
FICO	-0.0077***	-0.0057***	-0.0063***	-0.0043***	-0.0124***	-0.0023***	-0.0028***	-0.0141***
	(0.0005)	(0.0005)	(0.0004)	(0.0004)	(0.0007)	(0.0003)	(0.0004)	(0.0006)
Balance	0.0059***	0.0044***	0.0052***	0.0019***	0.0061***	0.0016***	0.0047***	0.0032***
	(0.0004)	(0.0003)	(0.0003)	(0.0004)	(0.0004)	(0.0002)	(0.0003)	(0.0004)
LTV	0.0068***	0.0053***	0.0050***	0.0090***	0.0172***	0.0010***	0.0026***	0.0132***
	(0.0005)	(0.0002)	(0.0003)	(0.0004)	(0.0007)	(0.0002)	(0.0002)	(0.0006)
ARM	0.0062***	0.0095***	0.0116***	-0.0153***	0.0191***	0.0009***	0.0093***	0.0025***
	(0.0007)	(0.0007)	(0.0007)	(0.0011)	(0.0012)	(0.0003)	(0.0009)	(0.0005)
Negative amortization	-0.0041**	-0.0043***	-0.0094***	0.0315***	-0.0247***	0.0051***	0.0070***	-0.0087***
	(0.0017)	(0.0014)	(0.0014)	(0.0019)	(0.0026)	(0.0007)	(0.0013)	(0.0013)
Option ARM	-0.0158**	-0.0251***	-0.0253***	-0.0107**	-0.0562***	-0.0061**	-0.0305***	-0.0084
	(0.0067)	(0.0048)	(0.0055)	(0.0053)	(0.0096)	(0.0029)	(0.0052)	(0.0051)
Prepayment penalty	-0.0073***	-0.0086***	-0.0076***	-0.0088***	-0.0102***	-0.0047***	-0.0126***	-0.0110***
	(0.0010)	(0.0009)	(0.0010)	(0.0009)	(0.0016)	(0.0005)	(0.0010)	(0.0010)
Low or no doc.	-0.0094***	-0.0077***	-0.0094***	-0.0089***	-0.0084***	-0.0040***	-0.0065***	-0.0018**
	(0.0007)	(0.0005)	(0.0005)	(0.0009)	(0.0011)	(0.0003)	(0.0006)	(0.0007)
Cash-out	0.0182***	0.0184***	0.0148***	0.0308***			0.0238***	0.0176***
	(0.0008)	(0.0006)	(0.0006)	(0.0010)			(0.0007)	(0.0007)
No-cash-out	0.0152***	0.0159***	0.0134***	0.0264***			0.0166***	0.0148***
	(0.0012)	(0.0007)	(0.0007)	(0.0011)			(0.0007)	(0.0011)
Investment	0.0259***	0.0257***			0.0256***	0.0192***	0.0235***	0.0316***
	(0.0012)	(0.0010)			(0.0012)	(0.0007)	(0.0010)	(0.0008)
Second-home	0.0133***	0.0161***			0.0169***	0.0102***	0.0149***	0.0196***
	(0.0017)	(0.0010)			(0.0013)	(0.0008)	(0.0008)	(0.0020)
Originator FE	Y	Y	Y	Y	Y	Y	Y	Y
State FE	Y	Y	Y	Y	Y	Y	Y	Y
Half-year FE	Y	Y	Y	Y	Y	Y	Y	Y
Observations	591,547	1,778,886	2,080,350	303,675	1,020,629	1,363,053	1,324,342	1,059,862
Adj. R^2	0.328	0.315	0.311	0.418	0.315	0.383	0.298	0.357

The table shows OLS estimates from regressions where the dependent variable takes a value of one if the loan is recognized as second-lien misrepresented and zero otherwise. We look at different subsamples in the following way: lower-class or middle-class (columns (1) and (2)), primary or non-primary (columns (3) and (4)), purchase or refinance (columns (5) and (6)), and high or low credit score (columns (7) and (8)). The high or low credit score subsamples are divided by a FICO score of 670. Variables are defined in Table 1. All continuous variables are winsorized at the 0.5 percent level and then standardized. All specifications include state, half-year, and originator fixed effects. Standard errors clustered at the county level are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

Table 6
Effect of Gambling Preference Associated Second-lien Misrepresentation on Delinquency

	(1)	(2)	(3)	(4)
Misrepresented	0.120***		0.120***	0.121***
	(0.004)		(0.004)	(0.003)
CPRATIO		-0.001	-0.001	-0.002
		(0.004)	(0.004)	(0.004)
Misrepresented*CPRATIO				0.014***
				(0.002)
Correctly reported	0.142***	0.127***	0.142***	0.141***
	(0.007)	(0.007)	(0.007)	(0.007)
REL	-0.009***	-0.009***	-0.009***	-0.009***
	(0.004)	(0.003)	(0.003)	(0.003)
Unemployment	0.006*	0.005*	0.006*	0.006*
	(0.003)	(0.003)	(0.003)	(0.003)
HPA	0.046***	0.047***	0.046***	0.046***
	(0.004)	(0.004)	(0.004)	(0.004)
Education	0.004	0.003	0.004	0.004
	(0.003)	(0.003)	(0.003)	(0.003)
Married	0.013***	0.013***	0.013***	0.013***
	(0.003)	(0.003)	(0.003)	(0.003)
Income	-0.012***	-0.012***	-0.012***	-0.012***
	(0.002)	(0.002)	(0.002)	(0.002)
Urban	0.007*	0.007*	0.007*	0.007*
	(0.004)	(0.004)	(0.004)	(0.004)
Total population	0.000	0.001	0.000	0.000
	(0.006)	(0.006)	(0.006)	(0.006)
Over65	0.003	0.003	0.003	0.003
	(0.003)	(0.003)	(0.003)	(0.003)
Male-female ratio	-0.002	-0.002	-0.002	-0.002
	(0.003)	(0.003)	(0.003)	(0.003)
Minority	0.005	0.005	0.005	0.005
	(0.004)	(0.004)	(0.004)	(0.004)
Interest rate	0.033***	0.032***	0.033***	0.033***
	(0.001)	(0.001)	(0.001)	(0.001)
FICO	-0.094***	-0.093***	-0.094***	-0.094***
	(0.002)	(0.002)	(0.002)	(0.002)
Balance	0.023***	0.021***	0.023***	0.023***
	(0.002)	(0.002)	(0.002)	(0.002)
LTV	0.052***	0.053***	0.052***	0.052***
	(0.002)	(0.002)	(0.002)	(0.002)
ARM	0.040***	0.044***	0.040***	0.040***
	(0.002)	(0.002)	(0.002)	(0.002)
Negative amortization	0.037***	0.029***	0.037***	0.037***
	(0.004)	(0.003)	(0.004)	(0.004)
Option ARM	0.047***	0.044***	0.047***	0.048***
	(0.007)	(0.007)	(0.007)	(0.007)
Prepayment penalty	0.055***	0.058***	0.055***	0.055***
	(0.002)	(0.002)	(0.002)	(0.002)
Low or no doc.	0.054***	0.053***	0.054***	0.054***
	(0.002)	(0.002)	(0.002)	(0.002)
Cash-out	-0.025***	-0.036***	-0.025***	-0.025***
	(0.002)	(0.002)	(0.002)	(0.002)
No-cash-out	0.010***	0.003	0.010***	0.010***
	(0.004)	(0.004)	(0.004)	(0.004)
Investment	0.045***	0.040***	0.045***	0.045***
	(0.004)	(0.004)	(0.004)	(0.004)
Second-home	0.008**	0.003	0.008**	0.008**
	(0.003)	(0.003)	(0.003)	(0.003)
Originator FE	Y	Y	Y	Y
State FE	Y	Y	Y	Y
Half-year FE	Y	Y	Y	Y
Observations	2,383,444	2,383,444	2,383,444	2,383,444
Adj. R^2	0.225	0.221	0.225	0.225

The table shows OLS estimates from regressions where the dependent variable takes a value of one if the loan becomes 90 days or more delinquent using the MBA method in the first three years after origination and zero otherwise. All continuous variables are winsorized at the 0.5 percent level and then standardized. All specifications include state, half-year, and originator fixed effects. Standard errors clustered at the county level are reported in parentheses. $*p < 0.10$, $**p < 0.05$, and $***p < 0.01$.

