Alternative Facts in Peer-to-Peer Loans? Borrower Misreporting Dynamics and Implications*

Vesa Pursiainen†

August 2019

Abstract

I study the implications and determinants of borrower misreporting in peer-to-peer (P2P) loans for credit card debt repayment and consolidation. I identify potential misreporting based on three behavior-based indicators: consistency of loan amount with outstanding credit balance, roundness of reported income, and roundness of chosen loan amount. A Misreporting Index constructed from these indicators has significant predictive power over the likelihood of default, and the additional default risk does not get compensated in the form of higher interest. I find evidence consistent with both intentional as well as innocent misreporting. Supporting the former, misreporting is more prevalent in areas with lower social capital, implying weaker social norms, and among borrowers whose professions are considered less honest, while borrowers with higher genuine income uncertainty are also more prone to misreport. Misreporting increases notably from Q2 2017 onwards, and the increase is larger among low-income and low-credit-grade borrowers.

JEL classification: D12, D91, G23, G41

Keywords: P2P loan, marketplace lending, misreporting, default, round number heuristic

*I thank Nuruddin Ahmed, Vimal Balasubramaniam, Li-Ting Chiu, Christopher Hansman, Ivika Jäger, Florian Kiesel, Michael King, Geng Li, Tse-Chun Lin, Christoph Merkle, Mario Milone, Kalle Rinne, Armin Schwienbacher, Katrin Tinn, Hans-Joachim Voth, Brian Wolfe, the conference participants at Toronto FinTech Conference 2019, European Retail Investment Conference (ERIC) 2019, EEA-ESEM 2019, Basel Workshop of Credit Risk 2019, and Private Capital Research Conference 2019, as well as the seminar participants at Imperial College London for helpful comments. I am grateful to Foundation for the Advancement of Finnish Securities Markets (Suomen Arvopaperimarkkinoiden Edistämissäätiö) for financial support. All errors are my own.

†Imperial College London and the University of Hong Kong. E-mail: vesa@hku.hk

1 Introduction

The market for peer-to-peer (P2P) loans has grown rapidly in recent years with the U.S. P2P consumer lending volume estimated at $21 billion in 2016.[1] Lending Club, the largest U.S. P2P loan platform, now accounts for 10% of outstanding U.S. personal loans.[2] P2P loan platforms enable investors to lend directly to borrowers and use technology to automate credit analysis and hence provide low-cost, speedy loan decisions.[3] This business model has obvious advantages. However, it also means that P2P lending relies heavily on self-reported data, for example on borrower income and the use of funds. Borrower misreporting thus represents a challenge for credit analysis.

As the growth of the sector in size and importance has coincided with signs of deteriorating credit quality of P2P loans, regulators have become increasingly concerned about the potential scale of defaults and their effects on consumers. The U.S. Treasury issued a white paper in 2016 (U.S. Department of the Treasury, 2016), highlighting concerns that the credit scoring models used in the sector have been developed during benign times and remain untested over the full credit cycle. Also in 2016, the Consumer Financial Protection Bureau (CFPB) began accepting complaints from consumers encountering problems with loans from online P2P lenders.[4] The Federal Deposit Insurance Corporation (FDIC) issued a note cautioning banks to consider the risks of engaging with online lending platforms (Federal Deposit Insurance Corporation, 2015). Such regulatory concerns, together with some notable governance issues, have also been reflected in investor worries about the sustainability of the P2P lenders' business models and the negative share price development of Lending Club over the last years.[5]

The potential for misreporting is not unique to P2P loans. Jiang, Nelson, and Vytlacil (2014) find that a significant part of the delinquency difference between full- and low-documentation mortgages is due to income falsification. Griffin and Maturana (2016) find evidence of widespread misrepresentation in the securitized loans market with around 48% of loans exhibiting at least one indicator of misrepresentation. The business model of P2P

[1] Estimate based on Ziegler, Reedy, Le, Zhang, Kroszner, and Garvey (2017).

[2] Estimate based on Lending Club Investor Day 2017 presentation.

[3] Throughout this paper, I use the term *peer-to-peer*, or *P2P*, lending. Some firms and analysts prefer to use the term *marketplace lending*, reflecting the fact that most of the funds come from institutional investors, not "peers".

[4] https://www.consumerfinance.gov/about-us/newsroom/cfpb-now-accepting-complaints-on-consumer-loans-from-online-marketplace-lender/

[5] In 2016, Lending Club CEO Renaud Laplanche resigned following an internal probe that found that the firm had misrepresented the characteristics of $22 million of loans sold to a single investor, in order to fit the investor's criteria. It was also found that a fund affiliated with Mr. Laplanche had bought 115 million worth of Lending Club loans.

loan marketplaces, with virtually no human interaction with borrowers, might make them less able to incorporate soft information and thus more vulnerable to misreporting than traditional lenders using loan officers. This idea is consistent with the findings of Berger, Miller, Petersen, Rajan, and Stein (2005), who show that smaller banks are better at incorporating soft information than large banks.

In this paper, I use a comprehensive sample of loans for credit card debt repayment and consolidation from Lending Club to study the implications and determinants of misreporting by borrowers. I construct a *Misreporting Index* for P2P loans based on three indicators that may signal untruthful or inaccurate information provided by the borrower: the consistency of loan amount with outstanding credit balance, the roundness of reported income, and the roundness of the chosen loan amount. These indicators combine behavioral cues and simple, observable indicators of potential misreporting into a single index variable. They may also capture both intentional as well as innocent elements of misreporting, so in the context of this study, misreporting should be understood broadly to mean any inaccurate information provided by the borrower.

The first indicator, consistency of loan amount with outstanding credit balance, is an attempt to capture observable discrepancies between the stated purpose of the loan and the requested loan amount. For the vast majority of Lending Club loans, the stated purpose of the loan is either credit card repayment or debt consolidation.[6] If the borrower is truthful in stating the purpose of the loan, the requested loan amount should be close to the current revolving credit balance of the borrower. That is, the amount borrowed should match the refinancing need. I calculate the ratio of loan amount to outstanding revolving credit balance and use that to identify suspect loan amounts. While this ratio has not been used in the literature, there is evidence of misreporting the intended use of proceeds in other contexts. For example, Karlan and Zinman (2012) find that poor entrepreneurs tend to under-report the non-enterprise uses of loan proceeds.

The second indicator, roundness of reported income, relies on behavioral cues that may be associated with misreporting. The general human tendency to use round numbers is well documented in the literature (e.g Baird, Lewis, and Romer, 1970; Dehaene and Mehler, 1992; Jansen and Pollmann, 2001). Experiments on numerosity show that when people are asked to estimate a value, they tend to provide round numbers (Kaufman, Lord, Reese, and Volkmann, 1949; Krueger, 1982; Lipton and Spelke, 2005). Furthermore, such a tendency to provide round numbers is more pronounced when lacking information or general knowledge (Ormerod and Ritchie, 2007; Kleven and Waseem, 2013; Whynes, Frew, Philips, Covey, and Smith, 2007). Garmaise (2015) shows that residential mortgage borrowers have a tendency

[6] These two categories account for 79% of all loans on Lending Club.

to misreport financial assets just above round number thresholds. Such misreporting is associated with significantly higher delinquency rates. Eid, Maltby, and Talavera (2016) were the first to report the positive relationship between the roundness of reported income and the likelihood of default in P2P loans.

The third indicator, roundness of the loan amount, is conceptually a mix between the first two indicators. If the stated purpose of the loan, credit card repayment or debt consolidation, is truthful, there is no reason why the required loan amounts should cluster at round numbers. Nevertheless, I find that loan amounts of Lending Club loans exhibit very clear clustering at round number thresholds. The distribution of outstanding loan balances shows no such clustering. This suggests that the uses of loan proceeds may not fully match the stated purpose. A round loan amount may thus signal a larger discrepancy between the funding need and the loan amount. This is consistent with the results of Lin and Pursiainen (2018b), who find that reward-based crowdfunding campaign goal amounts set by entrepreneurs on Kickstarter cluster strongly at round numbers, and that such use of round goal amounts is associated with significantly lower likelihood of funding success.

Based on these three indicators of potential misreporting, I construct a loan-level Misreporting Index. I define the index as the sum of five dummies indicating whether i) the loan amount/revolving balance ratio is below 0.9 or above 1.1, ii) the reported income is divisible by 5,000, iii) the reported income is divisible by 10,000, iv) the loan amount is divisible by 5,000, and v) the loan amount is divisible by 10,000. The maximum index value is thus five, and the minimum zero. I hypothesize that loans with a higher Misreporting Index value are associated with adverse outcomes for the investors.

To test this hypothesis, I perform a regression analysis of loan default rates, controlling for loan credit grades assigned by the Lending Club platform as well as for a large number of other observable loan and borrower characteristics. Consistent with the hypothesis, I find that loans with higher Misreporting Index values are associated with a significantly higher likelihood of default than other loans. Furthermore, the default likelihood increases monotonically with the index value. The economic magnitude of the effect is large. A loan with a Misreporting Index value of five (the highest value) is 5.0 %-points more likely to default than an otherwise similar loan with an index value of zero (the lowest value). This represents a 25% increase relative to the unconditional mean default rate in the sample.

Furthermore, loans with a high Misreporting Index value do not compensate investors for the higher default risk in the form of higher interest. Based on my regression analysis, there is no significant difference in the pricing of loans with high and low Misreporting Index values, although the estimated coefficients are negative, the opposite of what one might expect if the loans were correctly priced. I also find that loans with a higher Misreporting

Index value incur higher losses in cases where they default.

It is important to note that misreporting as measured in this paper can have multiple causes. It is possible that the borrower deliberately attempts to mislead lenders in order to get better loan terms. It is also possible that the index captures uncertainty or cognitive limitations of the borrower, both of which have been linked with the use of round numbers in the literature. To explore these different possibilities, I also analyze the determinants and dynamics of misreporting, as proxied by the Misreporting Index. I use two indicators that are likely to signal lower propensity to deliberately misreport: the strength of social norms, as proxied by regional social capital, and borrower honesty, based on the perceived trustworthiness of the borrower's stated profession. In addition, I explore the implications of genuine income uncertainty.

I find that the propensity of borrowers to misreport is lower in areas with higher social capital, suggesting that social norms may limit willingness to misreport.[7] This is consistent with a large literature suggesting that regional social capital can help generate trust and trustworthy behavior by enforcing social norms. For example, Guiso, Sapienza, and Zingales (2013) show the importance of social norms in decisions on whether to strategically default on mortgages. Guiso, Sapienza, and Zingales (2004) show that social capital is important for the spread of financial contracts and hence for the development of financial markets. Hasan, Hoi, Wu, and Zhang (2016) find that high social capital is associated with lower bank loan spreads, looser other loan terms, and lower at-issue bond spreads. Similarly, I use reported employment titles and Gallup survey data of the perceived honesty of different professions to assign an honesty score to a large subsample of borrowers. The survey is done on an annual basis and asks a large and representative sample of Americans to rate how honest they think people are in different professions. I find that misreporting is significantly more prevalent among borrowers with professions considered less honest.

To obtain an indicator of genuine uncertainty of expected income, I construct an occupation-based future income volatility measure using data from the Panel Survey of Income Dynamics (PSID). I match the expected income volatility variable to P2P borrowers using reported employment titles. I find that expected income volatility is associated with significantly higher levels of misreporting. This is also consistent with my robustness checks, where I find that round income is a significant predictor of default likelihood even when the income has been verified by the platform. Taken together, these findings suggest that the Misreporting Index captures both a deliberate as well as an inadvertent component of misreporting due to genuine uncertainty by the borrower.

[7] I use the county-level social capital index of Lin and Pursiainen (2018a) to measure social capital. I calculate population-weighted averages of social capital for each 3-digit zip code.

I also explore the time trends in misreporting and find that the propensity to misreport increases markedly following Q2 2017. This trend is clearly visible from an estimation of quarterly fixed effects from a regression analysis with Misreporting Index as the dependent variable. A closer investigation shows that the main driver of this increase in Misreporting Index is the roundness of loan amount. There are several things that might plausibly increase the aggregate propensity of borrowers to deliberately misreport, but also possibly incentivize the platform to approve more potentially dubious loans.

The shift coincides with the Trump presidency and a period of historically unparalleled media focus on "fake news" and lying by politicians. It seems possible that this dynamic could have caused some loosening of moral norms and hence an increase in misreporting, as captured by the index. On the other hand, it is equally possible that there was no fundamental change in borrower behavior, and that the shift resulted from changes made by the platform. In an e-mail exchange with Lending Club Investor Relations, I received a reply saying they have not made changes that should result in changes in loan amounts. It is impossible for me to verify this statement, but in case truthful, it would suggest that the shift is driven by borrower behavior. Without making any claims on causality, I perform a regression analysis using the Trump inauguration as a dividing point in time and I find that misreporting increases significantly following the inauguration. This increase in misreporting is significantly larger among low-income and low-credit-grade borrowers.

In my additional analysis, I show that each of the three indicators used for constructing the index separately has significant predictive power over the probability of default. Furthermore, the estimated effect of each of the components remains significant when including all of them separately in the same regression. This suggests that each captures separate aspects of misreporting. I also analyse the robustness of the *Suspect amount* variable by dividing it into *Suspect amount above*, a dummy taking the value one if the loan amount / revolving balance ratio is above 1.1, and *Suspect amount below*, taking the value one if the loan amount / revolving balance ratio is below 0.9. Interestingly, the predictive power of both over the default likelihood is of similar magnitude. This finding is noteworthy as I cannot observe the initial loan application amounts of the borrowers, and hence it is possible that a borrower might have been refused a larger loan, resulting in this ratio to be below 0.9.

I contribute to at least three strands of literature. First, I add to the literature on borrower misreporting by introducing two new easily identifiable indicators of potential misreporting: the consistency of loan amount with the existing credit balance to be refinanced, and the roundness of the loan amount. I combine these with income rounding to provide a more comprehensive and less noisy proxy for potential misreporting. Second, I contribute to the extensive literature on the use of the round number heuristic by individuals and the

adverse outcomes it entails. Third, I contribute to the literature on the determinants of default in P2P loans by providing behavioral indicators that can be calculated from the data that are readily available to all investors, but may not be considered as "obvious" inputs by many credit analysts.

It should also be noted that I cannot know what variables or what methodology Lending Club itself uses for generating credit grades for the loans on the platform. Therefore, I do not know to what extent Lending Club is aware of the findings that I report. However, I show that loan pricing does not appear to fully capture the increased risk related to loans with higher Misreporting Index values. Therefore, for an investor in this asset class, my findings provide valuable insights. They should also be of interest for anyone making credit decisions, whether they are based on algorithms or other decision-making methods.

2 Literature review and hypothesis development

2.1 Misreporting in loans

People are often not entirely truthful in reporting their finances. Zinman (2009) compares estimates of credit card use based on household data from the Survey of Consumer Finances (SCF) and industry data and finds that the SCF yields much lower estimates of revolving debt. Karlan and Zinman (2008) compare survey self-reports with administrative data and find that nearly 50% of recent borrowers do not report their high-interest consumer loans.

Misreporting also represents a fundamental problem for lenders. Jiang et al. (2014) show that a significant part of the delinquency difference between full- and low-documentation mortgages is due to income falsification. Griffin and Maturana (2016) find evidence of widespread misrepresentation in the securitized loans market with around 48% of loans exhibiting at least one indicator of misrepresentation. They also find that such misrepresentation is associated with a 51% higher likelihood of delinquency. Karlan and Zinman (2012) document poor entrepreneurs' tendency to under-report the non-enterprise uses of loan proceeds. Of course, firms may also engage in misreporting. Chava, Huang, and Johnson (2018) show evidence of reputational costs for firms for misreporting.

2.2 Round numbers as a behavioral cue

A substantial amount of literature suggests a general tendency to use round numbers (e.g Baird et al., 1970; Dehaene and Mehler, 1992; Jansen and Pollmann, 2001). They are the most cognitively accessible numbers (Schindler and Kirby, 1997) and act as reference points (Rosch, 1975). Experiments on numerosity show that when people are asked to estimate a

value, they tend to provide round numbers (Kaufman et al., 1949; Krueger, 1982; Lipton and Spelke, 2005). Furthermore, such a tendency to provide round numbers is more pronounced when lacking information or general knowledge (Ormerod and Ritchie, 2007; Kleven and Waseem, 2013; Whynes et al., 2007).

Financial decisions represent good examples of situations where assessing the exact value of the traded asset is a cognitively difficult task and hence a likely area to find evidence of the reliance on such round number heuristics. Correspondingly, the empirical literature has reported the clustering of prices at round numbers in a number of financial markets. Osborne (1962) is perhaps the first study documenting stock price clustering at round numbers, using closing prices from the New York Stock Exchange (NYSE). Niederhoffer (1965) shows similar clustering at limit orders. The same patterns are reported in a number of other equity markets (e.g., Harris, 1991; Aitken, Brown, Buckland, Izan, and Walter, 1996; Grossman, Miller, Cone, Fischel, and Ross, 1997; Hameed and Terry, 1998; Bhattacharya, Holden, and Jacobsen, 2012).

Ball, Torous, and Tschoegl (1985) show that prices at the London gold market cluster at round numbers, and the degree of clustering depends on the amount of information in the market. There are also a plethora of studies documenting price clustering in the foreign exchange markets (Goodhart and Curcio, 1991; DeGrauwe and Decupere, 1992; Grossman et al., 1997; Sopranzetti and Datar, 2002; Mitchell and Izan, 2006), stock index values (Donaldson and Kim, 1993; Koedijk and Stork, 1994; Ley and Varian, 1994), index futures and options markets (ap Gwilym, Clare, and Thomas, 1998), orders submitted by investors in Israeli IPOs (Kandel, Sarig, and Wohl, 2001), and real estate prices (Palmon, Smith, and Sopranzetti, 2004). In addition, Kahn, Pennacchi, and Sopranzetti (1999) show that bank deposit rates cluster at round numbers.[8]

2.3 Misreporting and P2P loan performance

There is a small but growing literature on peer lending and the performance of P2P loans. Serrano-Cinca, Gutirrez-Nieto, and Lpez-Palacios (2015) analyze the determinants of defaults on Lending Club loans and conclude that while grade assigned by the platform is the most important predictive factor of default, the accuracy of the model is improved by adding other information, especially the borrower's debt level. Emekter, Tu, Jirasakuldech, and Lu (2015) provide similar results.

