FEDERAL RESERVE BANK
PHILADELPHIA

Working Papers
RESEARCH DEPARTMENT

Recurring-Payment Sensitivity in Household Borrowing

Jing Xian Ng
The Wharton School and
Federal Reserve Bank of Philadelphia
Consumer Finance Institute

WP 25-22

PUBLISHED
August 2025

FRB CFI
PHILADELPHIA

ISSN: 1962-5361

Disclaimer: This Philadelphia Fed working paper represents preliminary research that is being circulated for discussion purposes. The views expressed in these papers are solely those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Philadelphia or the Federal Reserve System. Any errors or omissions are the responsibility of the authors. Philadelphia Fed working papers are free to download at: https://www.philadelphiafed.org/search-results/all-work?searchtype=working-papers.

DOI: https://doi.org/10.21799/frbp.wp.2025.22

Recurring-Payment Sensitivity in Household Borrowing

Jing Xian Ng*

July 2025

For the latest version of this paper, please click HERE

Abstract

This paper provides evidence of payment sensitivity in household borrowing decisions: Mortgage borrowers respond to the size of the recurring payment as opposed to discounted total loan costs when choosing between loan options. I develop a test for payment sensitivity that exploits differences in predicted bunching at kinks and notches generated by mortgage insurance requirements. I find that borrowing is substantially more responsive to nominal recurring payments than to the net present value of total costs. To rationalize the result, outside borrowing costs would have to be implausibly high, exceeding 40% a year. Payment sensitivity is the most likely explanation for observed borrowing choices as alternatives require implausible non-mortgage borrowing costs or household preferences. I develop a dynamic consumption-savings model and show that underlying preferences can generate the observed payment sensitivity only if borrowers initially have a high marginal utility of cash-on-hand that coincidentally and sharply falls by more than 50% in a narrow time window after loan origination. Payment sensitivity has important implications for regulation and policy. Lenders can manipulate loan features and shroud increases in total costs from payment sensitive borrowers, even while keeping fixed or even decreasing recurring payments. This type of shrouding could enable excessive borrowing and attenuate the transmission of monetary policy.

*XianNg@wharton.upenn.edu, The Wharton School, University of Pennsylvania and Consumer Finance Institute, Federal Reserve Bank of Philadelphia. The views expressed are those of the author and do not necessarily reflect those of the Federal Reserve Bank of Philadelphia, or the Federal Reserve System. This paper is based on work supported by an Institute of Consumer Money Management graduate student Fellowship on Consumer Financial Management. I would like to thank my dissertation committee, Benjamin Keys, Alex Rees-Jones, and Todd Sinai for their advice on this project. I would also like to thank Robert Hunt, Sasha Indarte, Lauren Lambie-Hanson, Benjamin Lockwood, Patrick Moran, Nikolai Roussanov, and seminar participants at the Federal Reserve Bank of Philadelphia for their insights.

In neoclassical models, borrowing is a tool for intertemporal consumption smoothing. Loans represent a trade-off between the benefit of immediate access to liquidity against the discounted cost of negative cash flow from making repayments. In practice, households often deviate from neoclassical predictions and make suboptimal borrowing and repayment decisions - they do not pay off highest cost debt first (Gathergood et al., 2019), they under-respond to distant payments (Stango and Zinman, 2009), and they fail to treat cash as fungible (Hastings and Shapiro, 2013). This paper presents new evidence of a systematic deviation from standard models with potentially large individual welfare and macroeconomic consequences, in the high-stakes, low-frequency mortgage borrowing setting: household borrowing is sensitive to the size of nominal recurring loan payments rather than to the discounted total cost of borrowing.[1]

The paper proposes a model of *payment sensitivity* to explain household borrowing choices. Payment sensitive borrowers are responsive to the size of recurring loan payments, as opposed to neoclassical borrowers who are responsive to the total cost of loan repayments. I discern between these two models of behavior by exploiting changes in mortgage insurance (MI) requirements at loan-to-value ratio (LTV) thresholds. The pattern of bunching and missing mass in borrower mortgage LTV choices, in response to kink and notch discontinuities generated by MI, is best explained by payment sensitivity and is inconsistent with neoclassical models. I develop a dynamic consumption-savings model to test mechanisms for payment sensitivity and find that neither liquidity constraints nor standard formulations of present-bias can explain the bunching result. In addition, any preference-based explanation requires an unusual pattern of utility weights, with high weight on initial and early periods that declines sharply at a specific time. Finally, I find potentially large macroeconomic consequences of payment sensitivity: I estimate that the growth of interest-only mortgages in the early 2000s housing expansion enabled payment sensitive borrowers to take out up to an additional $91 billion a year in loans during a rate-hiking cycle.