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

	(1) Lower-class	(2) Middle-class	(3) Primary	(4) Non-primary	(5) Purchase	(6) Refinance	(7) High FICO	(8) Low FICO
Misrepresented	0.134***	0.117***	0.121***	0.112***	0.106***	0.109***	0.094***	0.155***
	(0.004)	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)	(0.004)
CPRATIO	-0.009*	0.001	-0.001	0.000	-0.003	-0.002	-0.001	-0.005
	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)	(0.004)	(0.004)
Misrepresented*CPRATIO	0.016***	0.012***	0.015***	0.001	0.012***	0.017***	0.015***	0.015***
	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)	(0.004)	(0.003)	(0.003)
Correct presented	0.168***	0.134***	0.140***	0.149***	0.127***	0.145***	0.116***	0.192***
	(0.010)	(0.006)	(0.007)	(0.005)	(0.006)	(0.007)	(0.007)	(0.007)
REL	-0.004	-0.011***	-0.009***	-0.011***	-0.009***	-0.008***	-0.009***	-0.007**
	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Unemployment	0.003	0.009***	0.005	0.006**	0.007**	0.006*	0.009***	0.002
	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
HPA	0.045***	0.049***	0.048***	0.029***	0.052***	0.044***	0.044***	0.054***
	(0.004)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)	(0.005)
Education	0.003	0.005**	0.004	0.002	0.005**	0.004	0.003	0.005*
	(0.004)	(0.003)	(0.003)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)
Married	0.013***	0.014***	0.013***	0.011***	0.019***	0.007***	0.015***	0.008***
	(0.004)	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)
Income	-0.005	-0.013***	-0.012***	-0.006***	-0.019***	-0.005**	-0.012***	-0.010***
	(0.004)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.003)
Urban	0.010**	0.006	0.006	0.009**	0.011**	0.005	0.009**	0.005
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Total population	-0.003	0.002	-0.001	0.004	-0.001	-0.000	-0.001	-0.000
	(0.007)	(0.006)	(0.006)	(0.005)	(0.006)	(0.005)	(0.005)	(0.007)
Over65	-0.002	0.004	0.003	0.002	-0.000	0.006**	0.000	0.006**
	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Male-female ratio	-0.004	-0.001	-0.002	-0.001	-0.002	-0.001	-0.002	-0.001
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)
Minority	0.004	0.005	0.005	0.007*	0.017***	-0.002	0.005	0.005
	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.004)
Interest rate	0.029***	0.034***	0.032***	0.046***	0.054***	0.018***	0.040***	0.026***
	(0.002)	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)
FICO	-0.098***	-0.093***	-0.094***	-0.094***	-0.099***	-0.091***	-0.083***	-0.106***
	(0.003)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)	(0.003)
Balance	0.035***	0.019***	0.025***	0.011***	0.021***	0.025***	0.016***	0.036***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)
LTV	0.062***	0.049***	0.052***	0.055***	0.035***	0.062***	0.041***	0.068***
	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
ARM	0.035***	0.040***	0.041***	0.044***	0.052***	0.025***	0.059***	0.004
	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)
Negative amortization	0.031***	0.039***	0.031***	0.093***	0.060***	0.026***	0.044***	0.041***
	(0.006)	(0.004)	(0.004)	(0.006)	(0.006)	(0.004)	(0.005)	(0.005)
Option ARM	0.032***	0.056***	0.046***	0.063***	0.035***	0.055***	0.104***	0.024***
	(0.011)	(0.009)	(0.008)	(0.023)	(0.012)	(0.008)	(0.012)	(0.009)
Prepayment penalty	0.045***	0.059***	0.056***	0.049***	0.077***	0.037***	0.063***	0.036***
	(0.004)	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)
Low or no doc.	0.051***	0.056***	0.051***	0.061***	0.049***	0.059***	0.067***	0.056***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)
Cash-out	-0.045***	-0.018***	-0.033***	0.022***			-0.010***	-0.049***
	(0.003)	(0.002)	(0.002)	(0.003)			(0.003)	(0.002)
No-cash-out	-0.014***	0.016***	0.003	0.056***			0.031***	-0.039***
	(0.005)	(0.004)	(0.004)	(0.003)			(0.002)	(0.004)
Investment	0.051***	0.042***			0.020***	0.064***	0.025***	0.084***
	(0.006)	(0.004)			(0.004)	(0.004)	(0.004)	(0.005)
Second-home	-0.009	0.012***			0.003	0.028***	0.012***	-0.001
	(0.007)	(0.003)			(0.004)	(0.003)	(0.003)	(0.005)
Originator FE	Y	Y	Y	Y	Y	Y	Y	Y
State FE	Y	Y	Y	Y	Y	Y	Y	Y
Half-year FE	Y	Y	Y	Y	Y	Y	Y	Y
Observations	591,142	1,778,053	2,079,194	303,592	1,020,025	1,362,416	1,324,100	1,058,863
Adj. R^2	0.224	0.222	0.226	0.224	0.265	0.204	0.224	0.181

The table shows OLS estimates from regressions where the dependent variable takes a value of one if the loan becomes 90 days or more delinquent using the MBA method in the first three years after origination and zero otherwise. We look at different subsamples in the following way: lower-class or middle-class (columns (1) and (2)), primary or non-primary (columns (3) and (4)), purchase or refinance (columns (5) and (6)), and high or low credit score (columns (7) and (8)). The high or low credit score subsamples are divided by a FICO score of 670. All continuous variables are winsorized at the 0.5 percent level and then standardized. All specifications include state, half-year, and originator fixed effects. Standard errors clustered at the county level are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

Table 8
Discontinuities for Low-documentation Loans Around the Credit Threshold in All Areas

	(1) All loan	(2) Without simul. second	(3) Correctly reported second	(4) Misrepresented second
Panel A. Number of Loans				
Est. 620-	3147	2716	241	190
Est. 620+	7133	5214	1190	729
Est. 620+/620-	2.266	1.920	4.939	3.838
Panel B. Default Rate				
FICO$\geq$620 (β)	0.037	-0.003	0.108	0.014
t-stat	(3.60)	(0.50)	(3.48)	(-1.15)

The table reports the discontinuities for low-documentation loans around a FICO score of 620 in all areas. Panel A presents the estimates from regressions where the dependent variable is the number of loans at each FICO score. Using local linear regressions of the RDD approach, we estimate the number of loans at FICO 620^{-} and FICO 620^{+} and compute the ratio of the estimated number of loans at FICO 620^{+} to 620^{-}. Panel B presents the estimates from regressions where the dependent variable takes a value of one if the loan becomes 90 days or more delinquent using the MBA method in the first three years after origination and zero otherwise. Using local linear regressions of the RDD approach, we estimate the difference in default rates between FICO 620^{-} and FICO 620^{+}. We perform the estimation for all loans, loans without simultaneous seconds, loans with correctly reported simultaneous seconds, and loans with misrepresented simultaneous seconds, and the results are reported in columns (1) to (4), respectively. Bias-corrected t-values (standard errors clustered at the county level) following Calonico et al. (2014) are reported in parentheses.

Table 9
Discontinuities for Low-documentation Loans Around the Credit Threshold in High and Low Gambling Preference Areas

		(1) High	(2) Low	(3) High-Low
Panel A. Number of Loans				
All loan	Est. 620-	1460	1687	
	Est. 620+	3271	3862	
	Est. 620+/620-	2.240	2.289	-0.049
				(-0.52)
Without simul. second	Est. 620-	1275	1442	
	Est. 620+	2502	2712	
	Est. 620+/620-	1.963	1.881	0.081
				(1.39)
Correctly reported second	Est. 620-	94	147	
	Est. 620+	518	673	
	Est. 620+/620-	5.498	4.581	0.917
				(1.23)
Misrepresented second	Est. 620-	91	99	
	Est. 620+	251	477	
	Est. 620+/620-	2.763	4.826	-2.063
				(-2.83)
Panel B. Default Rate				
All loan		0.038	0.037	0.002
		(2.85)	(2.25)	(0.10)
Without simul. second		0.000	-0.006	0.006
		(1.14)	(-0.41)	(0.44)
Correctly reported second		0.061	0.131	-0.070
		(1.46)	(3.65)	(-1.62)
Misrepresented second		0.001	0.052	-0.052
		(-0.07)	(-0.86)	(-0.98)

The table reports the discontinuities for low-documentation loans around a FICO score of 620 separately in high and low gambling preference areas. Panel A presents the estimates from regressions where the dependent variable is the number of loans at each FICO score. Using local linear regressions of the RDD approach, we estimate the number of loans at FICO 620^{-} and FICO 620^{+} and compute the ratio of the estimated number of loans at FICO 620^{+} to 620^{-}. Panel B presents the estimates from regressions where the dependent variable is the default dummy variable. Using local linear regressions of the RDD approach, we estimate the difference in default rates between FICO 620^{-} and FICO 620^{+}. Using the diff-in-disc approach, we estimate the difference between high and low gambling preference areas. We perform the estimation for all loans, loans without simultaneous seconds, loans with correctly reported simultaneous seconds, and loans with misrepresented simultaneous seconds. The results for loans in high, low, and the difference between high and low gambling preference areas are reported in columns (1) to (3), respectively. For discontinuities, bias-corrected t-values (standard errors clustered at the county level) following Calonico et al. (2014) are reported in parentheses. For the difference between high and low gambling preference areas, t-statistics based on heteroskedasticity-consistent standard errors are reported in parentheses.

Table 10
Alternative Gambling Preference Measures

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Gam. Ind. Emp.	0.002***					0.003***	
	(0.001)					(0.001)	
Gam. Ind. Est.		0.005**					0.006**
		(0.002)					(0.002)
Casino Hotel Emp.			-0.001			-0.001	
			(0.001)			(0.001)	
Casino Hotel Est.				0.003*			0.002
				(0.001)			(0.002)
Nonemployer Gam. Est.					0.000	0.001	0.001
					(0.001)	(0.002)	(0.001)
Simul. Second	0.317***	0.317***	0.317***	0.317***	0.317***	0.317***	0.317***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
REL	0.002	0.001	0.000	0.001	0.000	0.002	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Unemployment	-0.001	-0.001	-0.001	-0.001	-0.001	-0.000	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
HPA	0.002	0.002	0.002	0.002	0.002	0.002	0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Education	0.001	-0.000	-0.000	0.001	0.000	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Married	0.001	0.001	0.001	0.001	0.001	0.001	0.002*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Income	-0.001	-0.001	-0.002	-0.002	-0.002	-0.001	-0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Urban	0.001	0.001	0.001	0.000	0.000	0.002	0.000
	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)
Total population	0.000	0.003*	0.002	0.003*	0.002	-0.000	0.004**
	(0.001)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.002)
Over65	-0.001**	-0.002**	-0.002**	-0.002***	-0.002**	-0.002***	-0.002***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Male-female ratio	0.001	0.001	0.001	0.001	0.001	0.000	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Minority	-0.002	-0.001	-0.002	-0.002*	-0.002	-0.002	-0.002
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Interest rate	-0.008***	-0.008***	-0.008***	-0.008***	-0.008***	-0.008***	-0.008***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
FICO	-0.006***	-0.006***	-0.006***	-0.006***	-0.006***	-0.006***	-0.006***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Balance	0.005***	0.005***	0.005***	0.005***	0.005***	0.005***	0.005***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
LTV	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
ARM	0.008***	0.008***	0.008***	0.008***	0.008***	0.008***	0.008***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Negative amortization	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Option ARM	-0.022***	-0.022***	-0.022***	-0.022***	-0.022***	-0.022***	-0.022***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Prepayment penalty	-0.009***	-0.009***	-0.009***	-0.009***	-0.009***	-0.009***	-0.009***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Low or no doc.	-0.008***	-0.008***	-0.008***	-0.008***	-0.008***	-0.008***	-0.008***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Cash-out	0.019***	0.019***	0.019***	0.019***	0.019***	0.019***	0.019***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
No-cash-out	0.016***	0.016***	0.016***	0.016***	0.016***	0.016***	0.016***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Investment	0.027***	0.027***	0.027***	0.027***	0.027***	0.027***	0.027***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Second-home	0.017***	0.017***	0.017***	0.017***	0.017***	0.017***	0.017***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Originator FE	Y	Y	Y	Y	Y	Y	Y
State FE	Y	Y	Y	Y	Y	Y	Y
Half-year FE	Y	Y	Y	Y	Y	Y	Y
Observations	1,050,588	1,050,588	1,050,588	1,050,588	1,050,588	1,050,588	1,050,588
Adj. R^2	0.318	0.318	0.318	0.318	0.318	0.318	0.318