A handful of papers focus on behavioral factors in P2P lending. Duarte, Siegel, and

[8]In a related strand of literature, Lacetera, Pope, and Sydnor (2012) find that left-digit bias leads to discontinuous drops in used-car sale prices at 10,000-mile odometer thresholds, along with smaller drops at 1,000-mile thresholds.

Young (2012) find that borrowers who appear more trustworthy have higher probabilities of having their loans funded, while Freedman and Jin (2014) find that borrowers with social ties are consistently more likely to have their loans funded and receive lower interest rates. Similarly, Lin, Prabhala, and Viswanathan (2013) find that online friendships of borrowers act as signals of credit quality. Friendships increase the probability of successful funding, lower interest rates on funded loans, and are associated with lower ex post default rates. Chen, Huang, and Yec (2018) find evidence that the amount of punctuation used in loan descriptions inuences the funding probability, borrowing rate, and default.

In the context of traditional bank loans, Garmaise (2015) shows that residential mortgage borrowers have a tendency to misreport financial assets just above round number thresholds. Such misreporting is associated with significantly higher delinquency rates.

In this paper, I combine three indicators of potential misreporting into a single Misreporting Index, partly relying on the existing literature on the connection between behavioral cues and borrower misreporting. Intuitively, and consistent with prior findings in the literature, I expect a negative relationship between misreporting and loan performance.

I thus propose my main hypothesis:

Hypothesis 1: *Loans with a higher Misreporting Index value, based on the consistency of loan amount with outstanding credit balance, roundness of reported income, and roundness of chosen loan amount, are associated with adverse outcomes for the investors.*

3 Data and methodology

3.1 P2P loan data

I use a comprehensive sample of loans from Lending Club, issued between 2007 and Q2 2018. Given one of my indicators of misreporting, consistency of loan amount with the outstanding credit balance, is only applicable for credit card and debt consolidation loans, I limit my analysis to those loans. Figure 1 shows the distribution of all Lending Club loans by stated purpose. Credit card and debt consolidation represent 79% of loans on the platform. These loans provide me with a sample of 1,584,127 loans. For the analysis of historical default rates and corresponding interest rates, I exclude current loans. resulting in a sample of 905,316 loans. Appendix B provides a summary of the sample filtering. Table 1 shows the number of campaigns by year in our sample, classified by loan status.

3.2 Misreporting Index

To proxy the likelihood of misreported information, I construct a loan-level *Misreporting Index*, defined as the sum of five dummy variables that indicate the consistency of loan amount with outstanding credit balance, the roundness of reported income, and the roundness of the chosen loan amount.

The first indicator variable, *Suspect amount*, is a dummy taking the value one if the loan amount / revolving credit ratio is below 0.9 or above 1.1. This variable is an attempt to capture observable discrepancies between the stated purpose of the loan and the requested loan amount. As mentioned above, the stated purpose of most loans on Lending Club loans is either credit card repayment or debt consolidation. If the borrower is truthful in stating the purpose of the loan, the requested loan amount should be close to the current revolving credit balance of the borrower. That is, the amount borrowed should match the refinancing need. I calculate the ratio of loan amount to outstanding revolving credit balance and use that to identify suspect loan amounts. Figure 2 shows the distribution of loans based on this ratio. As expected, the distribution for credit card and debt consolidation loans clearly centers at one, while this is not the case for other loan types.

Figure 3.A plots the average default rate against this ratio. It clearly suggests that loans with loan/revolving credit balance ratio of one or near one have substantially lower default rates than loans either above or below one. To check whether this pattern is captured the credit grades and interest rate by the platform, I regress the *Default* dummy on a set of dummies for each sub-credit grade (35 different grades) and the interest rate of the loan. Figure 3.B plots the residuals from this regression against the loan amount / revolving credit ratio. We see that the same pattern remains. I hence use this apparent discontinuity and define a variable *Suspect amount* as a dummy taking the value one if this ratio is below 0.9 or above 1.1.

As can be seen from Figure 4, there is no such discontinuity in loans taken for other reported purposes. This acts as a sanity check for the methodology, as there is no reason why this ratio should be consequential for other loan types.

The second indicator, roundness of reported income, is suggested as a proxy for dishonest income reporting by prior literature. Garmaise (2015) shows that residential mortgage borrowers have a tendency to misreport financial assets just above round number thresholds. Such misreporting is associated with significantly higher delinquency rates. Eid et al. (2016) were the first to report the positive relationship between the roundness of reported income and the likelihood of default in P2P loans. Figure 5 shows the distribution of reported annual incomes. This distribution exhibits very clear clustering at round numbers. I define two dummy variables, *Income divisible 5,000* and *Income divisible 10,000*, indicating different

levels of roundness.

The third indicator, roundness of the loan amount, is conceptually related to both of the first two indicators. If the stated purpose of the loan, credit card repayment or debt consolidation, is truthful, there is no reason why the required loan amounts should cluster at round numbers. However, Figure 6 shows that similar to reported incomes, loan amounts exhibit very clear clustering at round number thresholds. The distribution of outstanding loan balances, shown in Figure 7, shows no such clustering. This suggests that the uses of loan proceeds may not fully match the stated purpose. A round loan amount may thus signal a larger discrepancy between the funding need and the loan amount. This is consistent with the results of Lin and Pursiainen (2018b), who find that reward-based crowdfunding campaign goal amounts set by entrepreneurs on Kickstarter cluster strongly at round numbers, and that such use of round goal amounts is associated with significantly lower likelihood of funding success. Similar to income roundness, I define two dummy variables, *Loan amount divisible 5,000* and *Loan amount divisible 10,000*, indicating different levels of roundness.

I define the *Misreporting Index* as the sum of *Suspect Amount*, a dummy taking the value one if the loan / revolving balance ratio is outside the range 0.9-1.1, *Income divisible 5,000* and *Income divisible 10,000*, dummies taking the value one if the reported annual income is divisible by 5,000 and 10,000, respectively, and *Amount divisible 5,000* and *Amount divisible 10,000*, each taking the value one if the loan amount is divisible by 5,000 and 10,000, respectively. The index values thus range from zero (lowest likelihood of misreporting) to five (highest likelihood of misreporting).

3.3 Descriptive statistics

Table 2 shows summary statistics for the sample. The average annual interest rate is 13.0% and loan amount around $15,600. 30% of the loans have a term of five years, the remainder three years. The average loan / revolving balance ratio is 3.89, the median 1.06. 41% of the full sample are current loans performing normally. When excluding current loans, around 60% of historical loans have been prepaid, 19% repaid at maturity, and 20% defaulted.

The average Misreporting Index value is 1.47. Nearly 84% of loans are classified as having a suspect loan amount, based on the loan / revolving balance ratio. 29% of borrowers have a reported income divisible by 10,000 and 33% a loan amount divisible by 5,000.

On borrower demographics, the average reported income is 76,700 and the median 65,000. The average employment length is 5.6 years. 10% of borrowers own their home without mortgage, while another 49% have a mortgaged home, and the remainder rent their homes.

4 Main results

4.1 Likelihood of default

My main hypothesis predicts that loans with a higher Misreporting Index value are associated with a higher likelihood of default. I test this hypothesis by performing regressions of the following form:

$$Default_i = \alpha_0 + \alpha_1 \times Misreporting\ Index_i + \alpha_2 \times X_i + \epsilon_i \tag{1}$$

where $Default_i$ is a dummy taking the value one if the loan i defaulted, $Misreporting\ Index_i$ is a proxy for potential misreporting, described above, and X_i is a vector of control variables, including loan amount, annual income, loan amount / revolving credit ratio, debt/income ratio, employment length, outstanding revolving credit balance, revolving credit utilization rate, number of open accounts, number of total accounts on file, dummies indicating home ownership and mortgage, sub-grade-year joint fixed effects based on the credit grade assigned by Lending Club (35 grades) and allowing time variation as the Lending Club credit rating methodology may change over time, purpose fixed effects based on the stated purpose of the loan (either credit card or debt debt consolidation), 3-digit zip code fixed effects (955 zip codes) based on the address of the borrower, Year-month x Term fixed effects based on the issue date of the loan (133 months x two alternative terms). For this analysis, I exclude current loans, where the loan outcome cannot yet be known.

The results, shown in Table 3, provide support for my hypothesis. As shown in Panel A, loans with a higher Misreporting Index value are associated with significantly higher likelihood of default across all model specifications. Panel B shows the same regression, but instead of including the Misreporting Index as a continuous variable, I include dummies for each of the possible index values. The omitted value is zero, so the estimated coefficients are relative to loans with an index value of zero. Based on the estimated coefficients in column 5, loans with a Misreporting Index value of three (highest) are associated a 4.2 %-points higher likelihood of default than otherwise similar loans with a Misreporting Index value of zero. This represents a 21% increase to the sample average default rate.

4.2 Loan pricing and loss given default

A higher likelihood of default does not mean that a loan represents a worse investment. In theory, such increased default risk should be reflected in a higher interest rate charged on such loans. To test whether this is the case for the loans with a higher Misreporting Index, I perform regressions of the following form:

$$Interest_i = \alpha_0 + \alpha_1 \times Misreporting\ Index_i + \alpha_2 \times X_i + \epsilon_i \tag{2}$$

where $Interest_i$ is the annual interest rate charged on the loan i, $Misreporting\ Index_i$ is a proxy for potential misreporting, described above, and X_i is a vector of control variables, as described above.

Default in most cases does not mean losing 100% of the original investment. Therefore, it is possible that even correctly priced loans could exhibit higher default rates and lower interest rates, in case the returns to investors were higher for defaulting loans. To test for this possibility, I perform regressions of the following form:

$$Loss\ given\ default_i = \alpha_0 + \alpha_1 \times Misreporting\ Index_i + \alpha_2 \times X_i + \epsilon_i \tag{3}$$

where $Loss\ given\ default_i$ is the fraction of principal lost for the defaulted loan i, $Misreporting\ Index_i$ is a proxy for potential misreporting, described above, and X_i is a vector of control variables, as described above. The sample for this analysis includes all defaulted loans.

The results from both of these analyses are shown in Table 4. Loans with a higher Misreporting Index value do not pay higher interest than other loans. In fact, the estimated coefficients are negative. This finding is the opposite of what could be expected if the loans were correctly priced. Combined with the earlier result, these findings suggest that misreporting is associated with significantly higher likelihood of default but not reflected in loan pricing. That is, the loan pricing fails to compensate investors for the increased risk.

The results also do not suggest that loans with a higher Misreporting Index value provide better payoffs to investors in cases of defaults. The estimated coefficient for *Misreporting Index* is positive, the opposite of what should be expected, and statistically significant.

4.3 Social norms, honesty, and uncertainty

In principle, the Misreporting Index can capture both deliberate misreporting and genuine uncertainty on part of the borrower. In this section I explore whether there is evidence pointing to either direction. As indicators of the likelihood of deliberate misreporting, I use the strength of social norms, as proxied by regional social capital, and the perceived trustworthiness of the borrower's stated profession.

The first, the strength of social norms, is suggested by the literature to affect borrower behavior. For example, Guiso et al. (2013) show the importance of social norms in decisions on whether to strategically default on mortgages. The ability of social norms to generate trust and trustworthy behavior is often referred to as *social capital*. Guiso et al. (2004)

show that social capital is important for the spread of financial contracts and hence for the development of financial markets. They argue that this is because individuals' willingness to sign financial contracts depends "not only on the enforceability of contracts, but also on the extent to which they trust the counterpart." Millo and Pasini (2010) find evidence that social capital mitigates moral hazard in insurance markets.

In the context of firm behavior, Hasan, Hoi, Wu, and Zhang (2017) find that firms located in high-social-capital counties are less prone to engage in corporate tax avoidance. In another paper (Hasan et al., 2016), the same authors study bank loan data and find that high social capital is associated with lower bank loan spreads, looser other loan terms, and lower at-issue bond spreads. Their results suggest that social capital imposes behavioral norms on firms and hence mitigates the risk of opportunistic firm behavior against debtholders. Moreover, Jha and Chen (2015) find evidence that audit firms judge the trustworthiness of their clients based partly on where they are located, and firms headquartered in high-social-capital counties pay lower audit fees.

To study whether social capital reduces the propensity to misreport, I perform regressions of the following form:

$$Misreporting\ Index_i = \alpha_0 + \alpha_1 \times Social\ capital_i + \alpha_2 \times X_i + \epsilon_i \tag{4}$$

where $Misreporting\ Index_i$ is the misreporting index value for loan i. $Social\ capital_i$ is the social capital index by Lin and Pursiainen (2018a) at the borrower's home location. The original index is constructed at county level. For this analysis, I calculate population-weighted average values for each 3-digit zip code level. X_i is a vector of control variables, as described above. Given social capital is a location-specific variable and does not change much over time, I include additional controls for the population and wealth level (personal income per capita) of the borrower location (measured at the 3-digit zip code level).

The second indicator, borrower honesty, is based on the reported employment title of the borrower and a yearly Gallup survey that asks a large sample of adults living across all U.S. states how honest they believe people in different professions are. The precise interview question asked is: "Please tell me how you would rate the honesty and ethical standards of people in these different fields – very high, high, average, low, or very low?" I define the *Honesty* variable for each profession as the percentage of respondents that assessed the honesty of people within the profession as "very high" or "high".

The Lending Club data include a text field where the borrower can state their employment title. I manually go through the 500 most common employment titles and match them to the professions included in the Gallup survey. I complement this matching by adding additional identification rules that seem likely to capture words specific to only one profession. For

example, if the employment title includes the word "nurse", I classify the profession as nurse, or if the title includes the word "police", I classify it as police. This methodology allows me to match honesty scores with 402,432 loans. I then use the honesty variable to perform a similar regression analysis, but replacing *Social capital* with *Honesty*, the percentage of people assessing the borrower's profession as very high or high in terms of honesty.

As an indicator of genuine uncertainty, I construct a measure of expected earnings volatility using the reported employment titles and data from the Panel Study of Income Dynamics (PSID). I use PSID data for the period 2007-2015 and calculate the biyearly volatility of household income for each households included in the data for the full period. I then calculate the average volatility for each household based on the occupation of the head of the household in 2007. This metric hence represents a static estimate of expected volatility for each occupation in the data. Similar to the honesty data above, I manually match the most common employment titles to the occupations in PSID data to obtain the expected income volatility variables for P2P borrowers. I then perform the same regression as above using *Income volatility* as the independent variable.

The results of all these regressions are shown in Panel A of Table 5. Social capital and honesty are both associated with significantly lower Misreporting Index values, while expected income volatility is associated with higher misreporting propensity. These results suggest that there is both a deliberate and an inadvertent component captured by the Misreporting Index.

Panels B, C, and D show the same analysis, but using each of the components of the Misreporting Index separately as the dependent variable in the regressions. Interestingly, the pattern is the same for each of these components, although the relative magnitude and significance of the effects varies by Panel. Social capital and honesty are associated with a lower likelihood for each of the component dummies of the index to take the value one (apart from the case of *Honesty* and *Loan amount divisible 10,000*, where the estimated coefficient is not statistically significant). Similarly, expected income volatility significantly increases the likelihood of each of the component dummies to take the value one.

4.4 Time-varying propensity to misreport

It seems reasonable that borrower behavior might change over time. For example, Gross and Souleles (2002) find evidence of the time-varying propensity of borrowers to default on their credit card debt. Hence, it is possible that borrowers' tendency to misreport may also change over time. To test prediction, I first estimate quarterly fixed effects based on the following regression:

$$Misreporting\ Index_i = \alpha_0 + \alpha_1 \times Quarter_i + \alpha_2 \times X_i + \epsilon_i \quad (5)$$

where $Misreporting\ Index_i$ is the misreporting index value for loan i. $Quarter_i$ is a vector of dummies indicating the quarter during which the loan was issued. X_i is a vector of control variables, as described above.

The estimated quarter fixed effects are shown in Figure 15. There is a notable increase in suspect loans from the second half of 2017 onward. This increase appears to coincide with the inauguration of Donald Trump as the president of the United States in Q1 2017 and the subsequent debate on the truthfulness or lack thereof in politics and public debate more broadly. This naturally prompts the question whether this finding has something to do with the political climate and whether misreporting has become more socially acceptable during this period.

The psychology literature suggests that our perception of normal does affect our behavior. Bear and Knobe (2017) find that when people think about what is normal, they combine their sense of what is typical with their sense of what is ideal. Therefore, if the perceived typical level of truthfulness changes, this may not simply begin to be regarded as more typical, but also as more normal. It is therefore possible that the increased exposure to obvious lying by politicians and the related media coverage has lowered the bar for misreporting.

On the other hand, it is equally possible that borrower behavior did not fundamentally change during this period, but the platform itself changed its underwriting criteria or platform layout such that it facilitated loans with higher Misreporting Index values. There is no perfect way for me to check this, as the exact underwriting criteria are not disclosed by Lending Club. A more thorough investigation of the dynamics of the different index components, included in the Internet Appendix, suggests that the upward shift in misreporting comes primarily from the roundness of loan amounts. However, Lending Club Investor Relations, in an e-mail replying to my question, said that there had been no "major changes that would specifically affect the average loan amount". If this is accurate, it would suggest that the shift is driven by borrower behavior.

While I cannot identify the exact channel, I nevertheless test whether the propensity to misreport changes following the Trump inauguration, and whether there are cross-sectional differences depending on the borrower characteristics. I do this by performing the following regression:

$$
\begin{aligned}
Misreporting\ Index_i = &\alpha_0 + \alpha_1 \times Post_i \times ln(Annual\ income)_i \\
&+ \alpha_2 \times Post_i + \alpha_3 \times ln(Annual\ income)_i \\
&+ \alpha_4 \times X_i + \epsilon_i
\end{aligned} \tag{6}
$$

The results are shown in Panel A of Table 6. They show that misreporting increases significantly following the Trump inauguration. Furthermore, the increase is significantly larger within low-income borrowers. In Panel B, I replace the borrower income with an indicator variable of borrower credit grade. The results show that the increase in misreporting is significantly larger within low-credit-grade borrowers.

5 Additional analysis and robustness checks

5.1 Index components

Panel A of Table 7 shows the default regression analysis separately for each of the five Misreporting Index components. We see that each of the indicators is associated with significantly higher likelihood of default. The estimated coefficient is largest for *Suspect amount*, suggesting that it increases the likelihood of default the most of the three. In column 7, I include all components in the same regression. All five components remain statistically significant, suggesting that each indicator adds valuable information that is not captured by the other components of the index.