I contrast the model of sensitivity to nominal, recurring loan payments with a model of sensitivity to the net present value (NPV) of loan payments. The NPV of loan payments is a useful comparative benchmark as it distills the cost of reduced future consumption from loan payments into a single statistic. In fact, any forward-looking borrower with some ability and desire to smooth consumption should be somewhat sensitive to the NPV of loan payments when choosing between loan options. The baseline results show that payment sensitivity better explains empirical mortgage borrowing choices than a model of NPV sensitivity. In secondary analyses, alternative explanations for observed borrower behavior besides NPV sensitivity also fail to explain the empirical result;

[1] Mortgages are typically households' largest single borrowing decision. Outstanding mortgage debt is $13 trillion, representing 72% of total household debt balances (Federal Reserve Bank of New York, 2023).

instead, payment sensitivity remains a better model of borrower choice in the mortgage choice setting.

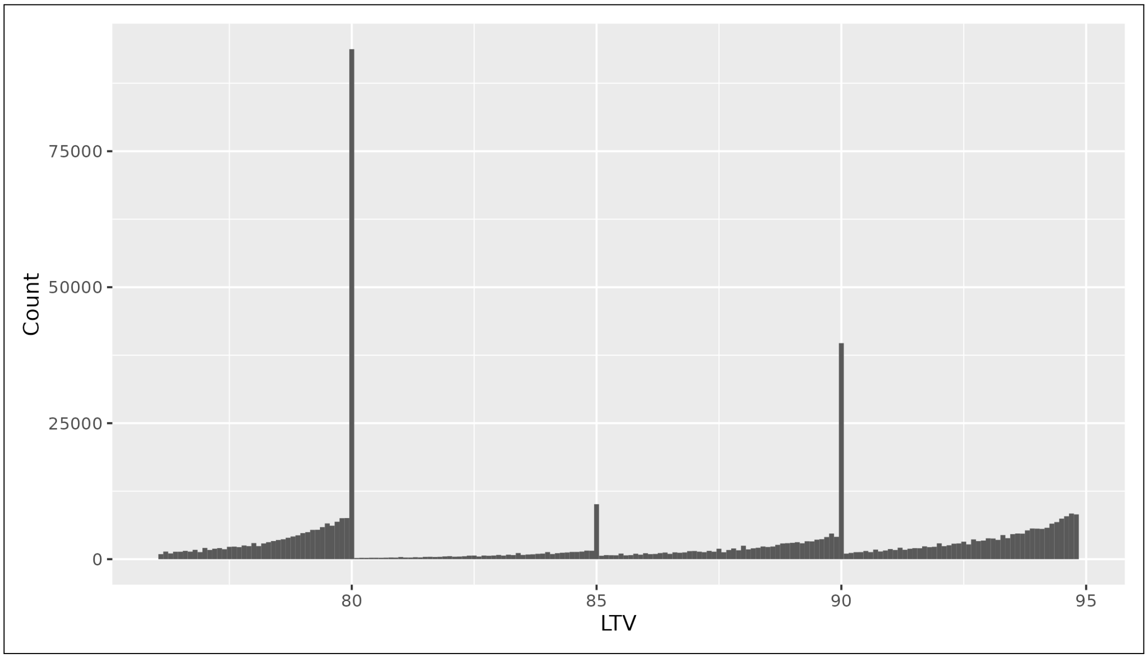
I develop a test for payment sensitivity by exploiting discontinuous mortgage insurance (MI) requirements at loan-to-value ratio (LTV) thresholds. The requirements affect the recurring payments and the NPV of payments differently, at two different LTV thresholds. MI modifies the schedule of recurring payments against initial LTV choice by introducing level discontinuities (notches) at 80% and 90% LTV, with a much larger notch at 80% compared to 90%. Simultaneously, MI modifies the schedule of NPV of payments against initial LTV by introducing a slope discontinuity (kink) at 80% LTV and a notch at 90%.

The bunching response to these MI-induced discontinuities distinguishes between models of borrower behavior. Borrowers responding to the recurring payment face a large notch in their budget constraint at 80% LTV and a smaller notch at 90% LTV. Per Kleven and Waseem (2013), notched budget constraints induce bunching at the notch point and missing mass above the notch point. As such, payment sensitive borrowers should bunch at both 80% and 90%, but there should be more missing mass just above 80% LTV. In contrast, borrowers responding to the NPV of payments face a kink in their budget constraint at 80% LTV, and a notch at 90% LTV. Per Saez (2010) and Chetty et al. (2011), kinks induce bunching at the threshold with no missing mass above. NPV sensitive borrowers would thus exhibit an *opposite pattern* to payment sensitive borrowers, with more missing mass above 90% LTV than 80% LTV. As such, the missing mass above the LTV thresholds is the key statistic used to test for payment sensitivity.