The table shows OLS estimates from regressions where the dependent variable takes a value of one if the loan is recognized as second-lien misrepresented and zero otherwise. Variables are defined in Table 1 while religiosity and the geographic controls are aggregated to CBSA level. All continuous variables are winsorized at the 0.5 percent level and then standardized. All specifications include state, half-year, and originator fixed effects. Standard errors clustered at the CBSA level are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

Table 11
Gambling Preference and Crime Rates

	Violent	Property	Arson	Fraud	Illegal Gambling
CPRATIO	-1.068**	-14.051***	-0.131***	-0.120	0.031**
	(0.450)	(2.324)	(0.029)	(0.469)	(0.014)
REL	-0.376	-8.568***	-0.088***	1.636***	-0.004
	(0.289)	(1.788)	(0.031)	(0.529)	(0.022)
Unemployment	0.447	10.944***	0.157***	-0.526	-0.084**
	(0.390)	(2.184)	(0.039)	(0.575)	(0.034)
HPA	0.667**	3.608*	0.035	0.595	-0.036*
	(0.321)	(1.877)	(0.032)	(0.376)	(0.021)
Education	-2.908***	-15.662***	-0.135***	-1.365***	-0.025
	(0.338)	(1.977)	(0.030)	(0.422)	(0.022)
Married	-4.559***	-35.001***	-0.112***	-0.556	-0.029
	(0.390)	(2.411)	(0.035)	(0.493)	(0.026)
Income	-0.203	-5.345*	-0.139***	-0.338	-0.047
	(0.446)	(2.863)	(0.041)	(0.704)	(0.035)
Urban	3.007***	48.773***	0.277***	0.256	-0.132***
	(0.371)	(2.075)	(0.038)	(0.530)	(0.029)
Total population	5.878***	41.713***	0.375***	-1.107*	0.142***
	(0.544)	(3.152)	(0.047)	(0.615)	(0.032)
Over65	-0.352	-3.069	-0.015	0.245	0.037
	(0.329)	(1.868)	(0.031)	(0.426)	(0.022)
Male-female ratio	-0.903***	-6.253***	-0.113***	-0.338	0.015
	(0.280)	(1.422)	(0.022)	(0.353)	(0.025)
Minority	7.457***	-2.647	0.170***	-0.962	0.060
	(0.474)	(2.627)	(0.043)	(0.707)	(0.041)
State FE	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y
Observations	7,505	7,505	7,505	7,505	7,505
Adj. R^2	0.550	0.532	0.257	0.237	0.376

The table shows OLS estimates from regressions where the dependent variables are crime rate (per 10,000 residents per year) for different types of crimianl activities. All continuous variables, except for the crime rates, are winsorized at the 0.5 percent level and then standardized. All specifications include state and year fixed effects. White robust standard errors are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

Table 12
Alternative Default Measures for Loan Performance Tests

	(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
Misrepresented	0.121***	0.066***	0.123***	0.109***
	(0.003)	(0.003)	(0.003)	(0.003)
CPRATIO	-0.002	-0.002	-0.001	-0.003
	(0.004)	(0.002)	(0.004)	(0.003)
Misrepresented*CPRATIO	0.014***	0.007***	0.014***	0.013***
	(0.002)	(0.002)	(0.003)	(0.003)
Correct presented	0.141***	0.079***	0.144***	0.113***
	(0.007)	(0.005)	(0.007)	(0.006)
REL	-0.009***	-0.005***	-0.009***	-0.007***
	(0.003)	(0.002)	(0.003)	(0.003)
Unemployment	0.006*	0.006***	0.004	0.007**
	(0.003)	(0.002)	(0.003)	(0.003)
HPA	0.046***	0.025***	0.046***	0.037***
	(0.004)	(0.003)	(0.004)	(0.003)
Education	0.004	0.002	0.003	0.003
	(0.003)	(0.001)	(0.003)	(0.002)
Married	0.013***	0.007***	0.013***	0.009***
	(0.003)	(0.001)	(0.003)	(0.002)
Income	-0.012***	-0.008***	-0.012***	-0.009***
	(0.002)	(0.001)	(0.002)	(0.002)
Urban	0.007*	0.004*	0.006	0.006*
	(0.004)	(0.002)	(0.004)	(0.003)
Total population	0.000	0.001	0.000	-0.001
	(0.006)	(0.003)	(0.006)	(0.005)
Over65	0.003	0.003	0.003	0.005
	(0.003)	(0.002)	(0.003)	(0.003)
Male-female ratio	-0.002	-0.001	-0.002	-0.002
	(0.003)	(0.002)	(0.003)	(0.002)
Minority	0.005	0.003	0.006	0.001
	(0.004)	(0.002)	(0.004)	(0.003)
Interest rate	0.033***	0.047***	0.027***	0.031***
	(0.001)	(0.001)	(0.001)	(0.001)
FICO	-0.094***	-0.068***	-0.113***	-0.058***
	(0.002)	(0.001)	(0.002)	(0.001)
Balance	0.023***	0.020***	0.020***	0.019***
	(0.002)	(0.001)	(0.002)	(0.002)
LTV	0.052***	0.027***	0.055***	0.043***
	(0.002)	(0.001)	(0.002)	(0.001)
ARM	0.040***	0.039***	0.033***	0.045***
	(0.002)	(0.002)	(0.002)	(0.002)
Negative amortization	0.037***	0.046***	0.034***	0.023***
	(0.004)	(0.003)	(0.004)	(0.005)
Option ARM	0.048***	0.028***	0.051***	0.036***
	(0.007)	(0.006)	(0.008)	(0.006)
Prepayment penalty	0.055***	0.019***	0.061***	0.025***
	(0.002)	(0.001)	(0.002)	(0.002)
Low or no doc.	0.054***	0.031***	0.058***	0.051***
	(0.002)	(0.001)	(0.002)	(0.002)
Cash-out	-0.025***	-0.025***	-0.020***	-0.026***
	(0.002)	(0.002)	(0.002)	(0.002)
No-cash-out	0.010***	0.001	0.014***	0.009***
	(0.004)	(0.003)	(0.003)	(0.003)
Investment	0.045***	0.028***	0.044***	0.058***
	(0.004)	(0.003)	(0.004)	(0.003)
Second-home	0.008**	0.001	0.009**	0.026***
	(0.003)	(0.002)	(0.004)	(0.003)
Originator FE	Y	Y	Y	Y
State FE	Y	Y	Y	Y
Half-year FE	Y	Y	Y	Y
Observations	2,383,444	2,383,444	2,383,444	2,383,444
Adj. R^2	0.225	0.156	0.236	0.155

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. All continuous variables are winsorized at 0.5 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 - Probit Model

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
CPRATIO	0.006**	0.014***	0.015***	0.012***	0.006***	0.006**	0.004**
	(0.002)	(0.003)	(0.002)	(0.002)	(0.001)	(0.003)	(0.002)
Geographic controls	N	Y	N	Y	Y	Y	Y
Loan chars controls	N	N	Y	Y	Y	Y	Y
Originator FE	N	N	N	N	Y	Y	Y
State FE	N	N	N	N	N	Y	Y
Half-year FE	N	N	N	N	N	N	Y
Observations	638,544	609,627	626,978	598,574	478,730	470,402	459,341
Pseudo R^2	0.000	0.005	0.024	0.026	0.223	0.221	0.258

The table shows the marginal effects of CPRATIO from probit models where the dependent variable takes a value of one if the loan is misrepresented and zero otherwise. All continuous variables, except for the crime rates, are winsorized at the 0.5 percent level and then standardized. All specifications include state, half-year, and originator fixed effects. Standard errors clustered at the county level are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

Table 14
Gambling Preference and Mortgage Misrepresentation - Causal Forest

	(1) Second-lien misrepresentation
CPRATIO	0.0013**
	(0.0006)

The table shows causal forest estimates from regressions where the dependent variable takes a value of one if the loan is misrepresented and zero otherwise. All control variables in Table 4 are used for growing trees and forests. Before growing the trees, all continuous variables are winsorized at the 0.5 percent level and then standardized. A half-year variable is created, starting from the first half of 2005 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 larger clusters receive more weight). 2000 trees are grown in the causal forest. The fraction used for determining splits (honesty) is 50 percent. Standard errors are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