Panel B shows the same analysis for loan interest rates. The only estimated coefficient that is significantly different from zero is that for *Income divisible 5,000*. However, this coefficient is negative, suggesting that the higher default risk receives negative compensation.

Panel C shows the results for loss given default. Here, the estimated coeficients are positive and significant for all variables except the two indicators of loan amount roundness, which are not statistically significant. The economic magnitude of all of these effects remains very small, suggesting the differences are unlikely to matter much for loan investor returns.

5.2 Sensitivity of index components to functional forms

The components of the Misreporting Index rely on discontinuities in the three variables that I use the construct the index: loan / revolving balance ratio, loan amount, and reported income. In my base regression specification, I control for each of these continuous variables. To make sure that the apparent discontinuities are not driven by misspecified functional form of the control variables, I perform two additional analyses.

First, I perform the default analysis with a flexible polynomial form of each continuous variable used to determine index variable thresholds. I include as controls polynomial terms up to sixth degree for each of the three index variables: loan/revolving balance ratio, annual income, and loan amount. The results are shown in Table 8. Each of the index components remains significantly positively associated with the likelihood of default.

Second, I perform the default likelihood regressions for each of the index component variables, but only include loans where the continuous variable is very near the round threshold. This mitigates the potential concern that the results could be driven by misspecification, as the functional form of the control variable should not matter much as long as the values are very close to each other. Given the sample for each index component includes a set of loans around each threshold, I include dummies for each threshold and assign each loan to the dummy indicating the nearest round threshold. In other words, for the analysis with *Income divisible 10,000*, I include the loans where the loan amount is divisible by 10,000, or where the loan amount is at maximum 1,000 below or above a multiple of 10,000. So the sample includes loan amounts in the ranges 9,000-11,000, 19,000-21,000, and so on. The results are shown in Table 9, with the bandwidth around thresholds indicated above each model. All of the index component variables remain statistically significantly associated with higher default risk.

These findings suggest that the index components are quite robust to different model specifications and represent genuine discontinuities in the index variables.

5.3 Income verification

Some of the borrowers provide a proof of income, typically a pay slip. This allows comparing borrowers with verified income to those without verification. Hence, I perform a regression analysis of the likelihood of default depending on income roundness separately for the subsample of loans where the income is verified, and for those where it is not.

The results are shown in Table 10. Panel A shows the results for the full sample of loans. Panel B includes only the loans where borrower income is verified, and Panel C includes the loans without income verification. Interestingly, round income remains associated with significantly higher likelihood of default even when the income is verified. There are at least two alternative explanations for this finding. First, it may be that there is genuine uncertainty about the annual income, even when in possession of a monthly pay slip. This is most obviously the case when the borrower income includes a meaningful variable component. Second, the income verification process by the platform necessarily allows some margin of error (or ambiguity), and hence enables borrowers to smireport income, as long as it is within

reasonable limits.

6 Conclusion

I find that easily observable indicators of potential misreporting, based on consistency of the stated purpose with loan characteristics and behavioral cues, can be used to identify suspicious P2P loans. The Misreporting Index I construct of these indicators has significant predictive power over loan default rates. I also show that the loan pricing on the Lending Club platform does not appear to account for the increased default risk associated with loans with a higher Misreporting Index value.

These findings can help the rapidly growing pool of investors active in marketplace lending to make better investment decisions by incorporating signs of misreporting into their decision making process. More generally, the findings should be of interest to anyone designing automated credit grading algorithms or making credit decisions.

My findings on the determinants of misreporting provide additional context to existing studies of the implications of the reliance on round numbers in other contexts. I show suggestive evidence that my Misreporting Index captures elements of both deliberate misreporting and genuine uncertainty. Of course, for purposes of credit analysis alone, one could argue that the disctinction is irrelevant, as long as the index can predict defaults. But for making generalizations of the findings for other purposes, this point is significant.

The time trends of misreporting are also intriguing. There seems to be a substantial increase in misreporting which coincides with the Trump inauguration and the subsequent public debate about lying for political purposes and "fake news". On the other hand, the P2P sector generally and Lending Club specifically have been going through challenging times during the last few years, which could have caused them to become more aggressive on loan underwriting. This, in principle, could have driven some of the shift as well. Whether this is a co-incidence, a sign of a moral erosion manifested in the domain of P2P loans, or a reflection of changes made in the platform remains an important topic for further research.

References

Aitken, Michael, Philip Brown, Christine Buckland, H.Y. Izan, and Terry Walter, 1996, Price clustering on the australian stock exchange, *Pacific-Basin Finance Journal* .

ap Gwilym, Owain, Andrew Clare, and Stephen Thomas, 1998, Extreme price clustering in the london equity index futures and options markets, *Journal of Banking & Finance* 22, 1193–1206.

Baird, John C., Charles Lewis, and Daniel Romer, 1970, Relative frequencies of numerical responses in ratio estimation, *Perception & Psychophysics* .

Ball, Clifford A., Walter N. Torous, and Adrian E. Tschoegl, 1985, The degree of price resolution: the case of the gold market, *Journal of Futures Markets* 5, 29–43.

Bear, Adam, and Joshua Knobe, 2017, Normality: Part descriptive, part prescriptive, *Cognition* 167, 25–37.

Berger, Allen N., Nathan H. Miller, Mitchell A. Petersen, Raghuram G. Rajan, and Jeremy C. Stein, 2005, Does function follow organizational form? evidence from the lending practices of large and small banks, *Journal of Financial Economics* 76, 237–269.

Bhattacharya, Utpal, Craig W. Holden, and Stacey Jacobsen, 2012, Penny wise, dollar foolish: Buysell imbalances on and around round numbers, *Management Science* 58, 413–431.

Chava, Sudheer, Kershen Huang, and Shane A. Johnson, 2018, The dynamics of borrower reputation following financial misreporting, *Management Science* 64, 4775–4797.

Chen, Xiao, Bihong Huang, and Dezhu Yec, 2018, The role of punctuation in p2p lending: Evidence from china, *Economic Modelling* 68, 634–643.

DeGrauwe, Paul, and Danny Decupere, 1992, Psychological barriers in the foreign exchange market, *CEPR Discussion Paper* .

Dehaene, Stanislas, and Jacques Mehler, 1992, Cross-linguistic regularities in the frequency of number words, *Cognition* 43, 1–29.

Donaldson, Glen, and Harold Y. Kim, 1993, Price barriers in the dow jones industrial average, *The Journal of Financial and Quantitative Analysis* 28, 313–330.

Duarte, Jefferson, Stephan Siegel, and Lance Young, 2012, Trust and credit: The role of appearance in peer-to-peer lending, *The Review of Financial Studies* 25, 2455–2484.

Eid, Nourhan, Josephine Maltby, and Oleksandr Talavera, 2016, Income rounding and loan performance in the peer-to-peer market, *Working paper* .

Emekter, Riza, Yanbin Tu, Benjamas Jirasakuldech, and Min Lu, 2015, Evaluating credit risk and loan performance in online peer-to-peer (p2p) lending, *Applied Economics* 47, 54–70.

Federal Deposit Insurance Corporation, 2015, Marketplace lending, *FDIC Supervisory Insights* .

Freedman, Seth, and Ginger Zhe Jin, 2014, The information value of online social networks: Lessons from peer-to-peer lending, *NBER Working Paper* .

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

Goodhart, Charles, and Riccardo Curcio, 1991, The clustering of bid/ask prices and the spread in the foreign exchange market, *Discussion Paper No. 110, LSE Financial Markets Group Discussion Paper Series* .

Griffin, John M., and Gonzalo Maturana, 2016, Who facilitated misreporting in securitized loans?, *The Review of Financial Studies* 29, 384–419.

Gross, David B., and Nicholas S. Souleles, 2002, An empirical analysis of personal bankruptcy and delinquency, *The Review of Financial Studies* 15, 319–347.

Grossman, Sanford J., Merton H. Miller, Kenneth R. Cone, Daniel R. Fischel, and David J. Ross, 1997, Clustering and competition in asset markets, *The Journal of Law & Economics* 40, 23–60.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales, 2004, The role of social capital in financial development, *American Economic Review* 94, 526–556.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales, 2013, The determinants of attitudes toward strategic default on mortgages, *The Journal of Finance* 68, 1473–1515.

Hameed, Allaudeen, and Eric Terry, 1998, The effect of tick size on price clustering and trading volume, *Journal of Business Finance & Accounting* 25, 849–867.

Harris, Lawrence, 1991, Stock price clustering and discreteness, *The Review of Financial Studies* 4, 389–415.

Hasan, Iftekhar, Chun Keung (Stan) Hoi, Qiang Wu, and Hao Zhang, 2016, Social capital and debt contracting: Evidence from bank loans and public bonds, *Journal of Financial and Quantitative Analysis* Forthcoming.

Hasan, Iftekhar, Chun Keung (Stan) Hoi, Qiang Wu, and Hao Zhang, 2017, Does social capital matter in corporate decisions? evidence from corporate tax avoidance, *Journal of Accounting Research* Forthcoming.

Jansen, C.J.M., and M.M.W. Pollmann, 2001, On round numbers: Pragmatic aspects of numerical expressions, *Journal of Quantitative Linguistics* 8, 187–201.

Jha, Anand, and Yu Chen, 2015, Audit fees and social capital, *The Accounting Review* 90, 611–639.

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

Kahn, Charles, George Pennacchi, and Ben Sopranzetti, 1999, Bank deposit rate clustering: theory and empirical evidence, *Journal of Finance* 54, 2185–2214.

Kandel, Shmuel, Oded Sarig, and Avi Wohl, 2001, Do investors prefer round stock prices? evidence from israeli ipo auctions, *Journal of Banking & Finance* 25, 1543–1551.

Karlan, Dean, and Jonathan Zinman, 2008, Lying about borrowing, *Journal of the European Economic Association* 6, 510–521.

Karlan, Dean S., and Jonathan Zinman, 2012, List randomization for sensitive behavior: An application for measuring use of loan proceeds, *Journal of Development Economics* 98, 71–75.

Kaufman, E. L., M. W. Lord, T. W. Reese, and J. Volkmann, 1949, The discrimination of visual number, *The American Journal of Psychology* 62, 498–525.

Kleven, Henrik J., and Mazhar Waseem, 2013, Using notches to uncover optimization frictions and structural elasticities: Theory and evidence from pakistan, *The Quarterly Journal of Economics* 128, 669–723.

Koedijk, Kees G., and Philip Stork, 1994, Should we care? psychological barriers in stock markets, *Economics Letters* 44, 427–432.

Krueger, Lester E., 1982, Single judgments of numerosity, *Perception & Psychophysics* 31, 175–182.

Lacetera, Nicola, Devin G. Pope, and Justin R. Sydnor, 2012, Heuristic thinking and limited attention in the car market, *American Economic Review* 102, 2206–2236.

Ley, Eduardo, and Hal R. Varian, 1994, Are there psychological barriers in the dow jones index?, *Applied Financial Economics* 4, 217–224.

Lin, Mingfeng, Nagpurnanand R. Prabhala, and Siva Viswanathan, 2013, Judging borrowers by the company they keep: Friendship networks and information asymmetry in online peer-to-peer lending, *Management Science* 59, 17–35.

Lin, Tse-Chun, and Vesa Pursiainen, 2018a, Fund what you trust? social capital and moral hazard in crowdfunding, *Working paper* .

Lin, Tse-Chun, and Vesa Pursiainen, 2018b, The round number heuristic and crowdfunding performance, *Working paper* .

Lipton, Jennifer, and Elizabeth Spelke, 2005, Preschool children's mapping of number words to nonsymbolic numerosities, *Child Development* 76, 978–88.

Millo, Giovanni, and Giacomo Pasini, 2010, Does social capital reduce moral hazard? a network model for non-life insurance demand, *Fiscal Studies* 31, 341–372.

Mitchell, Jason, and H.Y. Izan, 2006, Clustering and psychological barriers in exchange rates, *Journal of International Financial Markets, Institutions and Money* 16, 318–344.

Niederhoffer, Victor, 1965, Clustering of stock prices, *Operations Research* 13, 258–265.

Ormerod, Catrin, and Felix Ritchie, 2007, Issues in the measurement of low pay, *Economic & Labour Market Review, U.K. Ofce for National Statistics* 1, 27–45.

Osborne, M. F. M., 1962, Periodic structure in the brownian motion of stock prices, *Operations Research* 10, 345–379.

Palmon, Oded, Barton A. Smith, and Ben J. Sopranzetti, 2004, Clustering in real estate prices: Determinants and consequences, *Journal of Real Estate Research* 26, 115–36.

Rosch, Eleanor, 1975, Cognitive reference points, *Cognitive Psychology* 7, 532–547.

Schindler, Robert M., and Patrick N. Kirby, 1997, Patterns of rightmost digits used in advertised prices: Implications for nine-ending effects, *Journal of Consumer Research* 24, 192–201.

Serrano-Cinca, Carlos, Begoa Gutirrez-Nieto, and Luz Lpez-Palacios, 2015, Determinants of default in p2p lending, *PLoS One* 10.

Sopranzetti, Ben J., and Vinay Datar, 2002, Price clustering in foreign exchange spot markets, *Journal of Financial Markets* 5, 411–417.

U.S. Department of the Treasury, 2016, Opportunities and challenges in online marketplace lending, *Treasury Note* .

Whynes, David K., Emma J. Frew, Zoë N. Philips, Judith Covey, and Richard D. Smith, 2007, On the numerical forms of contingent valuation responses, *Journal of Economic Psychology* 28, 462–476.

Ziegler, Tania, E.J. Reedy, Annie Le, Bryan Zhang, Randall S. Kroszner, and Kieran Garvey, 2017, The 2017 americas alternative finance industry report, *Chicago Booth School of Business, Polsky Center for Entrepreneurship and Innovation, and Cambridge Centre on Alternative Finance* .

Zinman, Jonathan, 2009, Where is the missing credit card debt? clues and implications, *Review of Income and Wealth* 55, 249–265.

Appendix A: Definitions of variables

Variable	Definition
Interest	The loan interest, set by Lending Club.
Loan amount	The loan amount chosen by the borrower.
Term 5 years	Dummy taking the value 1 if the loan term is 5 years, zero if the loan term is 3 years.
Loan / revolving balance	Loan amount divided by the outstanding revolving balance of the borrower.
Current performing	Dummy taking the value 1 if the loan is still outstanding and performing.
Fully paid	Dummy taking the value 1 if the loan has been paid in full.
Loan late	Dummy taking the value 1 if the loan payment is late by up to 120 days.
Default	Dummy taking the value 1 if the loan status is classified as "default" or "charged off".
Prepaid	Dummy taking the value 1 if the loan has been paid in full before maturity.
Misreporting Index	Calculated as the sum of five dummies: *Suspect amount*, *Income divisible 5,000*, *Income divisible 10,000*, *Loan amount divisible 5,000*, and *Loan amount divisible 10,000.*
Suspect amount	Dummy taking the value 1 if the loan / revolving balance ratio is below 0.9 or above 1.1.
Suspect amount above	Dummy taking the value 1 if the loan / revolving balance ratio is above 1.1.
Suspect amount below	Dummy taking the value 1 if the loan / revolving balance ratio is below 0.9.
Income divisible 5,000	Dummy taking the value 1 if the reported borrower income is divisible by 5,000.
Income divisible 10,000	Dummy taking the value 1 if the reported borrower income is divisible by 10,000.
Loan amount divisible 5,000	Dummy taking the value 1 if the loan amount is divisible by 5,000.
Loan amount divisible 10,000	Dummy taking the value 1 if the loan amount is divisible by 10,000.
Annual income	The reported annual income of the borrower.
Debt / income	The borrower's debt/income ratio, calculated as monthly debt payments divided by monthly income.
Employment length	The reported length of the borrower's current employment.
No employment length	Dummy taking the value 1 if the length of employment is not reported.
Revolving balance	The outstanding revolving credit balance of the borrower.
Revolving utilization	The revolving credit utilization of the borrower.
Total accounts	The total number of credit lines in the borrower's credit file.
Open accounts	The number of open credit lines in the borrower's credit file.
Home owned	Dummy taking the value 1 of the home ownership status provided by the borrower is "own".
Home mortgaged	Dummy taking the value 1 of the home ownership status provided by the borrower is "mortgage".
Honesty	The percentage of respondents that rank the honesty of people in borrower's profession very high or high in an annual Gallup survey.
Income volatility	Expected income volatility based on borrower employment title, using occupation-level averages of biannual income volatility calculated with data from the Panel Study of Income Dynamics (PSID) in 2007-2015.

Appendix B: Summary of the sample

This table summarizes how I filter the sample. My initial raw data include all Lending Club loans issued until Q 2018. For my analysis, I only include loans where the stated purpose of the loan is either credit card repayment or debt consolidation. For the analysis of historical default rates, I exclude current loans.

	Number of loans	% of total
Lending Club – all loans 2007 - Q2 2018	2,004,058	
Of which Credit card & debt consolidation (full sample)	1,584,127	79%
Of which non-current (default analysis sample)	905,316	45%

Figure 1: Number of loans by purpose

Distribution of loans by stated purpose. Includes all loans issued on Lending Club from 2007 until Q2 2018.