Borrowers respond to MI requirements by bunching in their mortgage LTV choices in a manner consistent with payment sensitivity. Figure 1 shows the empirical distribution of borrowers' LTV choices, in the presence of distortions caused by MI at LTV multiples of five.[2] Focusing on the 80% and 90% LTV thresholds, borrowers bunch at both thresholds, but there is more missing mass of borrowers above 80% LTV than above the 90% LTV threshold, relative to plausible, undistorted, counterfactual distributions of LTV choices. This bunching pattern of more missing mass above 80% LTV is consistent with payment sensitive borrowers responding comparatively more to the larger notch in recurring payments at the 80% threshold. Under the preferred counterfactual estimate there is 2.5 times more missing mass above 80% than 90% LTV; this result is robust to a wide range of model specifications.

[2] All borrower and loan data are from a merge of Home Mortgage Disclosure Act (HMDA) and Equifax Credit Risk Insight Servicing McDash (CRISM) data from Equifax and ICE, McDash. Section 1.2 describes the data and the sample.

Figure 1: Distribution of Borrower Loan-to-Value Ratio at Origination



Note: The figure shows the count of mortgage borrowers by initial loan-to-value ratio in 0.1 point bins. The makeup of the sample is described in Section 1.2.

This differences-in-bunching approach isolates the effect of changes in recurring payments at the LTV thresholds, unconfounded by changes in other incentives such as the NPV of payments. The method differences out contamination from all other effects at the thresholds, such as other pricing discontinuities, round-number bunching, or selection of types across the thresholds, as long as these effects are of similar size at both the 80% and 90% thresholds.

The paper then considers other potential explanations for the observed pattern of bunching at LTV thresholds and shows that they are empirically or theoretically unlikely. First, I examine down payment constraints or frictions in adjusting LTV. For frictions to explain the observed pattern of bunching, borrowers would need to face adjustment frictions exceeding the implicit cost of locating above the LTV threshold and these costs would need to be differentially higher for borrowers around the 90% threshold. However, the effective interest rate when locating above the LTV threshold is very high, exceeding the prevailing rate on unsecured credit card borrowing. In addition, borrowers are relatively high income and have ample access to credit, making it unlikely for credit constraints to explain observed bunching. Finally, the pattern of selection across the thresholds shows no evidence of higher frictions at the 90% threshold. These results make it highly unlikely for frictions alone to rationalize observed bunching and bolster the case for payment sensitivity.

Second, I address a broad range of preference-based or behavioral explanations for the bunching pattern using a dynamic consumption-savings model. The observed pattern of bunching could be rationalized with NPV sensitivity if borrowers have substantially higher marginal utility in near periods compared to distant periods, or high time preference. The model reveals that preference-based or behavioral explanations for the bunching pattern must place very high utility weight on initial and some early periods, but the utility weight for future periods must decline precipitously within a narrow window and at a precise time. This rules out liquidity constraints, standard quasi-hyperbolic discounting, and income growth as explanations for the observed bunching pattern.

To illustrate the potential impact of payment sensitivity, I consider the US housing bubble in the early to mid 2000s and the interaction of payment sensitivity with interest-only (IO) mortgages. Substitution from amortizing loans to IO loans allowed payment sensitive borrowers to borrow more at the same monthly payment despite being more costly in NPV terms. I estimate that the growth of IO loans during the rate-hiking cycle enabled up to $91.2 billion per year of additional mortgage lending, representing more than 3% of the entire mortgage origination market.

This paper extends several areas of research: consumer financial behavior, the estimation of mortgage borrowing demand elasticities, and the use of bunching estimators. First, this paper provides strong evidence for payment sensitivity in household financial decision-making over high-stakes, infrequent choices. This result builds on findings from Juster (1964) which provides survey evidence for monthly payment sensitivity and Argyle et al. (2020) which finds that borrowers tend to adjust maturities on auto loans to target monthly payments. The paper also adds to a literature finding heuristics and biases in mortgage decisions, including Campbell (2006), Keys et al. (2016), Agarwal et al. (2017), and Bäckman et al. (2024). In particular, Bäckman et al. (2024) is the most closely related paper, finding that households adjust mortgage borrowing to avoid flow disutility from higher amortization payments. This paper uses a distinct methodology to develop a clean test for payment sensitivity, and establishes quantitative bounds on preference-based explanations for payment sensitivity. The finding of payment sensitivity fits in a broader literature on deviations from neoclassical predictions in household finance.[3]