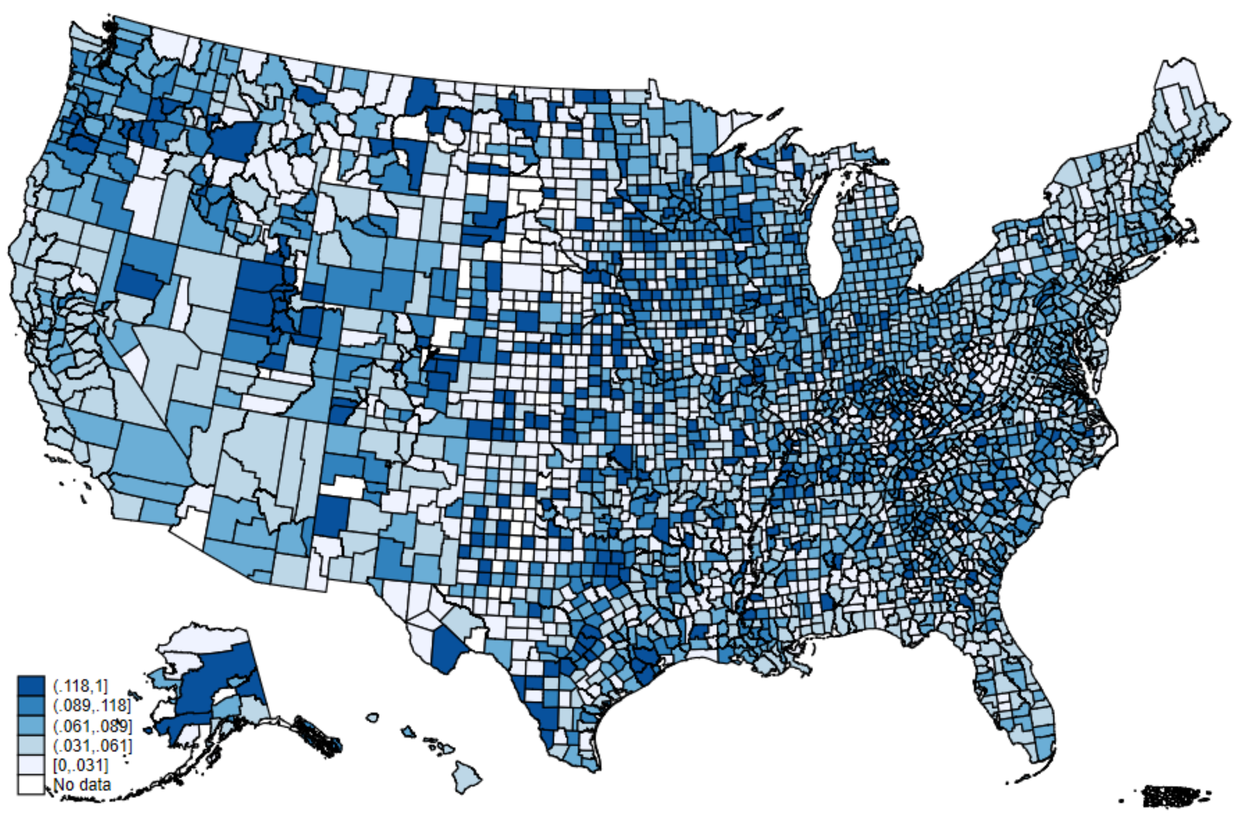

Figure 1
Second-lien misrepresentations distribution
The figure plots the county level proportion of second-lien misrepresentation in all loans.

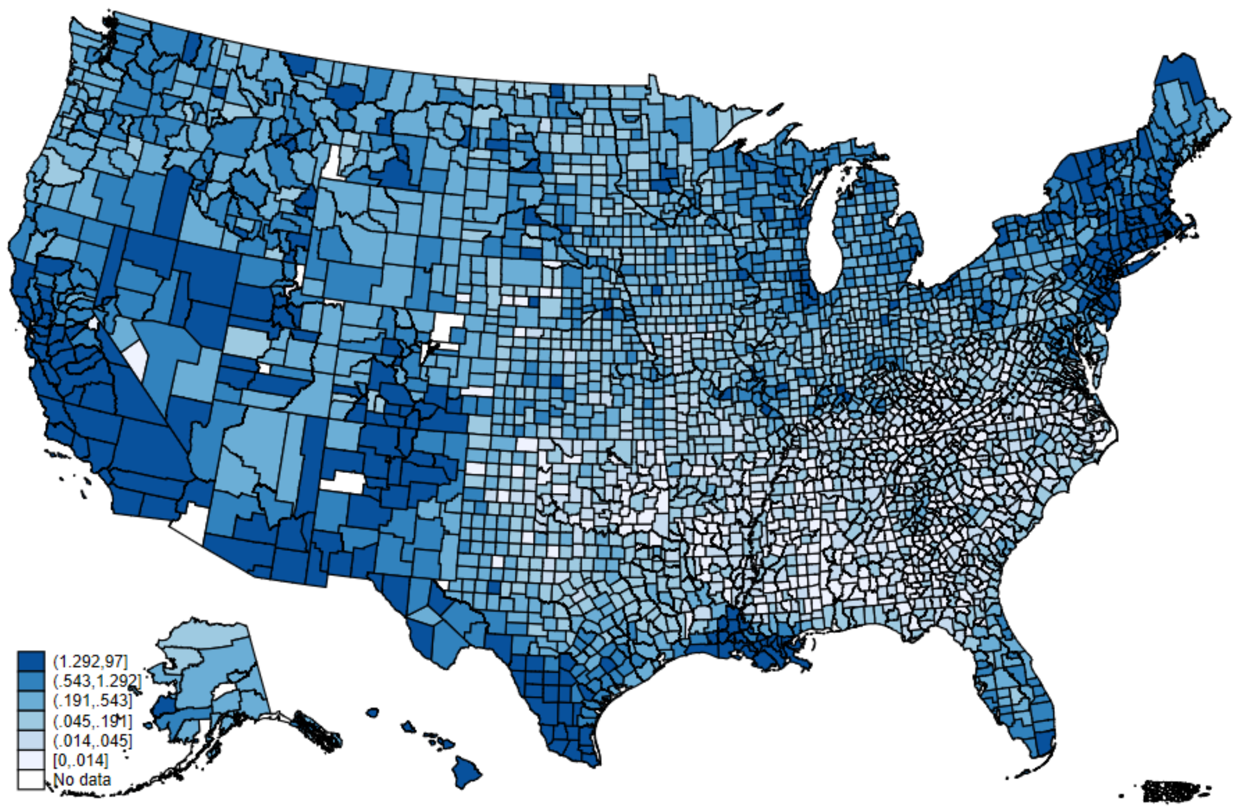

Figure 2
CPRATIO distribution
The figure plots the county level CPRATIO across the US.

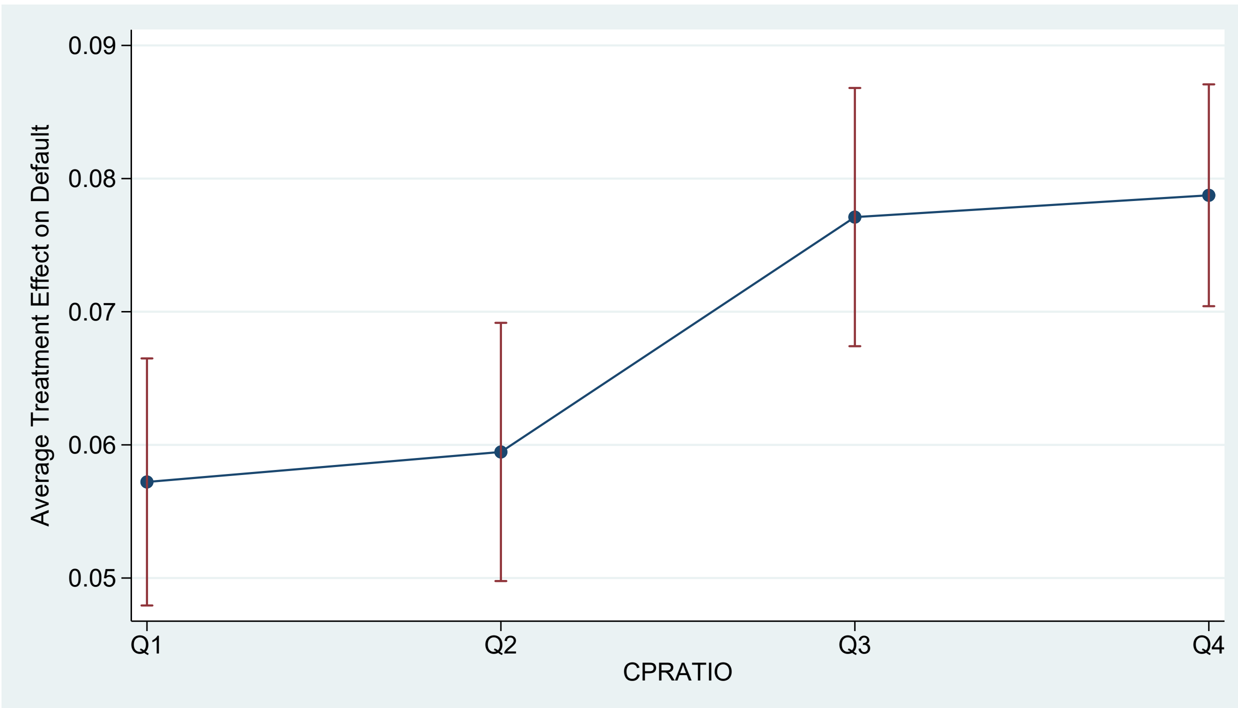

Figure 3
Heterogeneous Treatment Effects of Second-lien Misrepresentation on Default conditioned on Different Levels of Gambling Preference
The figure plots the heterogeneous treatment effects of second-lien misrepresentation on default, conditioned on different levels of gambling preference, using the causal forest approach. The dependent variable is an indicator that takes a value of one if the loan becomes 90 days or more delinquent using the MBA method in the first three years after origination. The average treatment effect of mortgage misrepresentation is estimated at different levels of local gambling preference (four quarters divided by CPRATIO). All control variables in column (3) of Table 6 are used for growing trees and forests. Before growing the trees, all continuous variables are winsorized at the 0.5 percent level and then standardized. A half-year variable is created, starting from the first half of 2005 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 larger clusters receive more weight). 2000 trees are grown in the causal forest. The fraction used for determining splits (honesty) is 50 percent.

Appendix

A Alternative Gambling Preference Measures

In this section, we provide detailed information regarding the use of alternative gambling preference measures.

A.1 Data and Measure Construction

Since CPRATIO may capture broader religious influences, such as attitudes toward forgiveness and dishonesty, we also employ more direct measures of local gambling preference. Specifically, we use per capita employment in the gambling industry and per capita number of gambling establishments to quantify local gambling engagement.