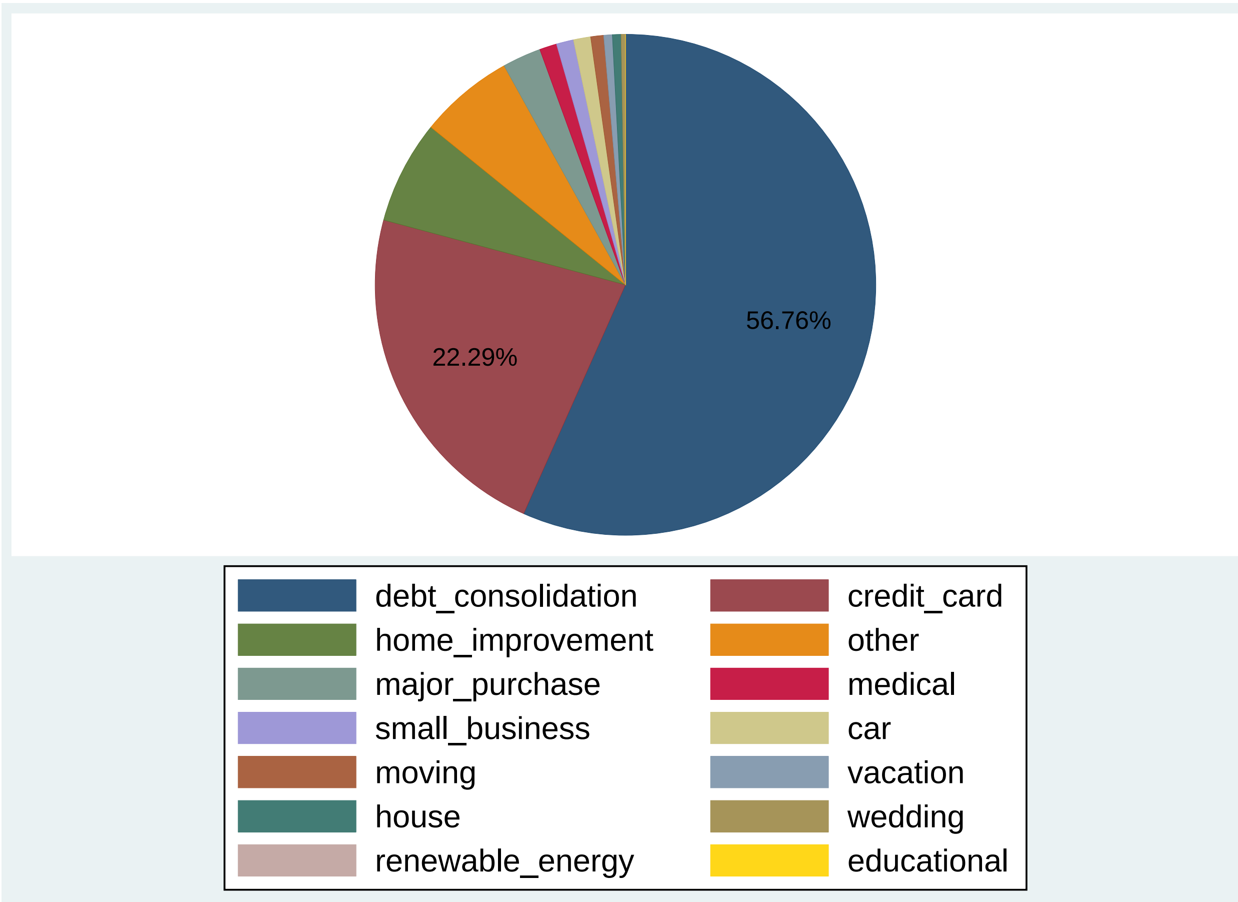


Figure 2: Loan amount / revolving balance – distribution

Distribution of loans by the loan amount / revolving balance ratio, shown separately for loans taken for credit card and debt consolidation, and for other purposes. Includes all loans issued on Lending Club from 2007 until Q2 2018. The red lines are drawn at 0.9 and 1.1. I define credit card and debt consolidation loan amounts outside this range as "suspicious".

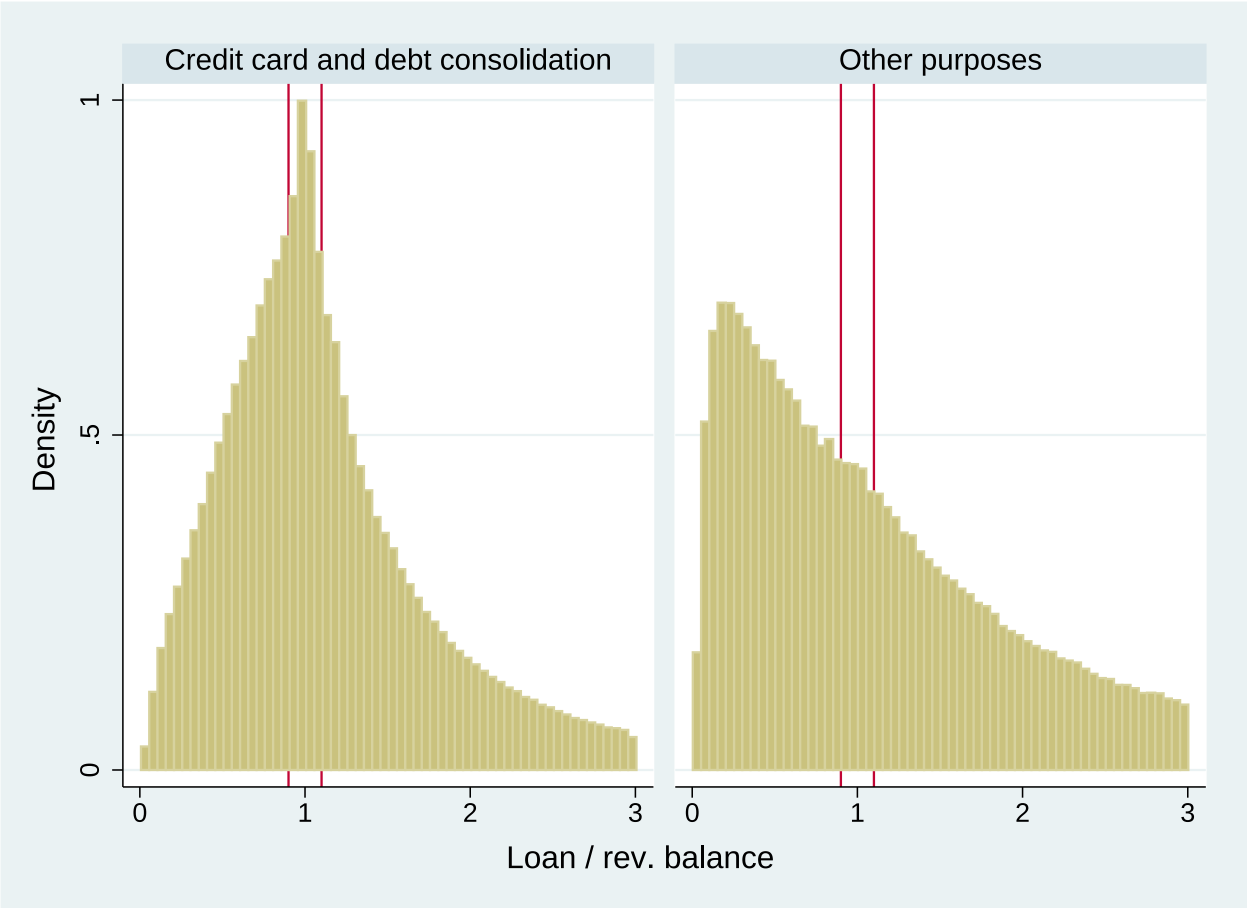
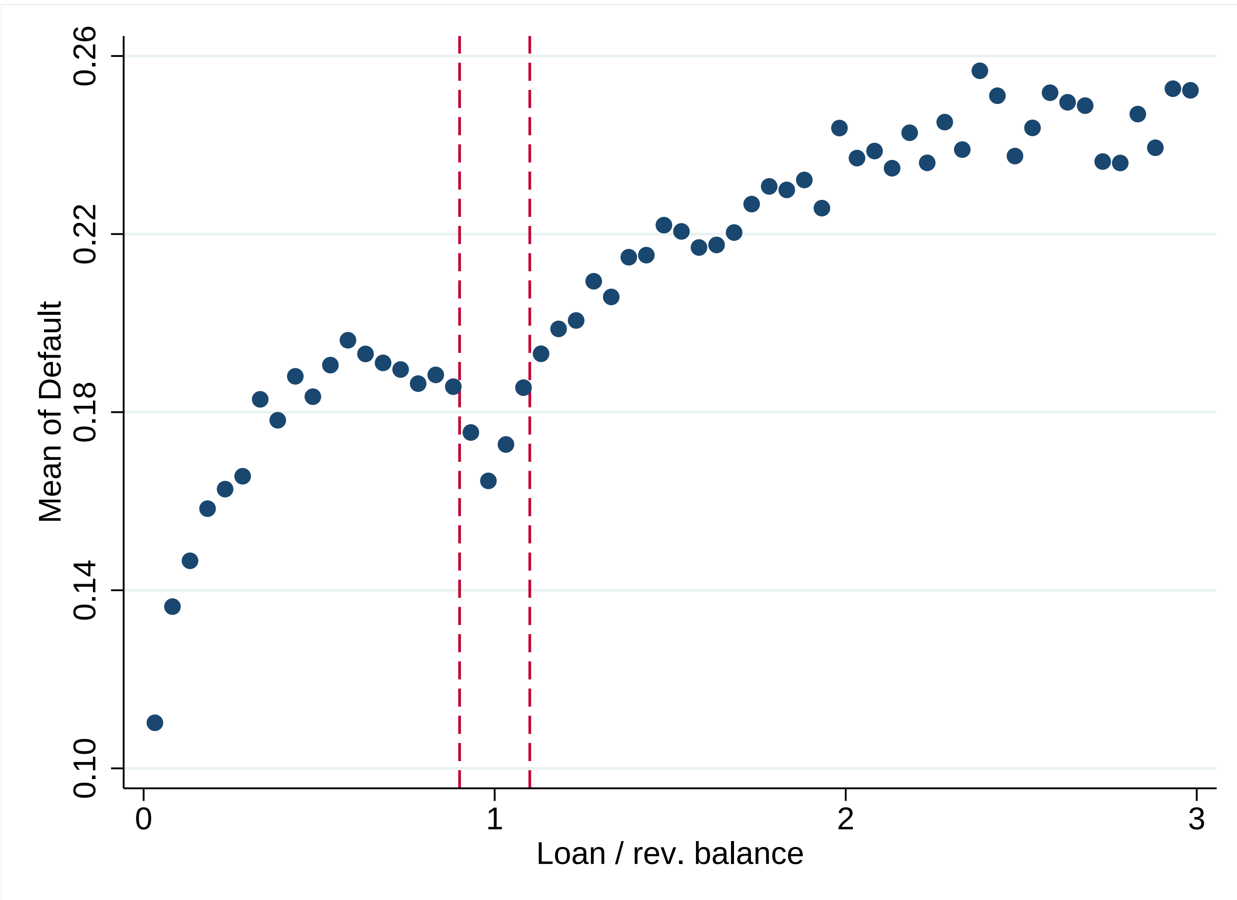


Figure 3: Mean default rate - credit card and debt consolidation

Average default rate of loans by the loan amount / revolving balance ratio. Includes all loans issued on Lending Club for the purpose of credit card or debt consolidation from 2007 until Q2 2018. The red lines are drawn at 0.9 and 1.1. I define credit card and debt consolidation loan amounts outside this range as "suspicious".

A: Mean default rate



B: Residual from regression of *Default* on subgrade dummies and interest

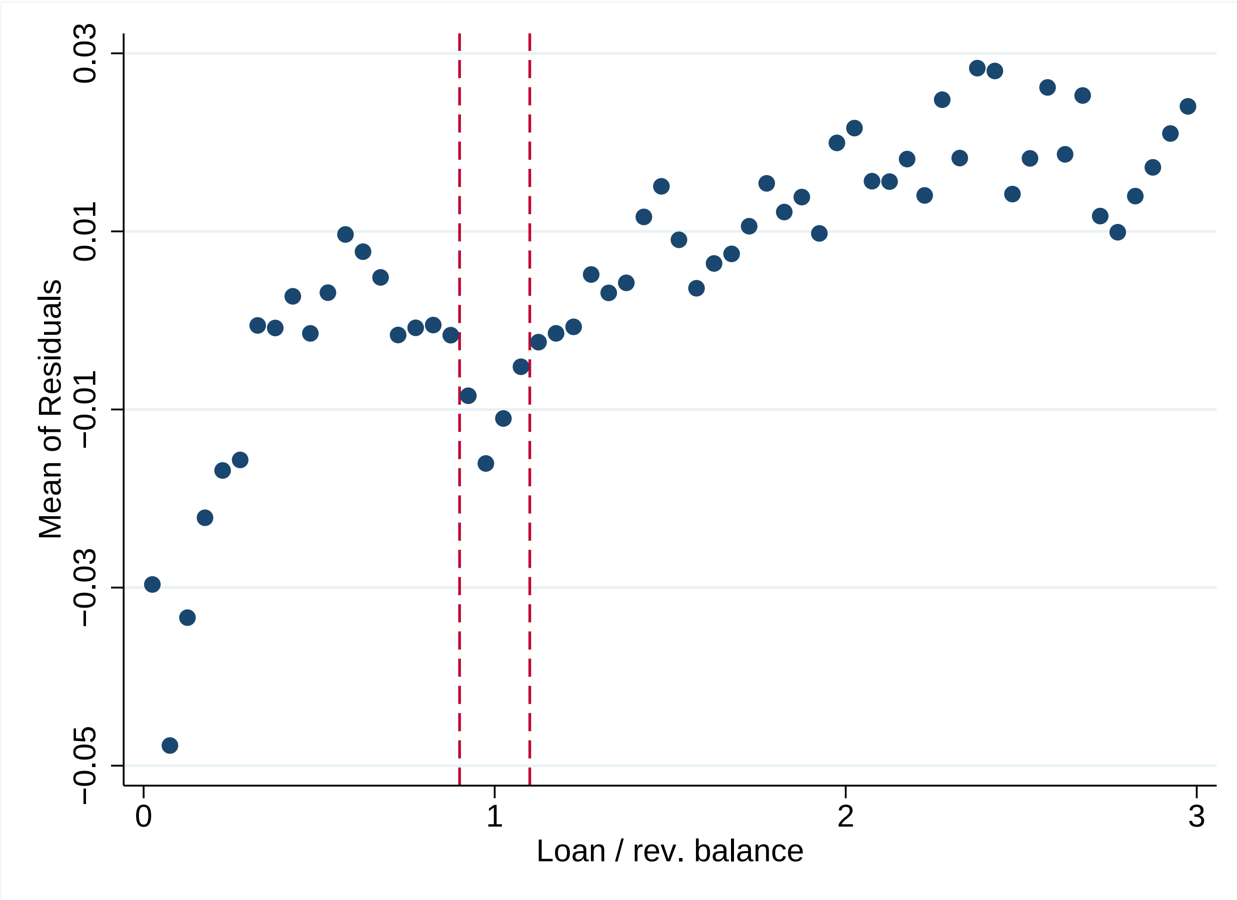


Figure 4: Mean default rate - other purposes

Average default rate of loans by the loan amount / revolving balance ratio. Includes all loans issued on Lending Club for purposes other than credit card or debt consolidation from 2007 until Q2 2018. The red lines are drawn at 0.9 and 1.1. I define credit card and debt consolidation loan amounts outside this range as "suspicious".

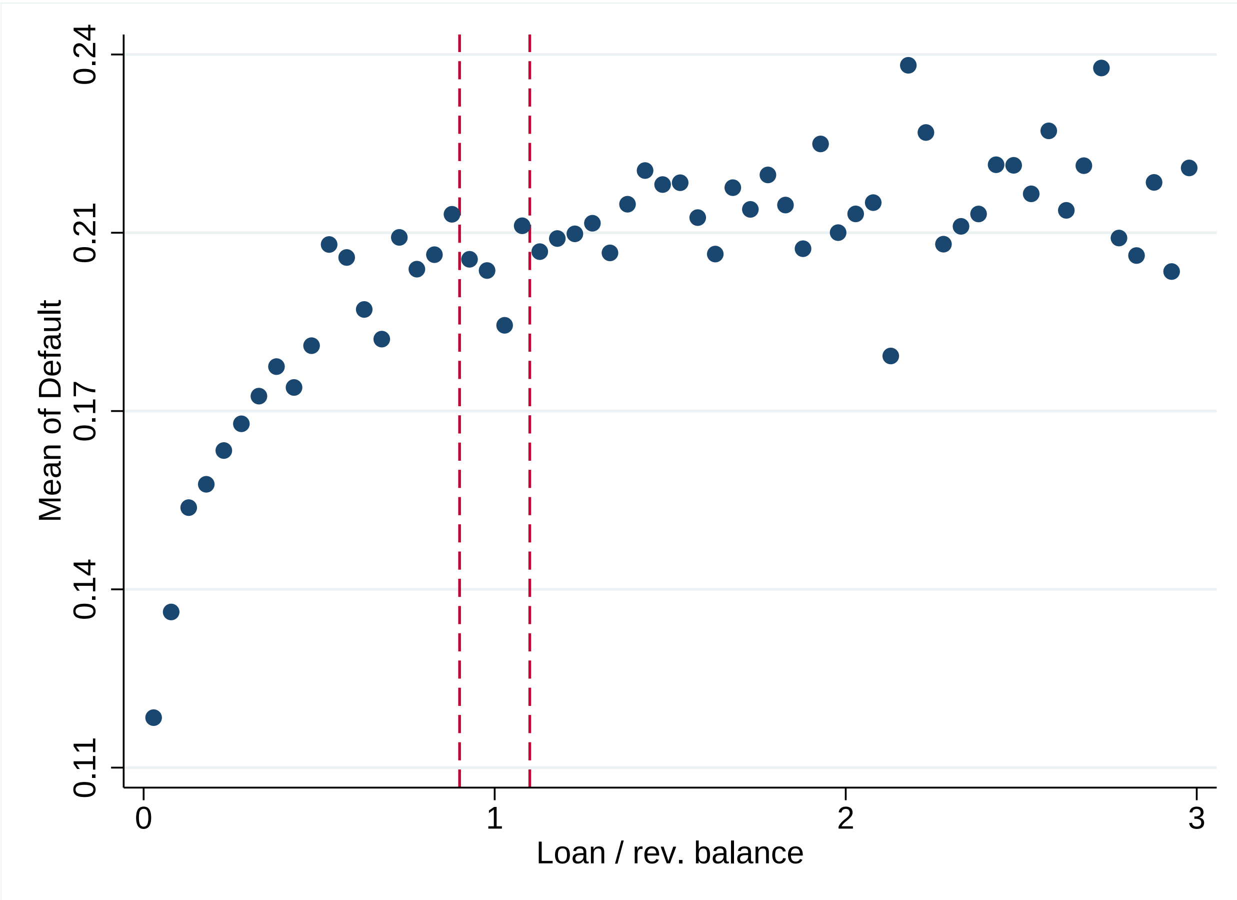


Figure 5: Annual income distribution

Distribution of loans by reported annual income of the borrower. Includes all credit card and debt consolidation loans issued on Lending Club from 2007 until Q2 2018.