This paper also extends the line of research estimating mortgage borrowing demand elasticities, including Follain and Dunsky (1997), Jappelli and Pistaferri (2007), DeFusco and Paciorek (2017), Hanson (2020), Best et al. (2020), and Fuster and Zafar (2021). A number of these papers use the bunching at loan-to-value ratio (LTV) thresholds in response to pricing discontinuities to estimate the response. This paper uses a similar bunching method and recovers some estimates of the bor-

[3]For an overview of the broad range of biases and behavioral phenomena observed in household financial decisions, see Beshears et al. (2018) and Stango and Zinman (2023).

rowing response, but is not primarily concerned with the demand elasticity of borrowing. Instead, the main innovation is to use the differences in borrowing responses to distinguish between models of borrower behavior.

Finally, I provide a methodological contribution building on the bunching estimators developed in Saez (2010), Chetty et al. (2011), Kleven and Waseem (2013), and Kleven (2016). These estimators have typically been used to estimate elasticities such as labor supply (with respect to post-tax income) or mortgage borrowing demand (with respect to the interest rate), where the determinants of utility are known (post-tax income, liquid cash). I propose a simple test that makes use of predicted bunching responses to kinks and notches in different quantities at the same thresholds, to distinguish between potential explanations for observed behavior in a setting where agents' motivations are not well known.

The rest of the paper is organized as follows: Section 1 provides an overview of pricing determinants in residential mortgage lending and describes the data. Section 2 describes the effect of mortgage insurance requirements on borrowing costs and payments, and develops a test to distinguish between the NPV sensitivity and payment sensitivity hypotheses using the empirical bunching missing mass. Section 3 describes the bunching counterfactual estimation procedure and results. Section 4 interprets the bunching results as evidence for payment sensitivity, and shows that frictions or selection effects do not rationalize the bunching result with NPV sensitivity. Section 5 develops the consumption-savings model and establishes conditions for preferences that can generate payment sensitivity in the MI setting. Section 6 presents a rough estimate of the potential impact of payment sensitivity by studying the expansion of interest-only mortgages during the early 2000s US housing expansion. Section 7 concludes.

1 Setting and Data

This section gives an overview of mortgage lending in the US, provides detail on mortgage insurance requirements and pricing, and describes the data used.

1.1 Mortgage Pricing and Private Mortgage Insurance

In the United States, the government-sponsored enterprises (GSEs) Fannie Mae and Freddie Mac facilitate a highly liquid market for residential mortgages and mortgage-backed-securities (MBS). More than $600 billion a year of mortgages, representing nearly 45% of all originations, are purchased by the GSEs (Urban Institute, 2024). The GSEs publish highly detailed selling guides, describing eligibility criteria and required price adjustments for “conforming loans,” i.e., loans eligible for sale to the GSEs. Many mortgage lenders operate partially or wholly on an “originate-to-distribute” model and intend to sell loans to the GSEs shortly after origination. As such, these lenders adhere strictly to GSE criteria and price adjustments when pricing loans.

The GSEs impose a mortgage insurance (MI) requirement for loans based on initial loan-to-value ratio (LTV). Borrowers who exceed 80% LTV at origination are required to pay for a MI policy that protects the lender in the case of default; MI is a pure cost with no direct benefit to the borrower.[4] Required MI premia are a step-function of initial LTV, with further rate increases for borrowers exceeding the 85%, 90%, and 95% LTV thresholds. I estimate the MI rates that borrowers of various credit scores face using the method in Bhutta and Keys (2022) and present them in Figure 2. The annual cost of MI ranges from 0.4% to 1.2% of the initial loan amount per year depending on borrower credit score and initial LTV. The initial rate step-up above 80% in the left panel is the largest at about 40-60 basis points per year, with smaller subsequent step-ups of about 10bps at higher LTVs.

An important characteristic of MI is that the rate at origination applies for the entire duration that MI is required. Mortgage insurance is required only for a limited time and thus acts as a *temporary* interest rate increase. Typically, MI premia must be paid until borrowers build up enough home equity for their remaining LTV to fall below 80%.[5] As such, for fixed-term amortizing loans the number of payments required to bring the remaining LTV below 80% and eliminate MI payments depends on the amount by which initial LTV exceeds 80%.