To construct these measures, we extract employment and number of establishments data at the core-based statistical area (CBSA) level[1] for the years 2005 through 2007 from the County Business Patterns (CBP) dataset provided by the U.S. Census Bureau. To identify gambling-specific industries, we retain observations with NAICS code 7132, which includes 71321 for Casinos (except Casino Hotels) and 71329 for Other Gambling Industries. These counts are then normalized by dividing by the total population of the area. Population figures for 2005 to 2007 are linearly interpolated based on CBSA-level data from the 2000 and 2010 Census.

In addition to the two gambling preference measures described above, we construct supplementary proxies. Because CBP classifies businesses based on their primary services, casino hotels (NAICS code 72110) are excluded from the gambling industry category. However, as these businesses also provide gambling services to the public, we extract employment and establishment data for casino hotels separately, normalizing them by total population as well.

Since the CBP dataset includes only employer-based businesses, some small gambling venues without formal employees may be omitted. To account for these, we incorporate data from the Nonemployer Statistics (NES) datasets, also maintained by the U.S. Census Bureau. We retain NES observations with NAICS code 7132[2]. The number of nonemployer gambling establishments is likewise normalized by CBSA population.

[1] This refers to the MSA File in the CBP dataset; however, the MSA identification codes correspond to CBSA codes.

[2] During our sample period, no casino hotel reports having employees; therefore, NAICS code 72110 does not appear in the NES data.

While the CBP dataset records the number of establishments without restrictions, the employment data suffers from a key limitation: many cells are suppressed to preserve confidentiality (Eckert et al., 2021)[3]. To ensure sufficient CBSA-level coverage, we apply the following method to approximate employment figures[4]. First, for each CBSA in which a gambling-related NAICS code appears, if the employment value is nonzero, we retain it as-is. Second, for cases where the employment value is reported as zero, we estimate it by multiplying the number of establishments in each employment-size interval by the median value of that interval. For instance, in CBSA 11460, there are two establishments, one in the 1–4 interval and one in the 5–9 interval, so the estimated employment is calculated as $1 \times 2.5 + 1 \times 7 = 9.5$. For establishments falling within the highest interval (i.e., open-ended, 5,000 or more), we conservatively assign the lower bound value of 5,000. Third, we compare the estimated employment figure with the lower bound of the corresponding employment level flag and retain the value that is greater[5].

The NES dataset also has similar restrictions: many cells are suppressed to preserve confidentiality or aviod quality issue. In other words, the number of establishment is recorded as zero but the establishment flag shows that the zero is not due to nonexistence but due to not willing to disclosure individual business or due to unreliable recording. To solve this issue, we replace zeros with flag D (disclosure concern) with 2[6] and replace zeros with flag S (reliability concern) with state average.

The NES dataset also contains limitations: numerous cells are suppressed to preserve confidentiality or to avoid quality issues. In such cases, the number of establishments is recorded as zero; however, the establishment flags indicate that the zero does not reflect true nonexistence. Instead, these suppressed values result from either unwillingness to disclose individual business information or concerns regarding data reliability. To address this issue, we implement the following imputation strategy. For observations with flag `D` (confidentiality concerns), we replace the zero with a value of 2[7]. For observations with flag `S` (reliability

[3] Another limitation discussed in Eckert et al. (2021) is the periodic reclassification of industries. However, since changes to gambling-related NAICS codes are infrequent and our sample period is short, this concern is minimal for our analysis.

[4] Eckert et al. (2021) proposes a solution for county-level data, but not for CBSA-level data. We adapt their core approach in a simplified form due to constraints that are difficult to satisfy at the CBSA level. Our primary strategy involves using the median value within employment intervals and verifying that the estimated employment remains within the bounds of the corresponding employment level flag interval.

[5] Refer to Tables 1 and 3 in Eckert et al. (2021) for definitions of employment intervals and employment level flag intervals, respectively. In our sample, all estimated values fall below the corresponding upper bounds of the employment level flags.

[6] We use 2 because the smallest nonzero value is 3. In untabulated results, using 1 doesn't change our conclusion.

[7] We use 2 since the smallest observable nonzero value is 3. In untabulated results, substituting with 1 does not materially affect our findings.

concerns), we substitute the zero with the state-level average for the corresponding industry category.

For geographic control variables, the U.S. Census Bureau does not provide CBSA-level data for the year 2000. To address this limitation, we aggregate county-level data up to the CBSA level. For ratio variables, we first aggregate the numerator and denominator separately across constituent counties, and then compute the ratio based on the aggregated values. For unemployment rates obtained from the U.S. Bureau of Labor Statistics, we apply the same aggregation procedure to the annual average data and implement a one-year lag in our regressions. For HPA, we use the CBSA-level HPI data directly from the FHFA.

Table A1 presents the summary statistics for all constructed gambling preference measures.

A.2 Additional Tests

To validate our alternative measures and explore their relation with CPRATIO, we do the following tests.

First, to examine the relationship between CPRATIO and both the alternative gambling preference measures and three additional control variables, we conduct a univariate sorting test. CBSAs are sorted into quintiles according to their CPRATIO values, and we calculate equal-weighted quintile averages for the two alternative measures of gambling preference as well as the three control variables. Table A2 summarizes the results. For the two alternative measures, we find that gambling industry employment tends to increase with CPRATIO, and the difference between the highest and lowest quintiles is statistically significant. This suggests that employment in the gambling industry captures a construct that is consistent with CPRATIO. In contrast, the relationship between gambling industry establishments and CPRATIO is hump-shaped, with the High quintile displaying slightly elevated values. This pattern may reflect legal constraints. For example, both a lottery venue and a casino venue are counted as a single establishment, even though casinos typically offer a broader array of gambling services. Moreover, lotteries are generally more easily approved due to their association with public funding for education and healthcare, whereas casinos face stricter legal barriers. As a result, in regions with low gambling preference where lotteries are permitted but casinos are prohibited, the number of recorded establishments may increase. Conversely, in areas where casinos are allowed, fewer establishments may exist, but each offers stronger access to gambling activities.

For the three control variables, employment and establishment counts in casino hotels do not show a substantial increase with CPRATIO. This outcome may stem from two factors. First, the presence of casino hotels is influenced not only by local gambling demand but also by tourist-oriented gambling activity. As a result, local gambling preference does not necessarily correlate with casino hotel employment or establishment density. Second, exceptions in gambling legality, such as tribal casinos, permit casino hotel operations in certain regions even when local gambling preference is low. In contrast, the number of nonemployer gambling establishments tends to increase with CPRATIO. This pattern suggests that such establishments complement minor local gambling demand and provide an additional proxy for community-level gambling activity.

Second, to further examine whether CPRATIO can explain variation in the alternative gambling preference measures and the three additional control variables, we estimate a series of OLS regressions. Each dependent variable is regressed on CPRATIO and CBSA-level geographic control variables with state and year fixed effects. All continuous variables are winsorized at the 0.5 percent level and standardized prior to analysis. Table A3 presents the regression results. For the two alternative measures, gambling industry employment and establishment, CPRATIO is positively and significantly associated with both outcomes, indicating that higher CPRATIO values correspond to greater levels of these supply-side indicators. This supports the use of CPRATIO as a proxy for local gambling preference. In contrast, results for casino hotels are mixed. While the per capita number of establishments is positively correlated with CPRATIO, the per capita employment in the sector is negatively correlated. This divergence reinforces our decision to treat casino hotel variables as controls for tourist-oriented gambling demand rather than proxies for local gambling preference. Finally, for nonemployer gambling establishments, which serve as a complementary measure for minor local gambling activity, we also observe a significantly positive coefficient on CPRATIO. This suggests that higher CPRATIO values are associated with increased presence of nonemployer gambling businesses, further corroborating CPRATIO's relevance as an indicator of local gambling preference.

A.3 Additional Illustration

In this section, we address several concerns related to the use of alternative measures of local gambling preference.

First, several limitations prevent us from using the alternative measures as the basis for our primary analysis. Although employment and establishment data in the gambling industry may more directly reflect local gambling preferences, these indicators represent supply-side dynamics rather than demand-side behavior. Furthermore, they are influenced by factors such as local legality and economic conditions, which introduces additional noise into the measurement. Another key limitation is the restricted coverage. These alternative measures span only 460 CBSAs, encompassing approximately 1,187 counties, and cover around 1 million loans. In contrast, CPRATIO includes data for nearly 3,000 counties and covers almost the entire sample of 3 million loans. This broader coverage is especially critical when analyzing lender-specific effects and borrower preference heterogeneity.

Second, we rely on CBSA-level data rather than county-level data to construct our gambling preference measures. The rationale for using a broader geographic and economic unit is that residents in most U.S. metropolitan areas can easily commute within the city. Consequently, individuals residing in counties with few gambling establishments may work or gamble in adjacent counties with greater access. As a result, low employment or establishment counts in a given county do not necessarily reflect low gambling preference among its residents. By contrast, CBSAs are sufficiently large that cross-CBSA commuting for work or gambling is relatively rare. Instances of inter-CBSA gambling, such as through tourism, are mostly captured by the casino hotel measures. Therefore, the CBSA level offers a more robust and stable proxy for local gambling preferences than county-level data.

Third, we restrict our sample to CBSAs where establishments exist in all three categories: gambling industry, casino hotels, and nonemployer gambling businesses. This selection improves the clarity and reliability of our gambling preference measures. In such areas, false zeros, resulting from legal prohibitions, data suppression, or missing supply, are minimized. In untabulated regressions, when we replace missing values with zeros and include all CBSAs, the coefficient on the gambling preference measure remains positive but becomes only marginally significant. Furthermore, when we split the sample into high and low per capita establishment groups, the coefficient is strongly significant in the high-establishment group but insignificant in the low group. This pattern suggests that second-lien misrepresentation does rise with elevated local gambling preference, whereas low values in the measure may reflect extraneous noise rather than genuinely low gambling tendencies.