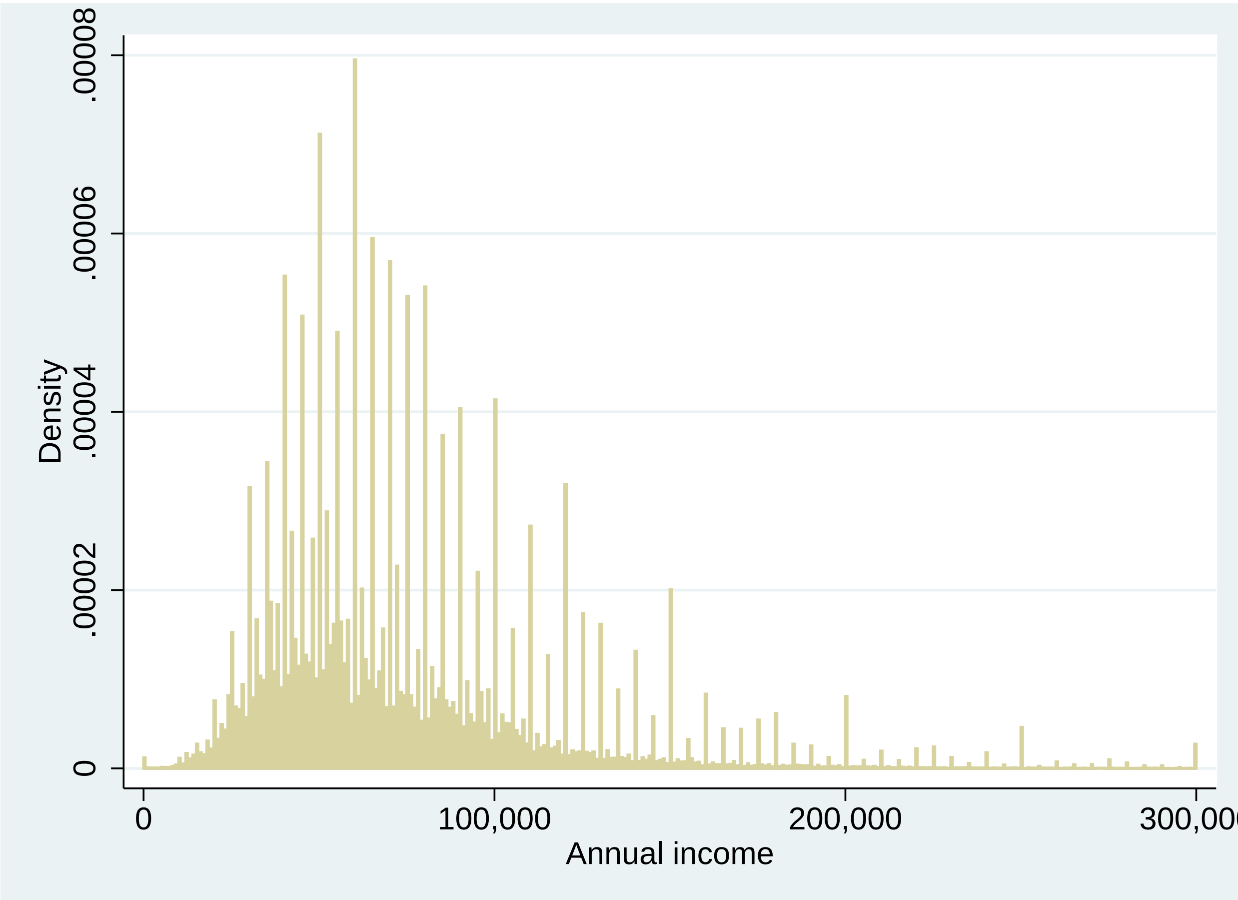


Figure 6: Loan amount distribution

Distribution of loans by loan amount. Includes all credit card and debt consolidation loans issued on Lending Club from 2007 until Q2 2018.

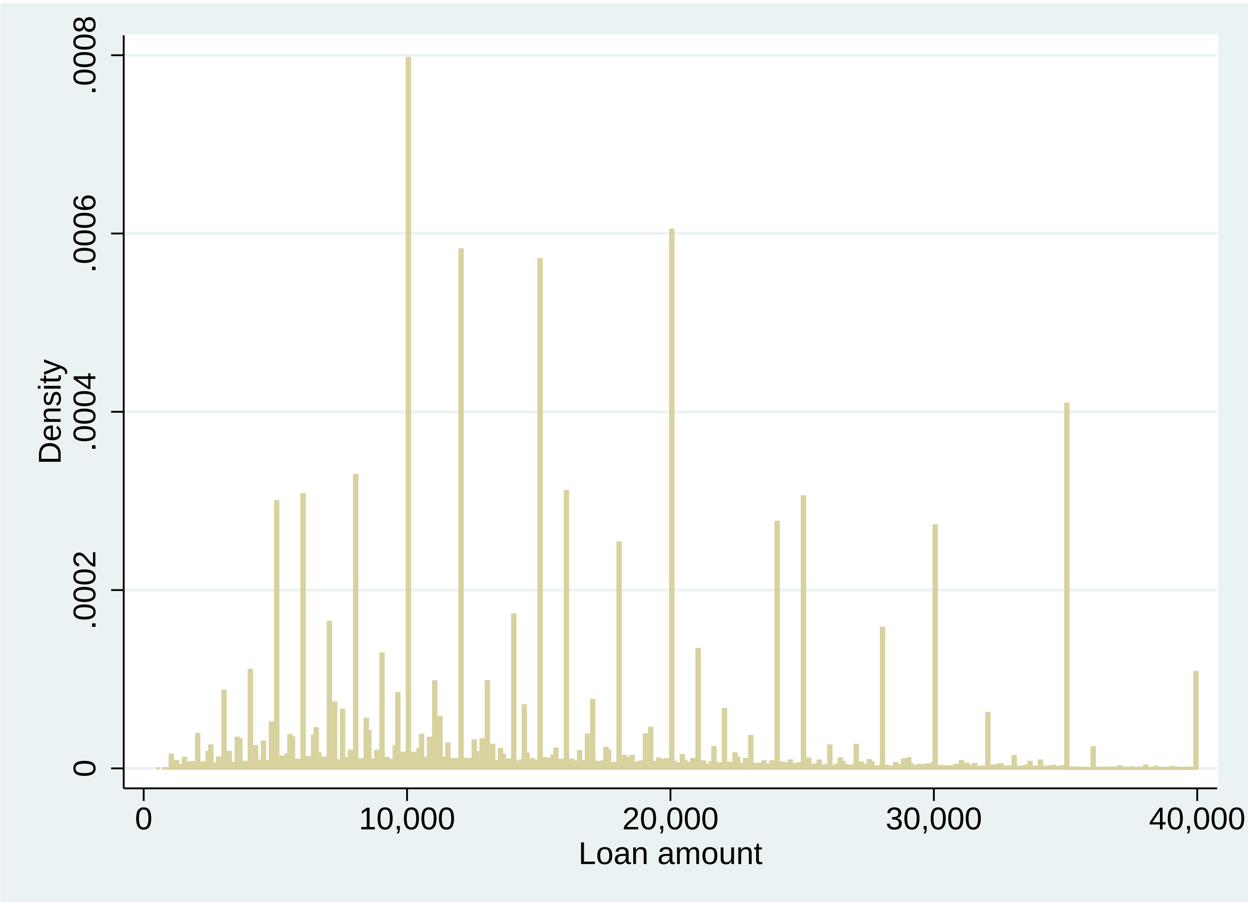


Figure 7: Revolving balance distribution

Distribution of loans by outstanding revolving balance of the borrower. Includes all credit card and debt consolidation loans issued on Lending Club from 2007 until Q2 2018.

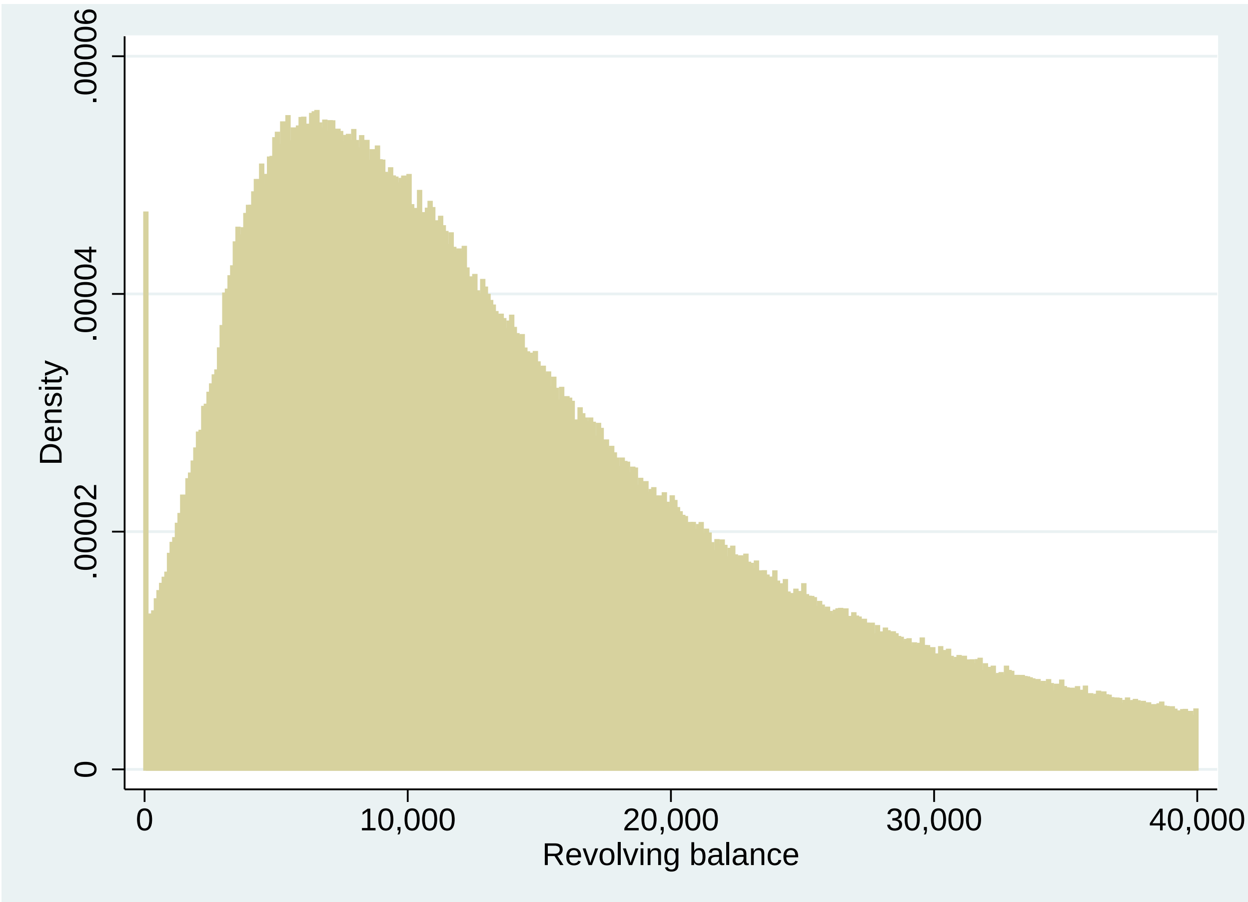


Figure 8: Misreporting Index - quarter fixed effects

Estimated quarter fixed effects from a regression of Misreporting Index on a full set of control variables on loan, borrower, and location characteristics. Includes all credit card and debt consolidation loans issued on Lending Club from 2010 until Q2 2018.

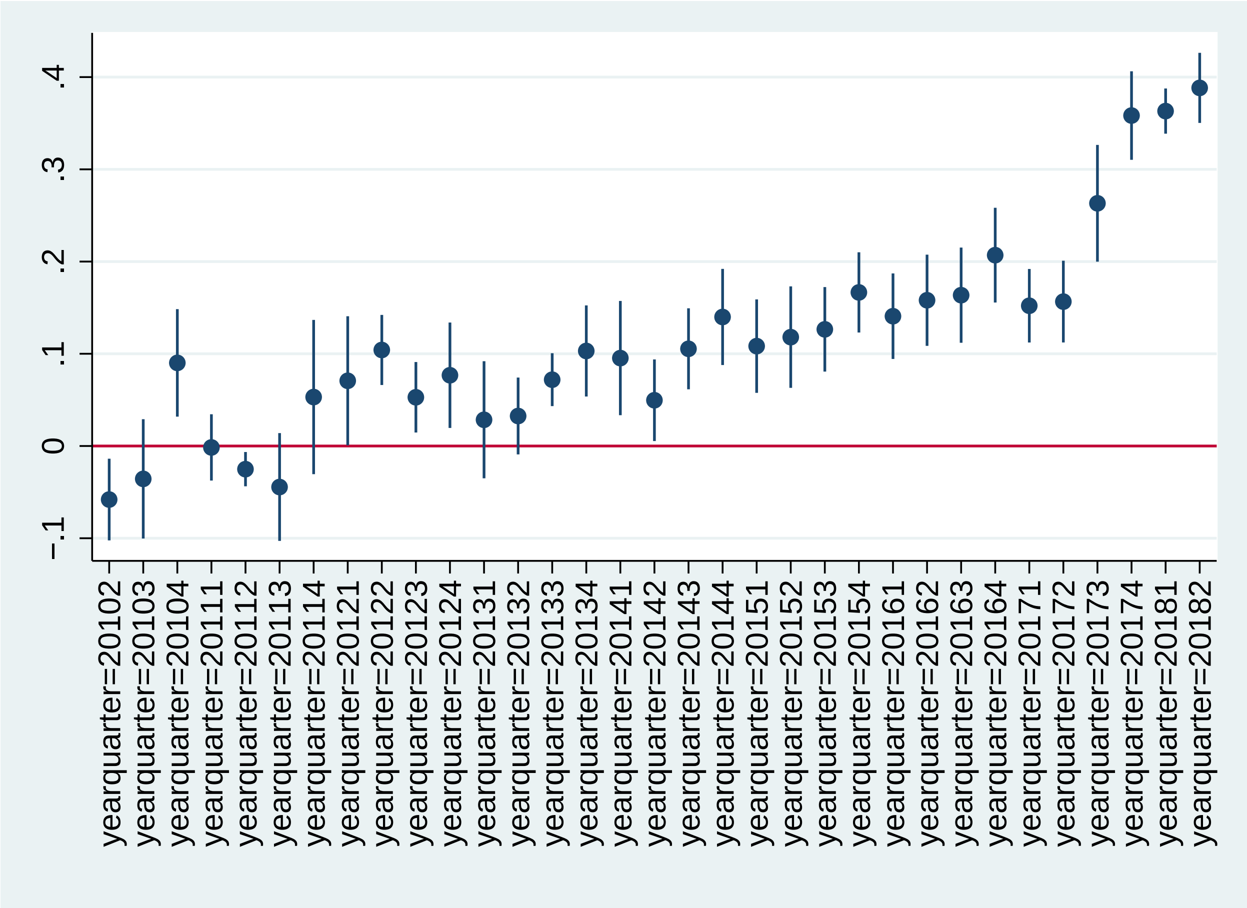


Table 1
Loans by year

Number of loans divided into suspicious and other loans. Includes all credit card and debt consolidation loans issued on Lending Club from 2010 until Q2 2018.

	Number of loans by status			
	Fully paid	Defaulted	Current	Total
2007	227	67		294
2008	1,109	266		1,375
2009	2,472	348		2,820
2010	6,306	968		7,274
2011	11,438	2,052		13,490
2012	34,503	6,673		41,176
2013	94,038	17,450	1,957	113,445
2014	148,393	33,837	16,298	198,528
2015	218,073	59,438	74,534	352,045
2016	128,669	45,030	166,809	340,508
2017	67,085	16,344	253,120	336,549
2018	9,905	625	166,093	176,623
Total	722,218	183,098	678,811	1,584,127

Table 2
Summary statistics

Summary statistics for the sample, including mean, standard deviation, and key percentiles. Variables are defined in Appendix A.

	Mean	Std	p25	p50	p75
Loan variables					
Interest (%)	13.043	4.710	9.670	12.620	15.610
Loan amount ('000)	15.653	8.857	9.000	14.000	20.300
Term 5 years	0.301	0.459	0.000	0.000	1.000
Loan / rev. balance	3.878	136.106	0.726	1.059	1.638
Loan status					
Current performing	0.408	0.491	0.000	0.000	1.000
Fully paid	0.456	0.498	0.000	0.000	1.000
Loan late	0.020	0.141	0.000	0.000	0.000
Default	0.116	0.320	0.000	0.000	0.000
Outcome (mature only)					
Prepaid	0.604	0.489	0.000	1.000	1.000
Fully paid	0.194	0.395	0.000	0.000	0.000
Default	0.202	0.402	0.000	0.000	0.000
Misreporting					
Misreporting Index	2.133	1.263	1.000	2.000	3.000
Suspect amount	0.839	0.368	1.000	1.000	1.000
Suspect amt. above	0.473	0.499	0.000	0.000	1.000
Suspect amt. below	0.365	0.482	0.000	0.000	1.000
Income div. 5,000	0.491	0.500	0.000	0.000	1.000
Income div. 10,000	0.293	0.455	0.000	0.000	1.000
Loan amt. div. 5,000	0.334	0.472	0.000	0.000	1.000
Loan amt. div. 10,000	0.177	0.382	0.000	0.000	0.000
Borrower					
Annual income ('000)	76.717	121.890	46.000	65.000	91.000
Debt / income (%)	19.251	13.214	12.600	18.380	24.840
Employment length (y.)	5.610	3.866	2.000	6.000	10.000
No employment length	0.060	0.237	0.000	0.000	0.000
Rev. balance ('000)	17.583	22.471	6.878	12.328	21.383
Revolving util. (%)	53.219	23.569	35.500	53.300	71.300
Total accounts	24.701	11.922	16.000	23.000	31.000
Open accounts	11.843	5.590	8.000	11.000	15.000
Home owned	0.104	0.306	0.000	0.000	0.000
Home mortgaged	0.487	0.500	0.000	0.000	1.000
Honesty	0.396	0.257	0.170	0.377	0.645
Income volatility	0.312	0.060	0.279	0.305	0.338
N	1,584,127				

Table 3
Likelihood of default vs. Misreporting Index

The dependent variable, *Default*, is a dummy taking the value one if the loan defaulted. Variables are defined in Appendix A. I include *Sub-grade x Year fixed effects* based on the credit grade assigned by Lending Club (35 grades times 12 years), *Purpose fixed effects* (credit card or debt consolidation), *3-digit zip fixed effects* (955 zip codes) based on the address of the borrower, and *Year-month x Term fixed effects* (133 months times two alternative terms). Includes all non-current credit card and debt consolidation loans. Heteroscedasticity-consistent standard errors, double-clustered by sub-grade and issue month, are shown in parentheses.