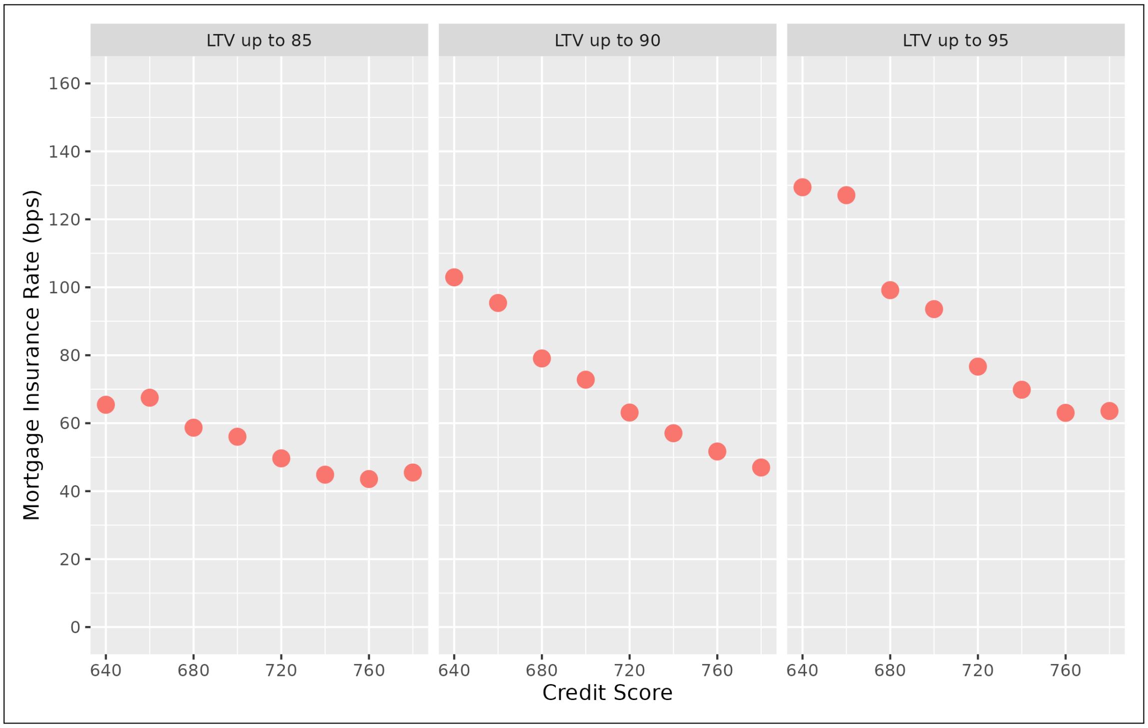
The step-function pricing of MI combined with fact that MI terminates when remaining LTV falls below 80% induces discontinuities in the relationship between initial LTV and required mortgage payments. I focus on two quantities: the initial monthly payment and the net present value (NPV) of payments. The initial monthly payment refers to the size of the first recurring mortgage payment, comprising principal and interest (PI) as well as any MI premia. The NPV of payments refers to the value of the entire stream of expected mortgage and MI payments, discounted to the present.

The effect of MI requirements on payments is illustrated in Figure 3. In panel (a), the bottom-most solid line represents scheduled payments for a typical borrower located at 80% LTV purchasing

[4]Acquiring MI coverage is a seamless process that typically requires no borrower input. Lenders use automated underwriting software that automatically attaches an MI policy to loans that require one, and approval is virtually guaranteed. Borrowers have no ability to shop around for pricing on the MI policy.

[5]Per the Homeowners Protection Act of 1999, owner-occupiers of single family homes may request termination of MI when the remaining loan balance falls below 80% of the original property value. Additionally, loan servicers must automatically terminate MI after the remaining loan balance falls below 78% of the original property value.

Figure 2: Estimated Mortgage Insurance Premia

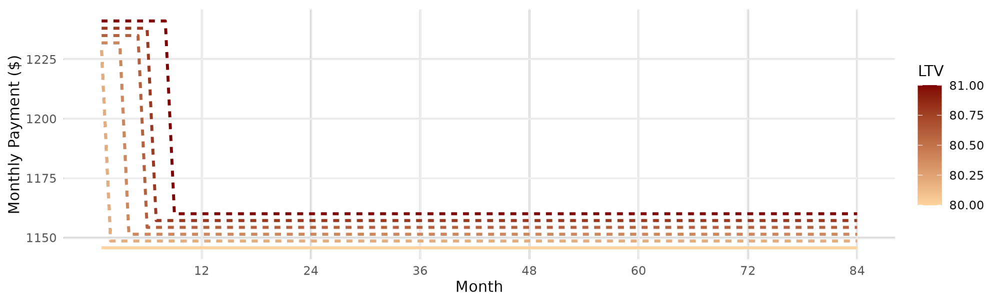


Note: The figure shows the estimated average mortgage insurance premium faced by borrowers in various 5-point LTV bins and 20-point credit score bins using the method from Bhutta and Keys (2022).