Table A1
Descriptive Statistics for Alternative Gambling Preference Measures

	N	Mean	SD	P10	P25	P50	P75	P90
Gambling Industries (NAICS 7132, CBSA-year level)								
Employment (per 10,000 residents)	1201	19.60	42.74	0.22	0.69	2.53	16.41	59.50
Establishment (per 10,000 residents)	1201	0.39	0.85	0.03	0.06	0.14	0.29	0.76
Casino Hotel (NAICS 72110, CBSA-year level)								
Employment (per 10,000 residents)	258	118.49	264.87	0.07	0.50	16.90	77.41	389.73
Establishment (per 10,000 residents)	258	0.24	0.60	0.00	0.01	0.07	0.21	0.55
Nonemployer Gambling (NAICS 7132, CBSA-year level)								
Establishment (per 10,000 residents)	1239	0.38	0.50	0.00	0.14	0.27	0.45	0.76

The table presents descriptive statistics for the alternative local gambling preference measures and its related controls, covering the sample period from 2005 to 2007. We report the number of observations, mean, standard deviation, 25th percentile, 50th percentile, and 75th percentile.

Table A2
Univariate Sorting

	Gambling Industries		Casino Hotel		Nonemployer Gambling
Quintile	Employment	Establishment	Employment	Establishment	Establishment
Low	16.26	0.26	114.20	0.18	0.28
Q2	11.77	0.47	140.31	0.79	0.35
Q3	18.67	0.57	49.83	0.11	0.43
Q4	26.47	0.34	95.67	0.21	0.40
High	24.92	0.31	219.91	0.25	0.46
High-Low	8.66	0.05	105.72	0.08	0.18
t-Statistic	2.02	0.78	1.72	1.10	4.01

The table presents univariate sorting results for five gambling preference related measures: per capita employment in the gambling industry (Gambling Industries Employment), per capita number of gambling establishments (Gambling Industries Establishment), per capita employment in the casino hotel industry (Casino Hotel Employment), per capita number of casino hotel establishments (Casino Hotel Establishment), and per capita number of nonemployer gambling establishments (Nonemployer Gambling Establishment). CBSAs are sorted into quintiles based on CPRATIO, and for each quintile, the equal-weighted average of the five gambling preference related measures is computed. The table reports White robust t-statistics for the difference in mean values between the High and Low CPRATIO quintiles.

Table A3
Alternative measures estimate

	Gambling Industries		Casino Hotel		Nonemployer Gambling
	Employment	Establishment	Employment	Establishment	Establishment
CPRATIO	0.112***	0.031*	-0.194***	0.194*	0.134***
	(0.043)	(0.019)	(0.068)	(0.113)	(0.028)
REL	-0.116**	-0.053	-0.084	-0.374*	-0.179***
	(0.059)	(0.042)	(0.147)	(0.206)	(0.061)
Unemployment	-0.013	-0.080***	-0.137	-0.100	0.027
	(0.061)	(0.022)	(0.090)	(0.072)	(0.072)
HPA	0.008	-0.039	0.108	-0.030	0.075
	(0.045)	(0.026)	(0.083)	(0.076)	(0.046)
Education	0.018	-0.073**	-0.703***	0.272	-0.084
	(0.058)	(0.033)	(0.214)	(0.198)	(0.063)
Married	0.059	-0.001	-0.480***	0.252	-0.017
	(0.047)	(0.031)	(0.150)	(0.153)	(0.048)
Income	-0.206***	-0.054*	0.950***	-0.292	0.162**
	(0.069)	(0.031)	(0.250)	(0.253)	(0.065)
Urban	0.035	0.070**	-0.397***	0.037	-0.008
	(0.063)	(0.032)	(0.102)	(0.129)	(0.044)
Total population	-0.138***	-0.132***	-0.023	-0.294***	-0.063
	(0.050)	(0.025)	(0.119)	(0.102)	(0.051)
Over65	0.113***	0.025	0.068	-0.107	0.038
	(0.043)	(0.027)	(0.110)	(0.110)	(0.044)
Male-female ratio	0.074	0.011	0.027	-0.081	-0.067**
	(0.050)	(0.017)	(0.080)	(0.074)	(0.026)
Minority	0.152**	0.008	0.169	0.173**	-0.032
	(0.063)	(0.020)	(0.110)	(0.087)	(0.042)
State FE	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y
Observations	1,118	1,118	240	240	1,147
Adj. R2	0.307	0.723	0.541	0.408	0.262

The table shows OLS estimates from regressions where the dependent variables are the alternative gambling preference measures and related control variables. Religiosity and the geographic controls are aggregated to CBSA level. All continuous variables are winsorized at the 0.5 percent level and then standardized. All specifications include state and year fixed effects. White robust standard errors are reported in parentheses. $*p < 0.10$, $**p < 0.05$, and $***p < 0.01$.

B Local Tolerance for Law-breaking Behaviors

In this section, we provide detailed information on the crime rate data.

We collect county-level crime data for the years 2005 to 2007 from the National Neighborhood Data Archive (NaNDA) and the Uniform Crime Reporting (UCR) Program Data available on the ICPSR website. From NaNDA, we extract county population (used to calculate county-level crime rates), total murders, total rapes, total robberies, total aggravated assaults, total burglaries, total larcenies, total auto thefts, and total arsons. Following the definitions in the same dataset, we construct violent crimes as the sum of murders, rapes, robberies, and aggravated assaults, and property crimes as the sum of burglaries, larcenies, and auto thefts. From the UCR Program Data, we obtain counts for fraud and gambling—total. For each crime category, we calculate the crime rate by dividing the number of incidents by the county population (as defined for county-level crime reports) and multiplying by 10,000, resulting in crime rates expressed as the number of crimes per 10,000 residents per year. Table B1 presents the summary statistics for all crime rates.

Table B1
Descriptive Statistics for Crime Rates

	N	Mean	SD	P10	P25	P50	P75	P90
Crime Rate (per 10,000 residents per year)								
Violent	8689	28.44	26.42	5.31	11.07	21.50	37.91	60.28
Property	8689	238.41	148.33	73.49	135.34	213.69	314.34	437.87
Arson	8689	1.69	2.17	0.00	0.00	1.19	2.42	3.97
Fraud	8689	18.25	92.24	0.00	1.92	5.91	17.32	39.40
Illegal Gamble	8689	0.48	4.31	0.00	0.00	0.00	0.00	0.27

The table presents descriptive statistics for crime rates, covering the sample period from 2005 to 2007. We report the number of observations, mean, standard deviation, 25th percentile, 50th percentile, and 75th percentile.

C Diff-in-disc concerns

To validate our application of the diff-in-disc approach in comparing the number of loans and default rates between high and low gambling preference areas, we address the following concerns in this section.

C.1 Assumptions for Applying Diff-in-disc

According to Grembi et al. (2016), the validity of the diff-in-disc approach depends on three assumptions. First, all potential outcomes must be continuous around the cutoff. In our setting, this means no manipulation of FICO scores around the cutoff point of 620. Since the data on the distribution of FICO scores in the U.S. population is not available to us, we argue this assumption by referencing the findings in Keys et al. (2010). They found that the distribution of FICO scores across the population is smooth using data from an anonymous credit bureau. Additionally, by exploring the reversal of anti-predatory lending laws in Georgia and New Jersey, they found that borrowers were either unaware of the differential screening around the threshold or unable to quickly manipulate their FICO scores[8].

Second, in the absence of treatment, the effect of the confounding policy around the cutoff is constant over the "diff" part. To fulfill this assumption, we check whether the pattern holds for full documentation loans around FICO 620, which have no treatment but similar confounding factors. Table C1 shows the results for the number of loans (panel A) and default rate (panel B) estimated using full-documentation loans. For full-documentation loans, the number of loans shows a minimal jump, and the difference between high and low gambling preference areas is slight. This evidence indicates that without the ease of securitization, having a FICO score below or above 620 would not cause the number of loans to increase differently between high and low gambling preference areas. Additionally, there is no jump in the default rate for all types of loans, and the difference in the increase of the default rate between high and low gambling preference areas is minor. These findings illustrate that without the ease of securitization, lenders do not have lax screening standards for any type of loan and do not screen differently between high and low gambling preference areas.

Third, the effect of the treatment around the cutoff does not depend on the confounding policy. This assumption states that there should be no interaction between the treatment and the confounding policy. In other words, the pre-determined outcomes and covariates should have similar jumps (or no jumps) between high and low gambling preference areas. We test this assumption by estimating equation 7 with pre-determined outcomes and covariates as the

[8] According to the rating agency (Fair Isaac), strategic manipulation of FICO scores is difficult.

dependent variable and also add other control variables in the regression. We choose the same bandwidth (10) and kernel (uniform) as the main tests. The results are reported in Table C2. Column (1) presents the results of loan characteristics, showing that most characteristics have similar jumps between high and low gambling preference areas. The proportions of negative amortization and second-home loans are significant, but we argue that this does not affect our conclusion for several reasons. First, although the absolute jump magnitude is different, the ratio of 620^+ to 620^- is close, i.e., negative amortization jumps from 6.74/2.97 percent to 15.82/6.61 percent (2.35/2.23 times) in high/low gambling preference areas. Second, the proportion of these specific loans is too low to have a meaningful effect on the number of loans with misrepresentation and the default rate (i.e., in this local range, second homes make up only 1.2 percent of all loans and only 0.6 percent of misrepresented loans). Column (2) reports the results of borrower characteristics, showing that most characteristics have similar jumps between high and low gambling preference areas. The ratios of older people, minorities, and the male-female ratio are significant, but their influence is minor because the magnitude is too small, i.e., the average proportion of older people, the male-female ratio, and minorities in the local range are 12.2 percent, 28.89 percent, and 0.96 percent, while the differences are only 0.319 percent, -0.534 percent, and 0.003 percent, respectively.

C.2 Sensitivity to Regression Choices

Beyond these three assumptions, we also check whether the findings are sensitive to the inclusion of control variables, the choice of bandwidth, and the choice of estimation kernel. We estimate models with and without covariates, choosing bandwidths of 8, 10, and 12, and using uniform or triangular kernels. Table C3 reports the ratio of estimated values at FICO 620^+ to 620^- for the number of loans[9]. The magnitude of jumps is similar to the baseline results: the number of different types of loans increases significantly due to the ease of securitization, and the jump for loans with misrepresented seconds is smaller in high gambling preference areas than in low gambling preference areas. These results confirm the robustness of the findings that lenders do not play a larger role in increasing second-lien misrepresentation in high gambling preference areas. Table C4 shows that the default rate of loans with misrepresented seconds is small and insignificant in all cases. Moreover, although not significant, the jump in high gambling preference areas is smaller than the jump in low gambling preference areas in most cases. These robustness tests show that lenders in high gambling preference areas are not more inclined to increase such misrepresentation than lenders in low gambling preference areas.