Panel A: Misreporting Index as continuous variable

	(1)	(2)	(3)	(4)	(5)
Misreporting Index	0.0048***	0.0077***	0.0074***	0.0079***	0.0082***
	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0005)
ln(Loan amount)		0.0413***	0.0411***	0.0259***	0.0248***
		(0.0053)	(0.0052)	(0.0034)	(0.0033)
Loan / rev. balance		−0.0000*	−0.0000	−0.0000*	−0.0000
		(0.0000)	(0.0000)	(0.0000)	(0.0000)
ln(Annual income)		−0.0424***	−0.0401***	−0.0395***	−0.0382***
		(0.0042)	(0.0039)	(0.0037)	(0.0035)
Debt / income (%)		0.0011***	0.0010***	0.0011***	0.0010***
		(0.0002)	(0.0002)	(0.0002)	(0.0002)
ln(1 + Emp. length)		0.0006	−0.0009	−0.0013*	−0.0012*
		(0.0006)	(0.0006)	(0.0007)	(0.0007)
No employment length		0.0698***	0.0663***	0.0714***	0.0718***
		(0.0048)	(0.0047)	(0.0045)	(0.0045)
ln(Revolving balance)		−0.0167***	−0.0155***	−0.0174***	−0.0177***
		(0.0023)	(0.0021)	(0.0023)	(0.0023)
Revolving util. (%)		0.0005***	0.0005***	0.0007***	0.0008***
		(0.0001)	(0.0001)	(0.0001)	(0.0001)
ln(Open accounts)		0.0514***	0.0507***	0.0567***	0.0579***
		(0.0051)	(0.0051)	(0.0055)	(0.0054)
ln(Total accounts)		−0.0154***	−0.0161***	−0.0215***	−0.0215***
		(0.0028)	(0.0029)	(0.0032)	(0.0031)
Home owned		−0.0269***	−0.0334***	−0.0337***	−0.0339***
		(0.0031)	(0.0037)	(0.0035)	(0.0035)
Home mortgaged		−0.0449***	−0.0491***	−0.0530***	−0.0538***
		(0.0044)	(0.0048)	(0.0048)	(0.0048)
Sub-grade FE	Yes	Yes	Yes	Yes	No
Sub-grade x Year FE	No	No	No	No	Yes
Purpose FE	No	Yes	Yes	Yes	Yes
3-digit zip FE	No	No	Yes	Yes	Yes
Year-Quarter FE	No	Yes	Yes	No	No
Year-month x Term FE	No	No	No	Yes	Yes
N	905,316	903,073	903,040	903,040	903,038
R^2	0.075	0.096	0.100	0.107	0.109

Significance levels: * 0.1, ** 0.05, * 0.01. Standard errors in parentheses.**

Panel B: Misreporting Index value dummies

	(1)	(2)	(3)	(4)	(5)
Misreporting Index=1	0.0140***	0.0199***	0.0201***	0.0201***	0.0195***
	(0.0019)	(0.0020)	(0.0020)	(0.0019)	(0.0019)
Misreporting Index=2	0.0161***	0.0261***	0.0258***	0.0261***	0.0262***
	(0.0021)	(0.0024)	(0.0024)	(0.0024)	(0.0024)
Misreporting Index=3	0.0189***	0.0320***	0.0316***	0.0320***	0.0320***
	(0.0019)	(0.0021)	(0.0022)	(0.0021)	(0.0021)
Misreporting Index=4	0.0263***	0.0394***	0.0388***	0.0415***	0.0423***
	(0.0030)	(0.0031)	(0.0031)	(0.0032)	(0.0032)
Misreporting Index=5	0.0347***	0.0479***	0.0472***	0.0496***	0.0504***
	(0.0037)	(0.0038)	(0.0037)	(0.0038)	(0.0037)
ln(Loan amount)		0.0416***	0.0414***	0.0261***	0.0250***
		(0.0053)	(0.0052)	(0.0034)	(0.0032)
Loan / rev. balance		−0.0000*	−0.0000	−0.0000	−0.0000
		(0.0000)	(0.0000)	(0.0000)	(0.0000)
ln(Annual income)		−0.0425***	−0.0402***	−0.0396***	−0.0383***
		(0.0042)	(0.0039)	(0.0037)	(0.0036)
Debt / income (%)		0.0011***	0.0010***	0.0011***	0.0010***
		(0.0002)	(0.0002)	(0.0002)	(0.0002)
ln(1 + Emp. length)		0.0005	−0.0010	−0.0014**	−0.0013*
		(0.0006)	(0.0007)	(0.0007)	(0.0007)
No employment length		0.0696***	0.0661***	0.0712***	0.0716***
		(0.0048)	(0.0048)	(0.0045)	(0.0045)
ln(Revolving balance)		−0.0166***	−0.0154***	−0.0173***	−0.0176***
		(0.0022)	(0.0021)	(0.0023)	(0.0023)
Revolving util. (%)		0.0005***	0.0005***	0.0007***	0.0008***
		(0.0001)	(0.0001)	(0.0001)	(0.0001)
ln(Open accounts)		0.0515***	0.0508***	0.0569***	0.0581***
		(0.0051)	(0.0051)	(0.0055)	(0.0054)
ln(Total accounts)		−0.0155***	−0.0162***	−0.0216***	−0.0217***
		(0.0028)	(0.0029)	(0.0032)	(0.0031)
Home owned		−0.0270***	−0.0335***	−0.0338***	−0.0339***
		(0.0031)	(0.0037)	(0.0035)	(0.0035)
Home mortgaged		−0.0450***	−0.0492***	−0.0531***	−0.0539***
		(0.0044)	(0.0048)	(0.0048)	(0.0048)
Sub-grade FE	Yes	Yes	Yes	Yes	No
Sub-grade x Year FE	No	No	No	No	Yes
Purpose FE	No	Yes	Yes	Yes	Yes
3-digit zip FE	No	No	Yes	Yes	Yes
Year-Quarter FE	No	Yes	Yes	No	No
Year-month x Term FE	No	No	No	Yes	Yes
N	905,316	903,073	903,040	903,040	903,038
R^2	0.075	0.096	0.100	0.107	0.109

Significance levels: * 0.1, ** 0.05, * 0.01. Standard errors in parentheses.**

Table 4
Interest rate and loss given default

The dependent variable is shown above each model. *Interest* is the annual interest rate of the loan. *Loss given default* is the proportion of loan principal lost following default. Control variables are the same as in Table 3. Columns 1-3 include all non-current credit card and debt consolidation loans. Columns 4-6 include all defaulted credit card and debt consolidation loans. Heteroscedasticity-consistent standard errors, double-clustered by sub-grade and issue month, are shown in parentheses.

	Interest (%)			Loss given default		
	(1)	(2)	(3)	(4)	(5)	(6)
Misreporting Index	−0.0075	−0.0028*	−0.0004	0.0088***	0.0014***	0.0014***
	(0.0067)	(0.0017)	(0.0003)	(0.0007)	(0.0004)	(0.0004)
ln(Loan amount)		0.0179**	0.0035		−0.0046***	−0.0039***
		(0.0076)	(0.0028)		(0.0013)	(0.0013)
Loan / rev. balance		−0.0000	−0.0000		0.0000	0.0000
		(0.0000)	(0.0000)		(0.0000)	(0.0000)
ln(Annual income)		−0.0309***	−0.0052		0.0104***	0.0103***
		(0.0102)	(0.0034)		(0.0016)	(0.0016)
Debt / income (%)		0.0013***	−0.0003**		0.0005***	0.0005***
		(0.0004)	(0.0001)		(0.0001)	(0.0001)
ln(1 + Emp. length)		0.0029*	0.0013**		−0.0084***	−0.0084***
		(0.0016)	(0.0006)		(0.0007)	(0.0008)
No employment length		−0.0084	0.0037		−0.0227***	−0.0231***
		(0.0066)	(0.0024)		(0.0025)	(0.0025)
ln(Revolving balance)		0.0144***	0.0006		0.0044***	0.0046***
		(0.0051)	(0.0010)		(0.0007)	(0.0007)
Revolving util. (%)		−0.0006**	0.0000		−0.0003***	−0.0003***
		(0.0003)	(0.0000)		(0.0000)	(0.0000)
ln(Open accounts)		−0.0107	0.0013		−0.0279***	−0.0284***
		(0.0078)	(0.0017)		(0.0014)	(0.0014)
ln(Total accounts)		−0.0029	−0.0034*		0.0173***	0.0176***
		(0.0054)	(0.0018)		(0.0014)	(0.0013)
Home owned		0.0012	0.0022*		0.0012	0.0013
		(0.0032)	(0.0012)		(0.0012)	(0.0012)
Home mortgaged		−0.0026	0.0011		0.0012	0.0014
		(0.0037)	(0.0010)		(0.0010)	(0.0010)
Sub-grade FE	Yes	Yes	No	Yes	Yes	No
Sub-grade x Year FE	No	No	Yes	No	No	Yes
Purpose FE	No	Yes	Yes	No	Yes	Yes
3-digit zip FE	No	Yes	Yes	No	Yes	Yes
Year-month x Term FE	No	Yes	Yes	No	Yes	Yes
N	905,316	903,040	903,038	183,098	182,663	182,651
R^2	0.953	0.979	0.996	0.109	0.307	0.310

Significance levels: * 0.1, ** 0.05, * 0.01. Standard errors in parentheses.**

Table 5
Honesty vs. uncertainty

The dependent variable is shown above each Panel. I include *Sub-grade fixed effects* based on the credit grade assigned by Lending Club (35 grades), *Purpose fixed effects* (credit card or debt consolidation), *3-digit zip fixed effects* (955 zip codes) based on the address of the borrower, and *Year-month x Term fixed effects* (133 months times two alternative terms). Includes all credit card and debt consolidation loans. Heteroscedasticity-consistent standard errors, double-clustered by sub-grade and issue month, are shown in parentheses.

Panel A: Misreporting Index

	(1)	(2)	(3)	(4)
Social capital (SK)	−0.0486***			−0.0478***
	(0.0038)			(0.0048)
Honesty		−0.2053***		−0.0880***
		(0.0105)		(0.0097)
Income volatility			1.6423***	1.5312***
			(0.0510)	(0.0590)
Controls	Yes	Yes	Yes	Yes
Sub-grade x Year FE	Yes	Yes	Yes	Yes
Purpose FE	Yes	Yes	Yes	Yes
State FE	Yes	No	No	Yes
3-digit zip FE	No	Yes	Yes	No
Year-month x Term FE	Yes	Yes	Yes	Yes
N	1,569,763	401,350	723,134	391,166
R^2	0.096	0.110	0.098	0.109

Panel B: Suspect amount

	(1)	(2)	(3)	(4)
Social capital (SK)	−0.0014			−0.0035
	(0.0012)			(0.0021)
Honesty		−0.0205***		−0.0159***
		(0.0028)		(0.0031)
Income volatility			0.0770***	0.0561***
			(0.0076)	(0.0073)
Controls	Yes	Yes	Yes	Yes
Sub-grade x Year FE	Yes	Yes	Yes	Yes
Purpose FE	Yes	Yes	Yes	Yes
State FE	Yes	No	No	Yes
3-digit zip FE	No	Yes	Yes	No
Year-month x Term FE	Yes	Yes	Yes	Yes
N	1,569,763	401,350	723,134	391,166
R^2	0.024	0.027	0.027	0.025

Significance levels: * 0.1, ** 0.05, * 0.01. Standard errors in parentheses.**

Panel C: Loan amount divisible

	Amount div. 5,000			Amount div. 10,000		
	(1)	(2)	(3)	(4)	(5)	(6)
Social capital (SK)	−0.0089***			−0.0059***		
	(0.0012)			(0.0009)		
Honesty		−0.0084**			0.0017	
		(0.0040)			(0.0030)	
Income volatility			0.1563***			0.0429**
			(0.0178)			(0.0189)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Sub-grade x Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Purpose FE	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	No	No	Yes	No	No
3-digit zip FE	No	Yes	Yes	No	Yes	Yes
Year-month x Term FE	Yes	Yes	Yes	Yes	Yes	Yes
N	1,569,763	401,350	723,134	1,569,763	401,350	723,134
R^2	0.093	0.103	0.096	0.046	0.046	0.046

Panel D: Annual income divisible

	Income div. 5,000			Income div. 10,000		
	(1)	(2)	(3)	(4)	(5)	(6)
Social capital (SK)	−0.0185***			−0.0138***		
	(0.0013)			(0.0013)		
Honesty		−0.1135***			−0.0646***	
		(0.0040)			(0.0029)	
Income volatility			0.7713***			0.5948***
			(0.0203)			(0.0198)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Sub-grade x Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Purpose FE	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	No	No	Yes	No	No
3-digit zip FE	No	Yes	Yes	No	Yes	Yes
Year-month x Term FE	Yes	Yes	Yes	Yes	Yes	Yes
N	1,569,763	401,350	723,134	1,569,763	401,350	723,134
R^2	0.060	0.069	0.059	0.043	0.056	0.046

Significance levels: * 0.1, ** 0.05, * 0.01. Standard errors in parentheses.**

Table 6
Misreporting vs. borrower income and credit grade

The dependent variable is *Misreporting Index*. *Post* is a dummy taking the value one if the loan is issued after Trump inauguration. Control variables are the same as in Table 3. Includes all credit card and debt consolidation loans. Heteroscedasticity-consistent standard errors, double-clustered by sub-grade and issue month, are shown in parentheses.

Panel A: By borrower income

	(1)	(2)
Post x ln(Income)	−0.0356**	−0.0345**
	(0.0140)	(0.0147)
Post	0.5558***	
	(0.1669)	
ln(Annual income)	0.5452***	0.5347***
	(0.0104)	(0.0099)
Controls	Yes	Yes
Sub-grade FE	Yes	No
Sub-grade x Year FE	No	Yes
Year-month x Term FE	No	Yes
Purpose FE	Yes	Yes
3-digit zip FE	Yes	Yes
N	1,578,919	1,578,918
R^2	0.093	0.098

Panel B: By credit grade

	(1)	(2)	(3)
Post x Low grade	0.0952***	0.0874***	0.0324***
	(0.0127)	(0.0128)	(0.0067)
Post	0.1414***	0.1427***	
	(0.0231)	(0.0232)	
Low grade	−0.1186***		
	(0.0179)		
Controls	Yes	Yes	Yes
Sub-grade FE	No	Yes	No
Sub-grade x Year FE	No	No	Yes
Year-month x Term FE	No	No	Yes
Purpose FE	Yes	Yes	Yes
3-digit zip FE	Yes	Yes	Yes
N	1,578,919	1,578,919	1,578,918
R^2	0.093	0.093	0.098

Significance levels: * 0.1, ** 0.05, * 0.01. Standard errors in parentheses.**

Table 7
Index components

The dependent variable in Panel A is *Default*, a dummy taking the value one if the loan defaulted, in Panel B*Interest*, the annual interest rate of the loan, and in Panel C *Loss given default*, the proportion of loan principal lost following default. Control variables are the same as in Table 3. Includes all non-current credit card and debt consolidation loans. Heteroscedasticity-consistent standard errors, double-clustered by sub-grade and issue month, are shown in parentheses.

Panel A: Likelihood of default

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Suspect amount	0.0221***						0.0220***	
	(0.0016)						(0.0016)	
Suspect amt. above		0.0195***						0.0193***
		(0.0013)						(0.0013)
Suspect amt. below		0.0248***						0.0249***
		(0.0025)						(0.0025)
Income div. 5,000			0.0140***				0.0118***	0.0118***
			(0.0009)				(0.0009)	(0.0009)
Income div. 10,000				0.0118***			0.0033***	0.0033***
				(0.0011)			(0.0012)	(0.0012)
Loan amt. div. 5,000					0.0084***		0.0044**	0.0044**
					(0.0017)		(0.0021)	(0.0021)
Loan amt. div. 10,000						0.0107***	0.0068***	0.0068***
						(0.0013)	(0.0016)	(0.0016)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sub-grade x Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Purpose FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
3-digit zip FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-month x Term FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	903,038	903,038	903,038	903,038	903,038	903,038	903,038	903,038
R^2	0.109	0.109	0.109	0.109	0.109	0.109	0.109	0.109

Significance levels: * 0.1, ** 0.05, * 0.01. Standard errors in parentheses.**

Panel B: Interest

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Suspect amount	−0.0000						−0.0000	
	(0.0010)						(0.0010)	
Suspect amt. above		−0.0009						−0.0009
		(0.0014)						(0.0014)
Suspect amt. below		0.0009						0.0009
		(0.0013)						(0.0013)
Income div. 5,000			−0.0020***				−0.0028***	−0.0028***
			(0.0006)				(0.0009)	(0.0009)
Income div. 10,000				−0.0005			0.0014	0.0014
				(0.0008)			(0.0011)	(0.0011)
Loan amt. div. 5,000					0.0005		0.0013	0.0013
					(0.0009)		(0.0013)	(0.0013)
Loan amt. div. 10,000						−0.0004	−0.0014	−0.0014
						(0.0010)	(0.0015)	(0.0015)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sub-grade x Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Purpose FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
3-digit zip FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-month x Term FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	903,038	903,038	903,038	903,038	903,038	903,038	903,038	903,038
R^2	0.996	0.996	0.996	0.996	0.996	0.996	0.996	0.996

Significance levels: * 0.1, ** 0.05, * 0.01. Standard errors in parentheses.**

Panel C: Loss given default

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Suspect amount	0.0091***						0.0090***	
	(0.0014)						(0.0014)	
Suspect amt. above		0.0031*						0.0030*
		(0.0016)						(0.0016)
Suspect amt. below		0.0157***						0.0157***
		(0.0013)						(0.0013)
Income div. 5,000			0.0033***				0.0034***	0.0034***
			(0.0009)				(0.0010)	(0.0010)
Income div. 10,000				0.0022**			−0.0001	−0.0001
				(0.0009)			(0.0011)	(0.0011)
Loan amt. div. 5,000					−0.0012		−0.0015	−0.0015
					(0.0011)		(0.0013)	(0.0013)
Loan amt. div. 10,000						−0.0006	0.0004	0.0005
						(0.0012)	(0.0015)	(0.0015)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sub-grade x Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Purpose FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
3-digit zip FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-month x Term FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	182,651	182,651	182,651	182,651	182,651	182,651	182,651	182,651
R^2	0.310	0.311	0.310	0.310	0.310	0.310	0.310	0.311

Significance levels: * 0.1, ** 0.05, * 0.01. Standard errors in parentheses.**

Table 8
Likelihood of default – flexible functional forms

The dependent variable, *Default*, is a dummy taking the value one if the loan defaulted. Control variables are the same as in Table 3, except that the models include first to sixth degree polynomial terms for the variables *Loan/rev. balance*, *Annual income*, and *Loan amount*, as indicated. Includes all non-current credit card and debt consolidation loans. Heteroscedasticity-consistent standard errors, double-clustered by sub-grade and issue month, are shown in parentheses.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Misreporting Index	0.0082***						
	(0.0006)						
Suspect amount		0.0221***					
		(0.0016)					
Suspect amt. above			0.0193***				
			(0.0012)				
Suspect amt. below			0.0250***				
			(0.0025)				
Income div. 5,000				0.0137***			
				(0.0009)			
Income div. 10,000					0.0116***		
					(0.0011)		
Loan amt. div. 5,000						0.0104***	
						(0.0016)	
Loan amt. div. 10,000							0.0096***
							(0.0015)
Loan/oustanding balance: 1st - 6th degree terms	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Annual income: 1st - 6th degree terms	Yes	No	No	Yes	Yes	No	No
Loan amount: 1st - 6th degree terms	Yes	No	No	No	No	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sub-grade x Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Purpose FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
3-digit zip FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-month x Term FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	903,038	903,038	903,038	903,038	903,038	903,038	903,038
R^2	0.109	0.109	0.109	0.109	0.109	0.109	0.109

Significance levels: * 0.1, ** 0.05, * 0.01. Standard errors in parentheses.**

Table 9
Likelihood of default – narrow bands around thresholds

The dependent variable, *Default*, is a dummy taking the value one if the loan defaulted. Control variables are the same as in Table 3. The bandwidth of the variable around the index component threshold is shown above each model. Includes all non-current credit card and debt consolidation loans. Heteroscedasticity-consistent standard errors, double-clustered by sub-grade and issue month, are shown in parentheses.