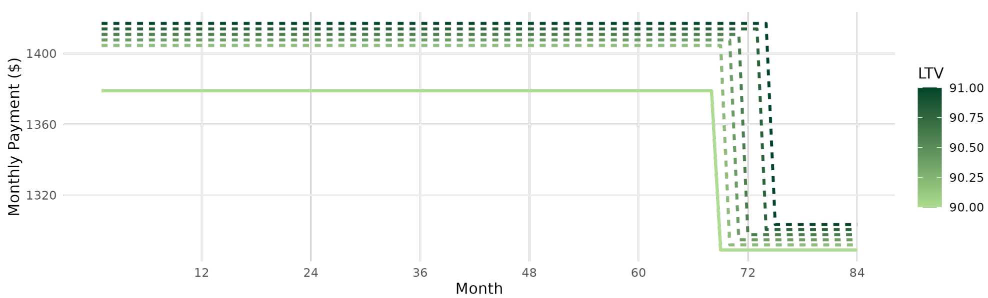
a \$300,000 house with an 80% LTV, fully amortizing fixed-rate loan. Since there is no MI at exactly 80%, the payments are constant throughout the lifetime of the loan. The darker dotted lines represent the payments at month t for the same borrower if they had instead chosen marginally higher initial LTVs. Relative to the 80% LTV loan, these larger loans have MI premia added to the monthly payment and have *discontinuously* higher total initial monthly payments. However, since the remaining LTV on these loans falls below 80% after several payments, MI premia are only paid for a short period. The NPV of loan repayments, proportional to the area under the payment schedule, increases *marginally* with respect to LTV.

Panel (b) of Figure 3 shows the effect of the step-up in MI rates at 90% LTV. The bottom-most solid line represents scheduled payments for a loan of 90% initial LTV. The required MI payments last for nearly 70 months until the remaining LTV falls below 80%, after which the monthly payment amount falls. A loan with initial LTV marginally above 90% has higher MI payments due to the MI rate step-up. This *discontinuous* increase in the initial payment is smaller than the discontinuity at 80% since the rate step-up is smaller. However, the higher rate applies until the remaining LTV falls below 80%. As such, the NPV of loan payments (the additional area under the schedule) increases *discontinuously* when crossing the 90% threshold, in contrast to the *marginal* effect of crossing the 80% threshold on the NPV of payments.

Figure 3: Effect of Mortgage Insurance on Payment Schedules



(a) 80% LTV



(b) 90% LTV

Note: The figure shows the schedule of monthly payments for loans located at and above LTV thresholds. The solid line in Panel (a) shows the monthly payments for an 80% LTV loan on a $300,000 property, while the dashed lines show payments for loans at marginally higher initial LTVs, incorporating required mortgage insurance (MI) payments. Panel (b) shows the monthly payments for loans at 90% (solid) and higher (dashed) LTVs, again incorporating required MI payments. The figures reflect the theoretical effect of mortgage insurance alone, and do not account for potential changes in loan interest rate based on LTV.

In summary, the relationship between initial LTV and the initial monthly payment exhibits a large notch at 80% LTV and a smaller notch at 90% LTV. On the other hand, the relationship between initial LTV and the NPV of payments is kinked at 80% LTV and notched at 90% LTV. These different discontinuities produce distinct predictions of borrowers' bunching response and are central in identifying borrowers' decision processes. Section 2 describes these predictions and the core identification strategy of this paper.

1.2 Mortgage and Borrower Data

The data on loan originations comprise two primary sources: the Home Mortgage Disclosure Act (HMDA) loan origination data, and the Equifax Credit Risk Insight Servicing McDash (CRISM) data from Equifax and ICE, McDash. The HMDA data are compiled from reports submitted by lending institutions for the purpose of assessing trends in the US mortgage market and ensuring fair lending practices. The data includes borrower demographics and loan characteristics, and cover approximately 90% of all mortgage applications and originations in the United States each year.

The CRISM data are an anonymized loan-level panel of mortgage servicing data from McDash Analytics, supplemented with credit report data from Equifax. The data include loan characteristics such as loan-to-value ratios, debt-to-income ratios, interest rates, the zip code of the mortgaged property, and applicant credit scores at origination. The credit report data include ongoing credit scores and credit variables such as credit card balances and limits. The panel stretches from 6 months prior to loan origination up to 6 months after termination.