[9]Since the number of loans is counted at the level of gambling preference areas while control variables are at the level of counties, control variables could not be included in the tests.

C.3 Placebo Test

Lastly, we conduct a placebo test to evaluate the possibility that our results are driven by chance. Specifically, following Goodman and Mayer (2018), we implement the diff-in-disc estimations at false FICO score thresholds below and above 620. Our inferences on the lender's role come from two findings from the diff-in-disc estimations: (1) the number of loans with misrepresented seconds increases much less in high gambling preference areas at FICO 620, and (2) the default rate of loans with misrepresented seconds either shows a small and insignificant jump in high gambling preference areas or a jump that is smaller in high gambling preference areas than in low gambling preference areas. Thus, at these false thresholds, we expect to find (1) no systematic difference in the increase of the number of loans with misrepresented seconds between high and low gambling preference areas as in our baseline results, and (2) no situation where the default rate of loans with misrepresented seconds shows a meaningful upward jump in high gambling preference areas and a jump that is larger in high gambling preference areas than in low gambling preference areas. To maintain a similar setting and avoid potential jumps in the misrepresentation rate found in the literature, we conduct the tests within the range of 601 to 639[10]. For the number of loans, we exclude the range of 615 to 625 to stay sufficiently away from the true threshold that could cause noise in jump estimation. Figure C1 plots the cumulative density function of the 28 placebo point estimates for the number of loans with misrepresented seconds. All of the placebo coefficients are greater than our estimated coefficient (-2.063) and smaller than the absolute value of our estimated coefficient with a small magnitude. We also perform a t-test on the placebo coefficients to check the null hypothesis that the true mean equals 0, and the p-value of this test is 0.819, which fails to reject the null hypothesis. These placebo tests support the idea that the difference in discontinuities for the number of loans with misrepresented seconds between high and low gambling preference areas is not due to chance. For the default rate, we only exclude the true threshold since no jump exists at the true threshold, so estimations at the scores near it would not be affected. Figure C2 plots the cumulative density function of the placebo point estimates for the default rate of loans with misrepresented seconds. Most placebo coefficients are smaller than the absolute value of our estimated coefficient (0.052). The larger ones are also insignificant, and they show an insignificant small jump in high gambling preference areas. The t-test on the placebo coefficients that checks whether the true mean equals 0 provides a p-value of 0.638. These placebo tests support the idea that lenders did not have laxer screening in high gambling preference areas.

[10]From Figure 3 in Griffin and Maturana (2016) about the misrepresentation rate, we can see clear jumps at FICO 600 and 640.

Table C1
Discontinuities of Full-documentation Loans Around the Credit Threshold

		(1) High	(2) Low	(3) All	(4) High-Low
Panel A. Number of loans					
All loan	Est. 620-	1565	2801	4367	
	Est. 620+	1948	3232	5180	
	Est. 620+/620-	1.244	1.154	1.186	0.090
					(2.27)
Without simul. second	Est. 620-	1169	1943	3112	
	Est. 620+	1461	2225	3686	
	Est. 620+/620-	1.249	1.146	1.185	0.104
					(2.19)
Correctly reported second	Est. 620-	267	585	852	
	Est. 620+	316	674	989	
	Est. 620+/620-	1.183	1.152	1.161	0.031
					(0.58)
Misrepresented second	Est. 620-	129	274	403	
	Est. 620+	171	333	505	
	Est. 620+/620-	1.325	1.216	1.251	0.109
					(1.03)
Panel B. Default rate					
All loan		-0.003	0.009	0.006	-0.012
		(0.42)	(0.03)	(0.34)	(-0.89)
Without simul. second		0.001	0.004	0.004	-0.003
		(1.10)	(-0.00)	(0.73)	(-0.23)
Correctly reported second		-0.008	0.014	0.007	-0.022
		(-0.22)	(0.06)	(-0.15)	(-0.61)
Misrepresented second		-0.020	0.029	0.018	-0.049
		(-0.48)	(0.03)	(-0.02)	(-1.15)

The table presents the results of the estimations in Table 8 and Table 9 using full-documentation loans. Panel A reports the estimates for the number of loans at each FICO score. Panel B reports the estimates for loans that become 90 days or more delinquent using the MBA method in the first three years after origination. Using local linear regressions of the RDD approach, we estimate the number of loans and the default rate at FICO 620^- and FICO 620^+ and compute the ratio of the estimated number of loans at FICO 620^+ to 620^-. Using the diff-in-disc approach, we estimate the difference between high and low gambling preference areas. We perform the estimation for all loans, loans without simultaneous seconds, loans with correctly reported simultaneous seconds, and loans with misrepresented simultaneous seconds. The results for loans in high gambling preference areas, loans in low gambling preference areas, loans in all areas, and the difference between high and low gambling preference areas are reported in columns (1) to (4), respectively. For discontinuities, bias-corrected t-values (standard errors clustered at the county level) following Calonico et al. (2014) are reported in parentheses. For the difference in the ratio of the number of loans/default rate between high and low gambling preference areas, t-statistics based on heteroskedasticity-consistent standard errors/standard errors clustered at the county level are reported in parentheses.

Table C2
Estimates of the discontinuities in observable covariates

	(1) loan characteristics		(2) borrower characteristics
Interest rate (%)	0.001	REL	-0.005
	(0.041)		(0.004)
Balance	0.017	Unemployment (%)	0.056
	(0.013)		(0.071)
LTV (%)	0.128	HPA	-0.004
	(0.367)		(0.004)
ARM	-0.011	Education (%)	-0.276
	(0.013)		(0.242)
Negative amortization	0.025***	Married (%)	-0.052
	(0.007)		(0.107)
Option ARM	0.001	Income	0.005
	(0.002)		(0.007)
Prepayment penalty	0.004	Urban (%)	0.539
	(0.015)		(0.335)
Cash-out	0.002	Total population (ln)	0.031
	(0.013)		(0.025)
No-cash-out	0.002	Over65 (%)	0.319***
	(0.007)		(0.105)
Investment	0.007	Male-female ratio	0.003**
	(0.007)		(0.001)
Second-home	0.007**	Minority (%)	-0.534*
	(0.003)		(0.314)

The table reports the results of diff-in-disc regression for pre-determined outcomes and covariates using Equation 7 with other controls. The dependent variables are the pre-determined outcomes and covariates. Column (1) shows the tests for loan characteristics, and column (2) shows the tests for borrower characteristics. As in the main test, the bandwidth is set to 10 and the kernel is uniform. Standard errors clustered at the county level are reported in parentheses.

Table C3
Sensitivity of Results for Number of Loans

	Uniform				Triangular			
	High	Low	All	H-L	High	Low	All	H-L
Bandwidth = 8								
All loan	2.230	2.242	2.237	-0.013	2.250	2.204	2.225	0.046
				(-0.13)				(0.44)
Without simul. second	1.950	1.811	1.875	0.140	1.970	1.745	1.846	0.225
				(2.45)				(4.10)
Correctly reported second	5.469	4.693	4.999	0.776	5.761	5.235	5.446	0.526
				(0.96)				(0.56)
Misrepresented second	2.709	5.095	3.957	-2.385	2.496	5.128	3.826	-2.632
				(-2.85)				(-3.21)
Bandwidth = 10								
All loan	2.240	2.289	2.266	-0.049	2.240	2.234	2.237	0.006
				(-0.52)				(0.06)
Without simul. second	1.963	1.881	1.920	0.081	1.961	1.788	1.867	0.172
				(1.39)				(3.26)
Correctly reported second	5.498	4.581	4.939	0.917	5.616	4.971	5.228	0.645
				(1.23)				(0.74)
Misrepresented second	2.763	4.826	3.838	-2.063	2.609	5.133	3.900	-2.524
				(-2.83)				(-3.20)
Bandwidth = 12								
All loan	2.199	2.245	2.224	-0.047	2.237	2.262	2.250	-0.025
				(-0.50)				(-0.27)
Without simul. second	1.940	1.859	1.897	0.081	1.960	1.831	1.891	0.129
				(1.52)				(2.44)
Correctly reported second	5.209	4.602	4.846	0.607	5.520	4.845	5.114	0.676
				(0.91)				(0.84)
Misrepresented second	2.656	4.388	3.590	-1.732	2.668	4.971	3.858	-2.302
				(-2.60)				(-3.20)
Bandwidth = 14								
All loan	2.197	2.199	2.198	-0.002	2.222	2.257	2.241	-0.036
				(-0.02)				(-0.39)
Without simul. second	1.946	1.842	1.891	0.104	1.952	1.842	1.893	0.110
				(2.02)				(2.12)
Correctly reported second	5.278	4.666	4.915	0.612	5.424	4.788	5.043	0.636
				(0.95)				(0.84)
Misrepresented second	2.515	3.940	3.275	-1.426	2.645	4.755	3.747	-2.110
				(-1.99)				(-3.18)

The table reports the sensitivity of RDD and diff-in-disc regression for the number of loans to the choice of bandwidth and estimation kernel. Bias-corrected t-values (standard errors clustered at the county level) are reported in parentheses. $*p < 0.10$, $**p < 0.05$, and $***p < 0.01$.