	Loan/rev. balance	Annual income		Loan amount	
	(1)	(2)	(3)	(4)	(5)
	± 0.2	$\pm 1,000$	$\pm 1,000$	$\pm 1,000$	$\pm 2,000$
Suspect amount	0.0043***				
	(0.0011)				
Income div. 5,000		0.0065***			
		(0.0006)			
Income div. 10,000			0.0078***		
			(0.0013)		
Loan amt. div. 5,000				0.0037***	
				(0.0011)	
Loan amt. div. 10,000					0.0064***
					(0.0016)
Controls	Yes	Yes	Yes	Yes	Yes
Round threshold dummies	No	Yes	Yes	Yes	Yes
Sub-grade x Year FE	Yes	Yes	Yes	Yes	Yes
Purpose FE	Yes	Yes	Yes	Yes	Yes
3-digit zip FE	Yes	Yes	Yes	Yes	Yes
Year-month x Term FE	Yes	Yes	Yes	Yes	Yes
N	639,958	1,090,780	594,935	978,273	518,989
R^2	0.092	0.098	0.099	0.096	0.101

Significance levels: * 0.1, ** 0.05, * 0.01. Standard errors in parentheses.**

Table 10
Income roundness and verification status

The dependent variable, *Default*, is a dummy taking the value one if the loan defaulted. I include *Sub-grade x Year fixed effects* based on the credit grade assigned by Lending Club (35 grades times 12 years), *Purpose fixed effects* (credit card or debt consolidation), *State fixed effects* or *3-digit zip fixed effects* (955 zip codes) based on the address of the borrower, and *Year-month x Term fixed effects* (133 months times two alternative terms). Includes all non-current credit card and debt consolidation loans. Heteroscedasticity-consistent standard errors, double-clustered by sub-grade and issue month, are shown in parentheses.

Panel A: All loans

	(1)	(2)	(3)
Income div. 5,000	0.0140***		0.0120***
	(0.0009)		(0.0010)
Income div. 10,000		0.0118***	0.0034***
		(0.0011)	(0.0012)
Controls	Yes	Yes	Yes
Sub-grade x Year FE	Yes	Yes	Yes
Purpose FE	Yes	Yes	Yes
3-digit zip FE	Yes	Yes	Yes
Year-month x Term FE	Yes	Yes	Yes
N	903,038	903,038	903,038
R^2	0.109	0.109	0.109

Panel B: Income verified

	(1)	(2)	(3)
Income div. 5,000	0.0116***		0.0074***
	(0.0010)		(0.0016)
Income div. 10,000		0.0125***	0.0071***
		(0.0016)	(0.0023)
Controls	Yes	Yes	Yes
Sub-grade x Year FE	Yes	Yes	Yes
Purpose FE	Yes	Yes	Yes
3-digit zip FE	Yes	Yes	Yes
Year-month x Term FE	Yes	Yes	Yes
N	288,601	288,601	288,601
R^2	0.115	0.115	0.115

Significance levels: * 0.1, ** 0.05, * 0.01. Standard errors in parentheses.**

Panel C: Income not verified

	(1)	(2)	(3)
Income div. 5,000	0.0154***		0.0144***
	(0.0013)		(0.0013)
Income div. 10,000		0.0116***	0.0017*
		(0.0012)	(0.0009)
Controls	Yes	Yes	Yes
Sub-grade x Year FE	Yes	Yes	Yes
Purpose FE	Yes	Yes	Yes
3-digit zip FE	Yes	Yes	Yes
Year-month x Term FE	Yes	Yes	Yes
N	614,410	614,410	614,410
R^2	0.103	0.103	0.103

Significance levels: * 0.1, ** 0.05, * 0.01. Standard errors in parentheses.**

IA.1 Internet Appendix

Table IA.1
Summary statistics – means by Misreporting Index value

Means of variables classified by Misreporting Index value.

	Suspect Index					
	0	1	2	3	4	5
Loan variables						
Interest (%)	12.707	13.455	12.971	12.897	12.696	12.380
Loan amount ('000)	13.492	13.593	15.938	16.297	20.277	19.357
Term 5 years	0.258	0.286	0.304	0.311	0.334	0.322
Loan / rev. balance	0.999	3.419	3.679	4.507	4.928	6.094
Loan status						
Current performing	0.388	0.373	0.407	0.424	0.461	0.498
Fully paid	0.491	0.481	0.458	0.444	0.410	0.382
Loan late	0.016	0.020	0.020	0.021	0.023	0.022
Default	0.105	0.127	0.115	0.111	0.106	0.098
Outcome (mature only)						
Prepaid	0.598	0.594	0.610	0.611	0.613	0.613
Fully paid	0.225	0.197	0.190	0.190	0.181	0.183
Default	0.177	0.209	0.200	0.200	0.205	0.204
Suspect loans						
Misreporting Index	0.000	1.000	2.000	3.000	4.000	5.000
Suspect amount	0.000	0.902	0.783	0.947	0.891	1.000
Suspect amt. above	0.000	0.502	0.442	0.536	0.512	0.592
Suspect amt. below	0.000	0.400	0.341	0.412	0.380	0.408
Income precision						
Income div. 5,000	0.000	0.062	0.659	0.741	1.000	1.000
Income div. 10,000	0.000	0.000	0.132	0.613	0.635	1.000
Loan amt. div. 5,000	0.000	0.036	0.365	0.418	1.000	1.000
Loan amt. div. 10,000	0.000	0.000	0.062	0.281	0.474	1.000
Borrower						
Annual income ('000)	59.456	64.172	76.823	83.582	101.087	102.157
Debt / income (%)	19.465	20.175	19.120	18.791	18.131	17.775
Employment length (y.)	5.264	5.520	5.636	5.681	5.821	5.781
No employment length	0.080	0.080	0.051	0.048	0.039	0.041
Rev. balance ('000)	13.547	15.011	17.584	18.833	23.809	22.223
Revolving util. (%)	55.708	53.202	53.250	52.724	53.658	52.100
Total accounts	23.143	24.134	24.907	25.047	25.987	25.400
Open accounts	11.365	11.530	11.959	11.985	12.444	12.218
Home owned	0.103	0.104	0.103	0.104	0.108	0.108
Home mortgaged	0.417	0.462	0.492	0.506	0.536	0.535
Honesty	0.423	0.416	0.394	0.387	0.370	0.373
N	93,834	520,909	340,629	416,385	133,715	78,655

Figure 9: Suspect amount - quarter fixed effects

Estimated quarter fixed effects from a regression of Misreporting Index on a full set of control variables on loan, borrower, and location characteristics. Includes all credit card and debt consolidation loans issued on Lending Club from 2010 until Q2 2018.

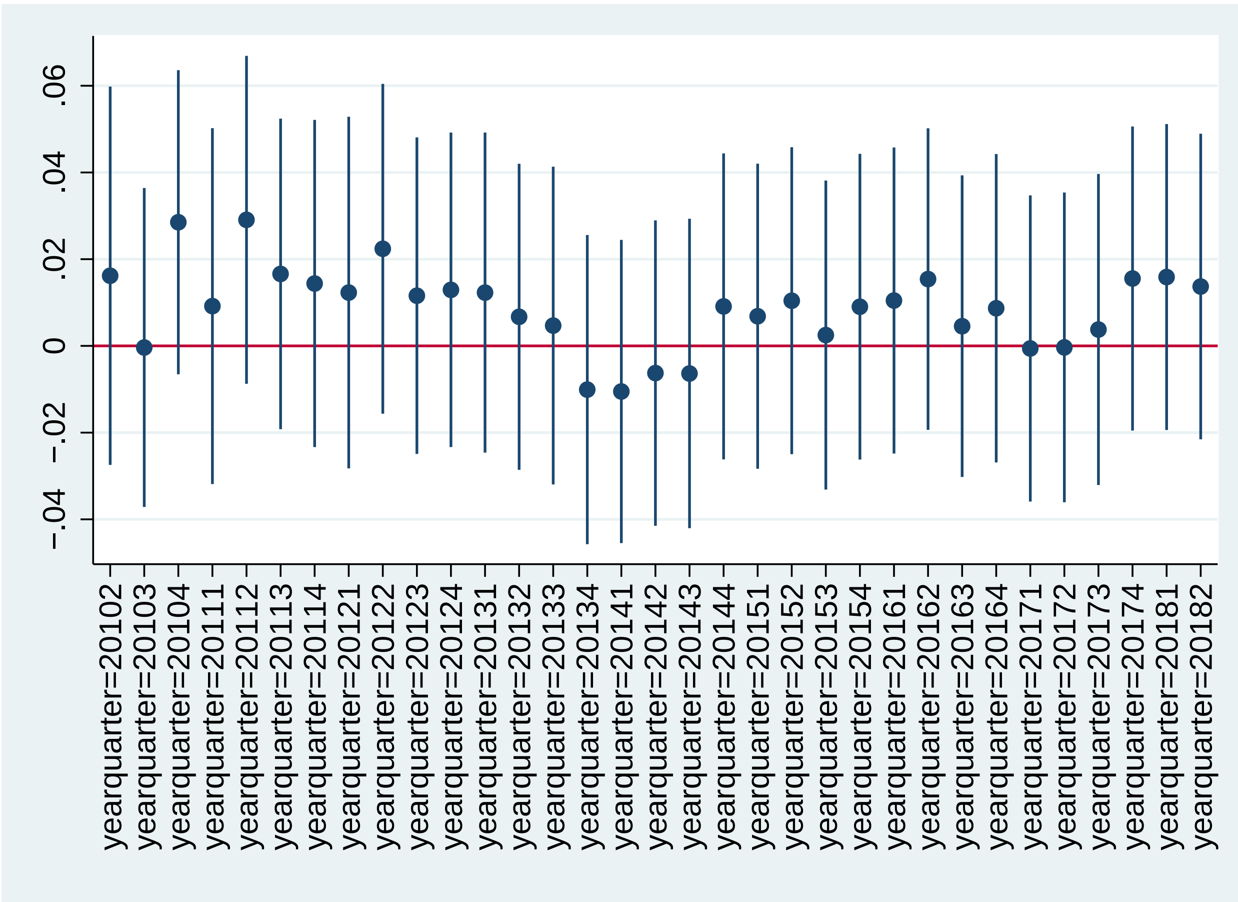


Figure 10: Suspect amount above - quarter fixed effects

Estimated quarter fixed effects from a regression of Misreporting Index on a full set of control variables on loan, borrower, and location characteristics. Includes all credit card and debt consolidation loans issued on Lending Club from 2010 until Q2 2018.

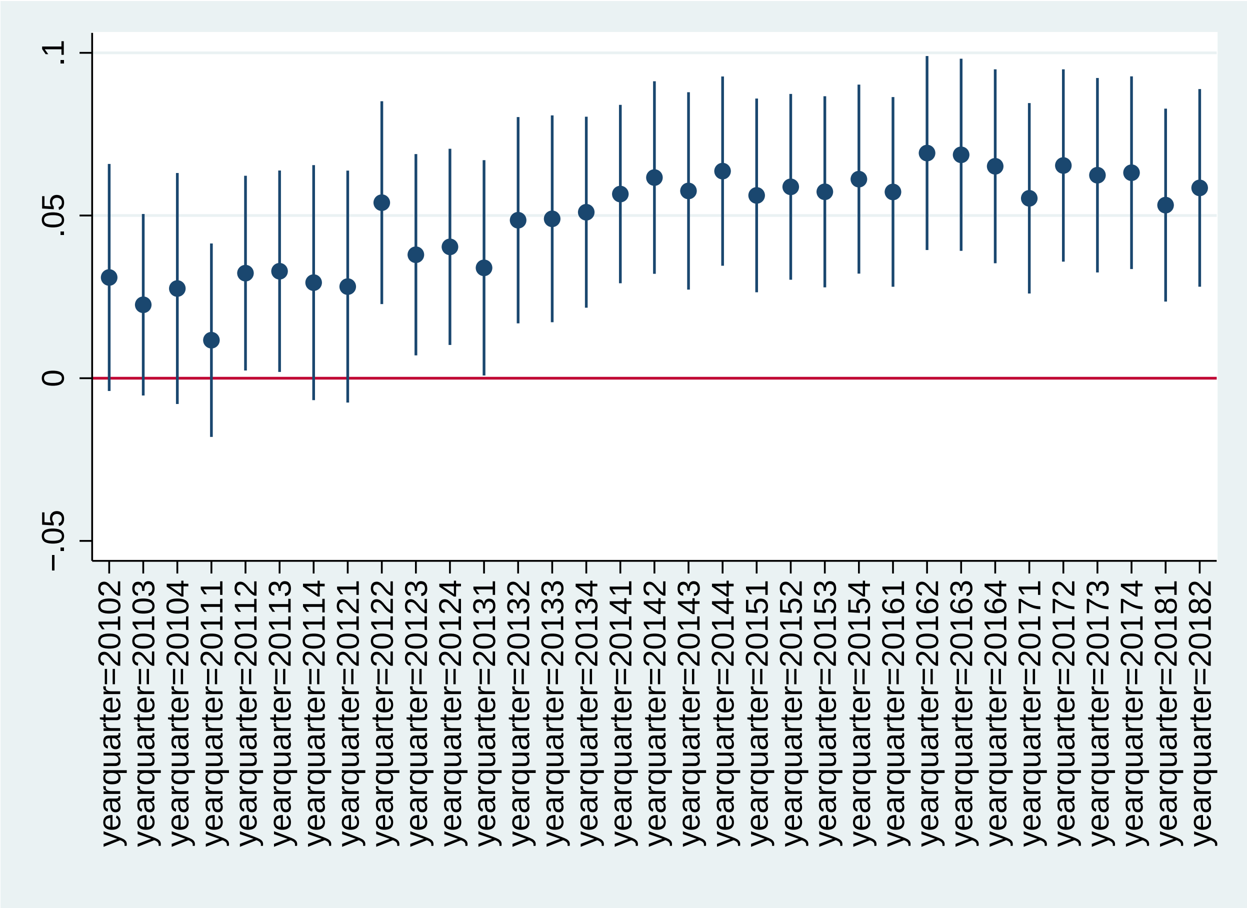


Figure 11: Suspect amount below - quarter fixed effects

Estimated quarter fixed effects from a regression of Misreporting Index on a full set of control variables on loan, borrower, and location characteristics. Includes all credit card and debt consolidation loans issued on Lending Club from 2010 until Q2 2018.

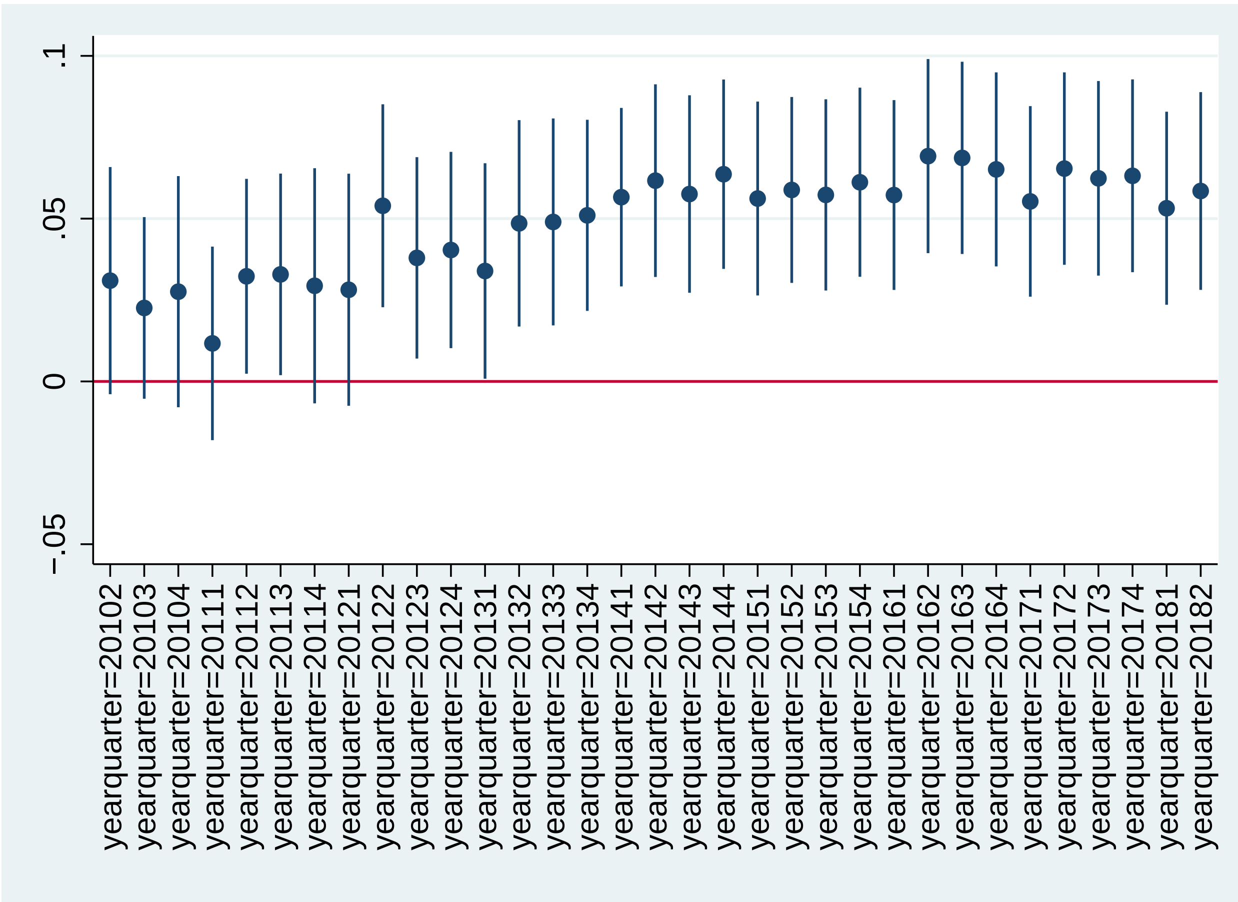


Figure 12: Income divisible 10,000 - quarter fixed effects

Estimated quarter fixed effects from a regression of Misreporting Index on a full set of control variables on loan, borrower, and location characteristics. Includes all credit card and debt consolidation loans issued on Lending Club from 2010 until Q2 2018.