The HMDA and CRISM data are merged using origination date, origination amounts, property ZIP codes, and loan types. In the merged dataset (the "HMDA-McDash-CRISM match") the date of mortgage origination is rounded to the nearest quarter and property location is reported at the ZIP code level to protect borrower confidentiality. All loan or borrower information and all results in this paper are from this merged dataset unless indicated otherwise. All credit scores used in the analysis are scores at origination from ICE, McDash unless otherwise indicated.

The sample is restricted to mortgages for the purchase of owner-occupied, single-family homes. It includes only 30-year, fixed-rate, fully amortizing, conventional loans with monthly payments. Additionally, the sample includes only borrowers with an initial LTV between 76% and 95%, and with credit scores between 640 and 780. The final sample includes loans originated from 2014-2018 and comprises 591,159 unique loans. Table 1 displays summary statistics for these loans and the primary borrowers associated with each loan.

The borrowers in this sample are higher-income and more creditworthy than the average American, and borrowers at the 80% and 90% LTV thresholds are similar to each other. Table 1 shows average income of $100,000, about double that of the US median (Board of Governors of the Federal Reserve System (U.S.) (2024)). The average loan amount is around $250,000, and property values are around $300,000.

Table 1: Loan and Borrower Summary Statistics

	All	80% LTV	90% LTV
Income ($)	99090	109300	102000
Interest Rate (Basis Points)	424	419	423
Credit Score at Origination	732	735	733
Debt-to-Income (%)	34.8	33.8	34.8
Loan Amount ($)	248000	269000	261900
Appraisal Value ($)	290000	337100	291700

Note: The table reports averages for 591,159 30-year, fixed-rate, fully amortizing conventional loans securitized by Fannie Mae or Freddie Mac, with loan-to-value (LTV) ratios in the (76,95) range and credit scores in the [640, 780] range. All variables are from the matched HMDA-McDash-CRISM dataset, with income from HMDA and all other variables from CRISM. Debt-to-Income refers to the ratio of monthly housing payments to income. Appraisal Value refers to the appraised value of the property the loan is secured by.

2 Identifying Payment Sensitivity from Bunching

This section quantifies the effect of mortgage insurance (MI) on borrowing costs and payments and develops predictions of expected bunching behavior. Section 2.1 proposes two hypotheses for borrower behavior: NPV sensitivity, under which borrowers respond to the net present value of loan repayment costs; and payment sensitivity, under which borrowers respond to the size of the recurring payment obligation. Section 2.2 quantifies the effect of MI requirements on the NPV of payments and the initial recurring payment. Section 2.3 uses a model of borrower optimization to generate predictions of empirical bunching under each of these hypotheses and develops a test to discriminate between them. Section 2.4 presents the main identification assumption.

2.1 Effect of Mortgage Insurance on Initial Monthly Payment & Net Present Value of Payments

The net present value (NPV) of loan payments is a natural quantity for borrowers to consider when taking out loans. Under the assumption of frictionless adjustment and complete markets, borrowers aiming to maximize lifetime consumption utility need only consider the NPV of payments as a sufficient statistic when evaluating loan options. Even under more realistic assumptions, the NPV of payments serves as a useful metric for evaluating loan options: all forward-looking borrowers who have some capacity and desire to smooth consumption across time should be somewhat sensitive to the NPV of payments.

Another quantity borrowers may be sensitive to when choosing loans is the initial monthly payment. At first glance, optimizing agents should not be highly responsive to the monthly payment as it need not be correlated with the total cost of borrowing. For example, an amortizing loan with a very long term could offer much lower payments with higher total costs than a loan for the same amount with a shorter repayment term. However, there is some evidence that the monthly payment does matter to borrowers: Hastings and Shapiro (2013) find that consumers use mental accounts for various budget categories, which could generate sensitivity to the size of the monthly payment, while Juster (1964) finds survey evidence of sensitivity to the monthly payment. Personal finance advice, advertising, and even government policy also provide suggestive evidence for the importance of monthly payments; auto loans are often advertised based on the monthly payment, regardless of loan term, while the federal government defines cost-burden for housing as making monthly shelter payments greater than 30% of income (Department of Housing and Urban Development, 2014).

For the baseline results, I consider two quantities that change discontinuously at the 80% and 90% LTV thresholds: the initial monthly payment and the net present value (NPV) of payments. The initial monthly payment refers to the first of the series of recurring loan repayment obligations, while the NPV of payments refers to the value of the stream of expected payment obligations, discounted at 5% per year.[6]

A key empirical goal of this paper is to distinguish between two hypotheses: payment sensitivity and NPV sensitivity. The payment sensitivity hypothesis states that borrowers act as if loans trade off access to immediate liquidity against the nominal size of the recurring repayment obligation. The NPV sensitivity null hypothesis states that borrowers act as if a loan is a trade-off between access to liquidity today and the NPV cost of repayments. This hypothesis is a useful benchmark as it distills the utility cost of future loan payments into a single statistic, which should be relevant for any forward-looking borrower with some ability to move consumption across time. The primary empirical approach tests between the payment sensitivity and NPV sensitivity hypotheses, while secondary analyses test payment sensitivity against other potential explanations such as NPV sensitivity augmented with adjustment frictions or alternative preferences.