Table C4
Sensitivity of Results for Default Rate of Loans with Misrepresented Second

	Uniform				Triangular			
	High	Low	All	H-L	High	Low	All	H-L
Bandwidth = 8								
No control	0.005	0.019	-0.004	-0.014	-0.007	-0.016	-0.031	0.009
	(-0.24)	(-0.83)	(-1.18)	(-0.23)	(0.81)	(-1.43)	(-0.89)	(0.13)
With control	0.026	0.048	0.033	-0.022	-0.001	-0.012	-0.007	0.010
	(-0.39)	(-1.27)	(-1.27)	(-0.43)	(0.45)	(-1.73)	(-0.97)	(0.16)
Bandwidth = 10								
No control	0.001	0.052	0.014	-0.052	-0.004	0.011	-0.014	-0.015
	(-0.07)	(-0.86)	(-1.15)	(-0.98)	(0.32)	(-1.15)	(-1.07)	(-0.25)
With control	0.013	0.078	0.045	-0.065	0.011	0.025	0.016	-0.014
	(0.16)	(-0.92)	(-0.73)	(-1.41)	(0.06)	(-1.43)	(-1.06)	(-0.26)
Bandwidth = 12								
No control	-0.022	0.065	0.014	-0.087	-0.003	0.033	0.000	-0.036
	(0.30)	(-0.30)	(-0.50)	(-1.71)	(0.09)	(-1.00)	(-1.14)	(-0.65)
With control	-0.004	0.082	0.042	-0.085	0.011	0.050	0.030	-0.038
	(0.33)	(0.07)	(0.15)	(-1.95)	(0.12)	(-1.12)	(-0.82)	(-0.80)
Bandwidth = 14								
No control	0.007	0.071	0.033	-0.064	-0.007	0.046	0.007	-0.053
	(-0.47)	(0.11)	(-0.72)	(-1.39)	(0.06)	(-0.59)	(-0.88)	(-1.03)
With control	0.017	0.081	0.051	-0.064	0.008	0.060	0.035	-0.052
	(-0.29)	(0.79)	(0.26)	(-1.53)	(0.13)	(-0.43)	(-0.35)	(-1.18)

The table reports the sensitivity of RDD and diff-in-disc regression for the default rate of loans with misrepresented seconds to the inclusion of control variables, the choice of bandwidth, and the choice of estimation kernel. Bias-corrected t-values (standard errors clustered at the county level) following Calonico et al. (2014) are reported in parentheses.

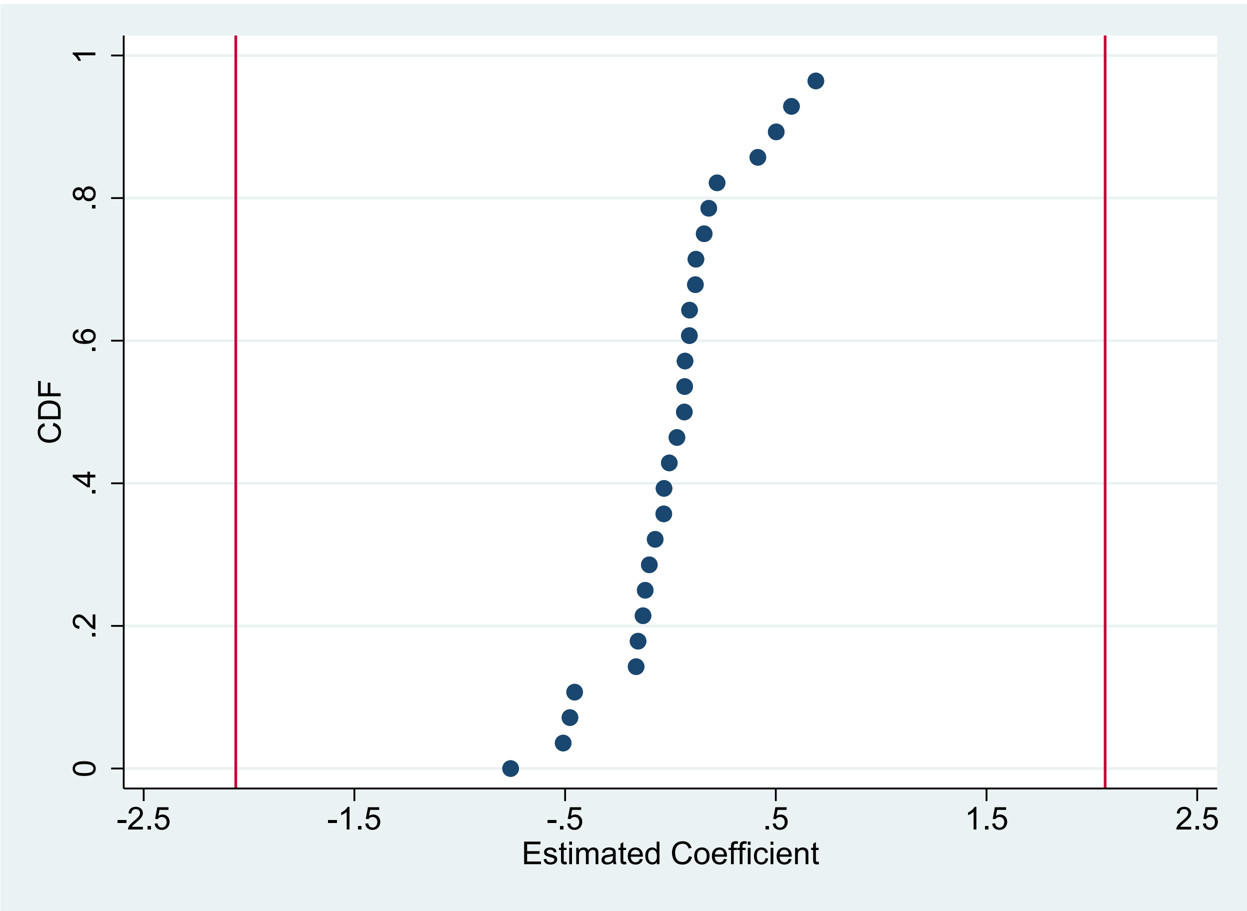

Figure C1
Placebo Tests for Number of Loans with Misrepresented Second
The figure plots the empirical c.d.f. of the estimated coefficient for the number of loans with misrepresented seconds from a set of diff-in-disc estimations at false FICO score thresholds below and above 620 (i.e., any score from 601 to 614 and any score from 626 to 639). The vertical lines show the benchmark estimate (-2.063) and its positive value (2.063).

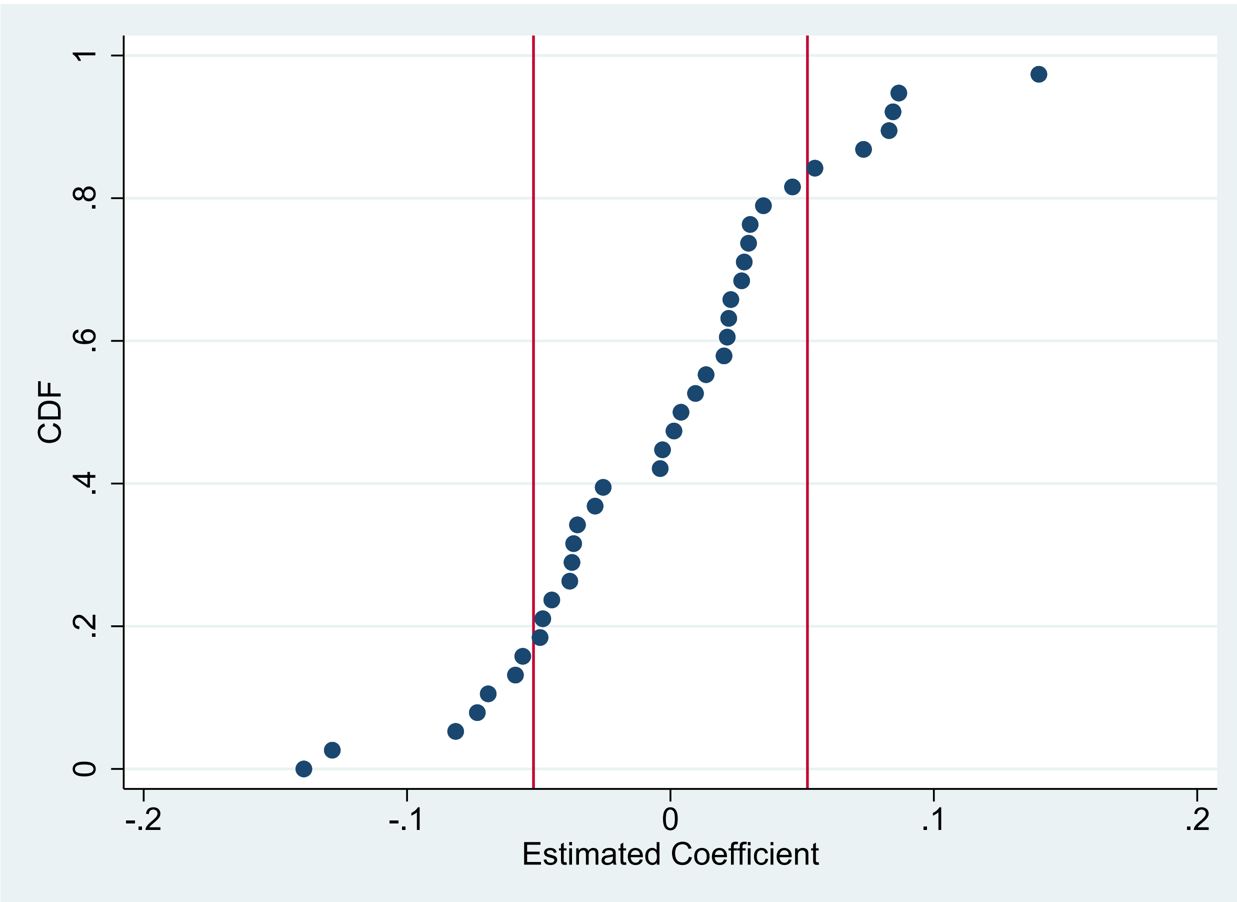

Figure C2
Placebo Tests for Default Rate of Loans with Misrepresented Second
The figure plots the empirical c.d.f. of the estimated coefficient for the default rate of loans with misrepresented seconds from a set of diff-in-disc estimations at false FICO score thresholds below and above 620. The vertical lines show the benchmark estimate (-0.052) and its positive value (0.052).