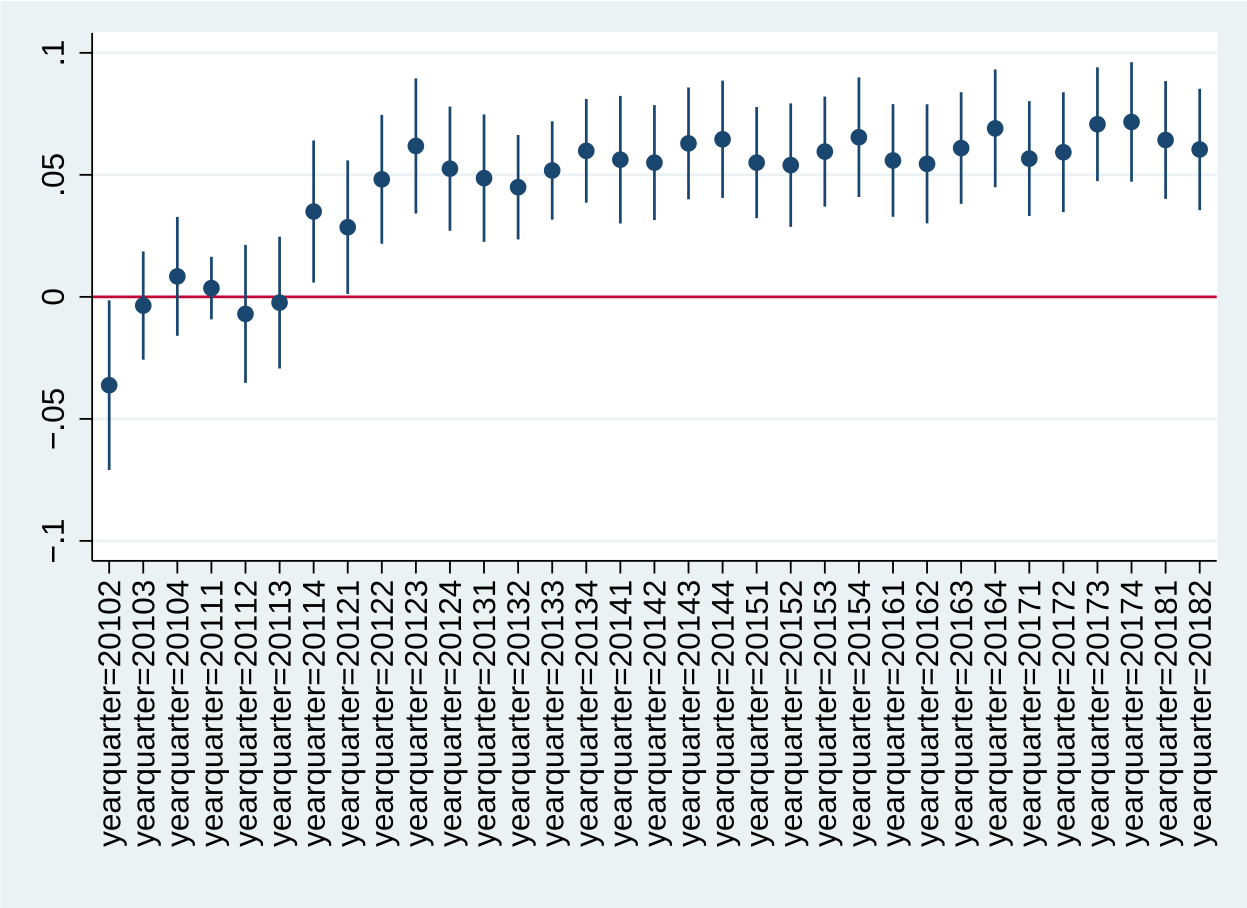


Figure 13: Income divisible 5,000 - quarter fixed effects

Estimated quarter fixed effects from a regression of Misreporting Index on a full set of control variables on loan, borrower, and location characteristics. Includes all credit card and debt consolidation loans issued on Lending Club from 2010 until Q2 2018.

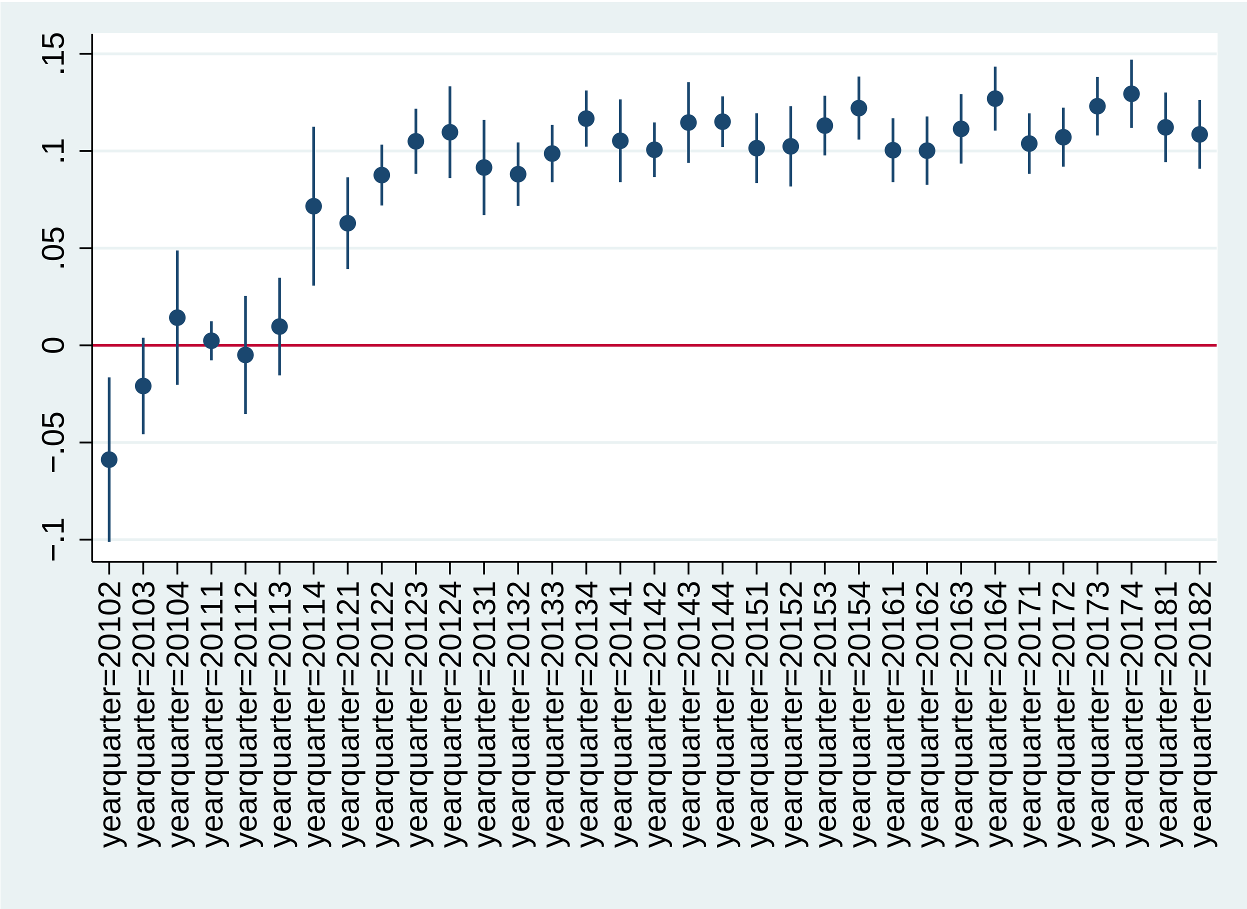


Figure 14: Loan amount divisible 5,000 - quarter fixed effects

Estimated quarter fixed effects from a regression of Misreporting Index on a full set of control variables on loan, borrower, and location characteristics. Includes all credit card and debt consolidation loans issued on Lending Club from 2010 until Q2 2018.

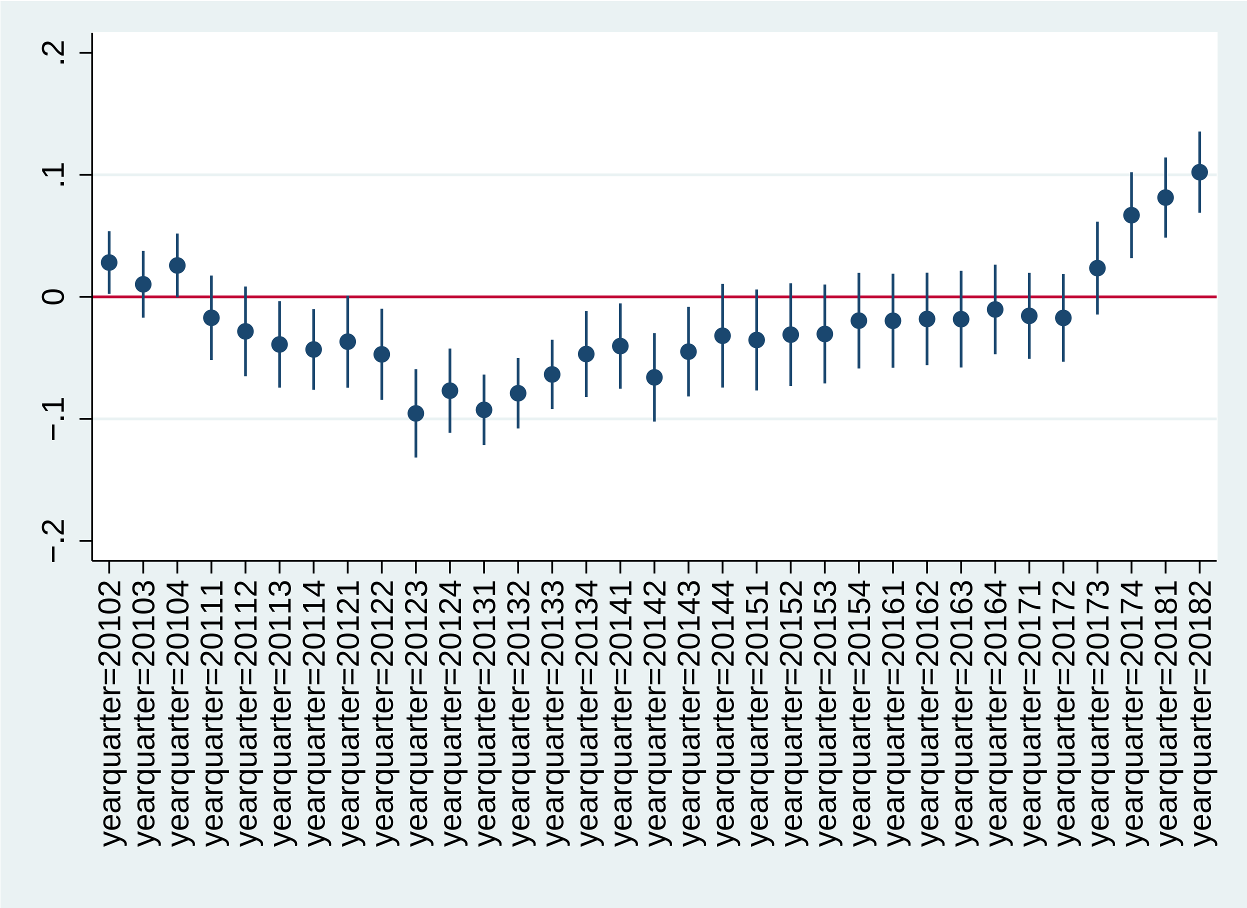


Figure 15: Loan amount divisible 10,000 - quarter fixed effects

Estimated quarter fixed effects from a regression of Misreporting Index on a full set of control variables on loan, borrower, and location characteristics. Includes all credit card and debt consolidation loans issued on Lending Club from 2010 until Q2 2018.

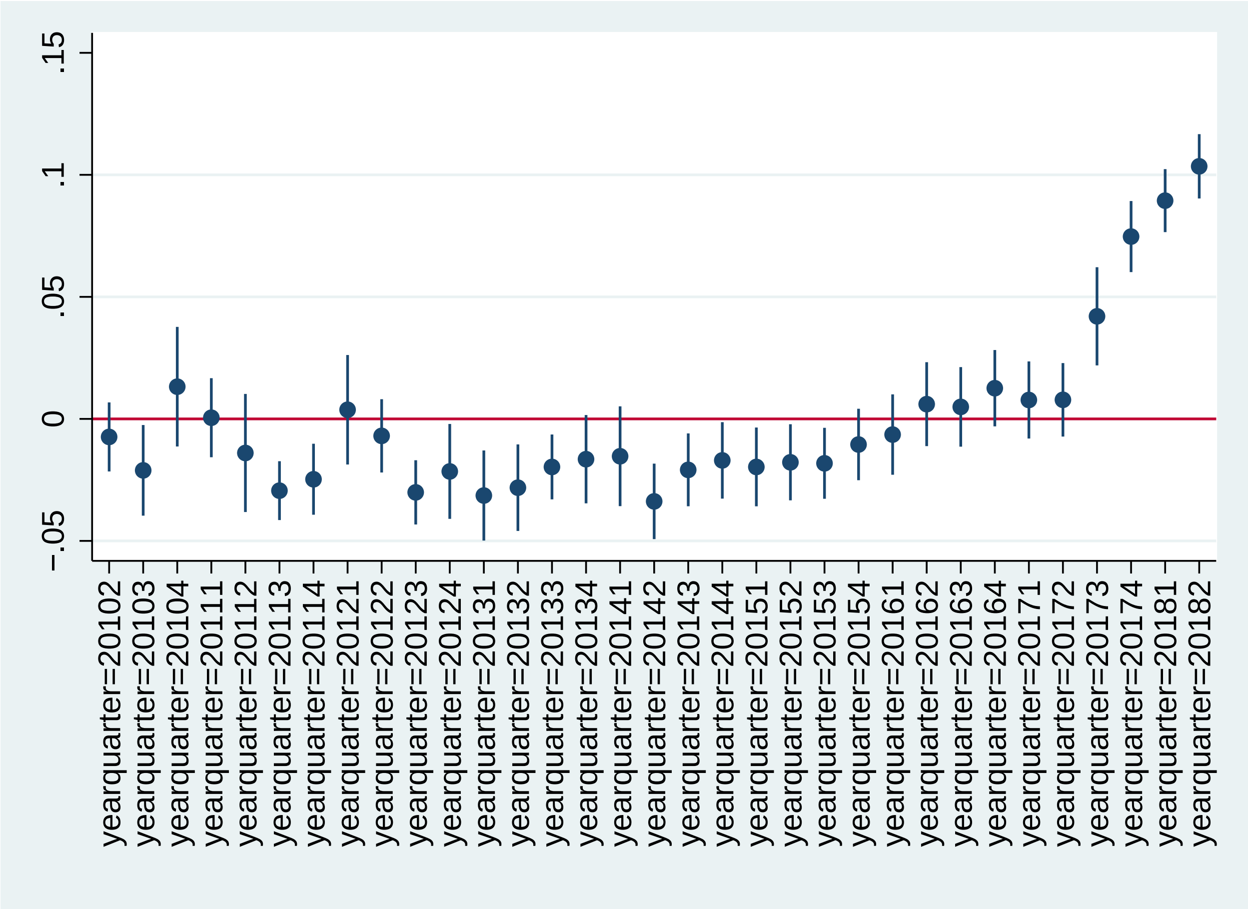


Table IA.2
Robustness check: Loans at least 3 months past maturity only

The dependent variable is shown above each model, *Default*, is a dummy taking the value one if the loan defaulted. Variables are defined in Appendix A. I include *Sub-grade x Year fixed effects* based on the credit grade assigned by Lending Club (35 grades times 12 years), *Purpose fixed effects* (credit card or debt consolidation), *3-digit zip fixed effects* (955 zip codes) based on the address of the borrower, and *Year-month x Term fixed effects* (133 months times two alternative terms). Includes all non-current credit card and debt consolidation loans. Heteroscedasticity-consistent standard errors, double-clustered by sub-grade and issue month, are shown in parentheses.

	(1) Default	(2) Default	(3) Interest
Misreporting Index	0.0053***		−0.0002
	(0.0004)		(0.0004)
Misreporting Index=1		0.0161***	
		(0.0018)	
Misreporting Index=2		0.0184***	
		(0.0026)	
Misreporting Index=3		0.0234***	
		(0.0021)	
Misreporting Index=4		0.0277***	
		(0.0028)	
Misreporting Index=5		0.0365***	
		(0.0035)	
Controls	Yes	Yes	Yes
Sub-grade x Year FE	Yes	Yes	Yes
Purpose FE	Yes	Yes	Yes
3-digit zip FE	Yes	Yes	Yes
Year-month x Term FE	Yes	Yes	Yes
N	320,422	320,422	320,422
R^2	0.056	0.056	0.996

Significance levels: * 0.1, ** 0.05, * 0.01. Standard errors in parentheses.**