2.2 Effect of Mortgage Insurance on Payments

MI requirements generate different discontinuities in the initial monthly payment and the NPV of payments as functions of the initial loan-to-value ratio (LTV). As described in Section 1.1 and illustrated in Figure 3, MI generates a *notch* in initial payments at the 80% LTV and 90% thresholds,

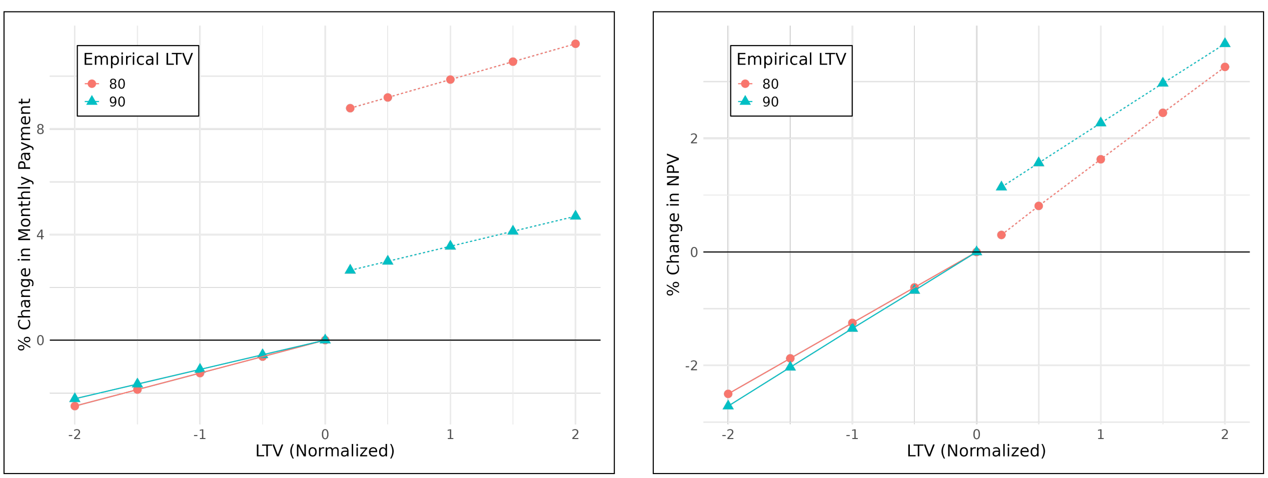
[6]The 5% discount rate is used throughout unless otherwise specified. Section 5.2 considers alternative preferences and finds that the payment sensitivity result is robust to a wide range of discount rates.

with a discontinuous increase in the level when exceeding the threshold. On the other hand, the NPV of payments experiences a *kink* at the 80% threshold: the marginal rate at which NPV grows with respect to LTV increases when crossing the threshold. At the 90% threshold, NPV is *notched* with respect to LTV: there is a discontinuous increase in the NPV of payments when crossing the 90% threshold.

Figure 4 shows the relative sizes of the kinks and notches generated by MI that borrowers in the data face. Panel (a) shows the effect of MI on initial monthly payments, using data from borrowers bunching at 80% (represented in red circles) and borrowers bunching at 90% (represented in green triangles). For each borrower the monthly payment and LTV are normalized to zero. Then, for a range of counterfactual LTVs, an alternative loan is constructed taking into account changes in MI, and the new initial monthly payment is computed. The figure reports the average change in the payment at different LTVs, relative to the actual payment when located at 80% or 90% LTV. The figure shows that above 80% LTV the loan menu has a larger notch in initial payments, while the notch in initial payments is smaller when exceeding the 90% LTV threshold.

Panel (b) of Figure 4 shows the effect on the NPV of payments. Using the same counterfactual loans, the figure reports the change in NPV of payments if bunchers had located at different LTVs. Borrowers at the 80% threshold experience a kink: a change in slope of the NPV:LTV schedule. Borrowers at the 90% threshold experience a notch: a discontinuous jump in the NPV:LTV schedule.

Figure 4: Effect of Mortgage Insurance on Initial Payment and NPV of Payments

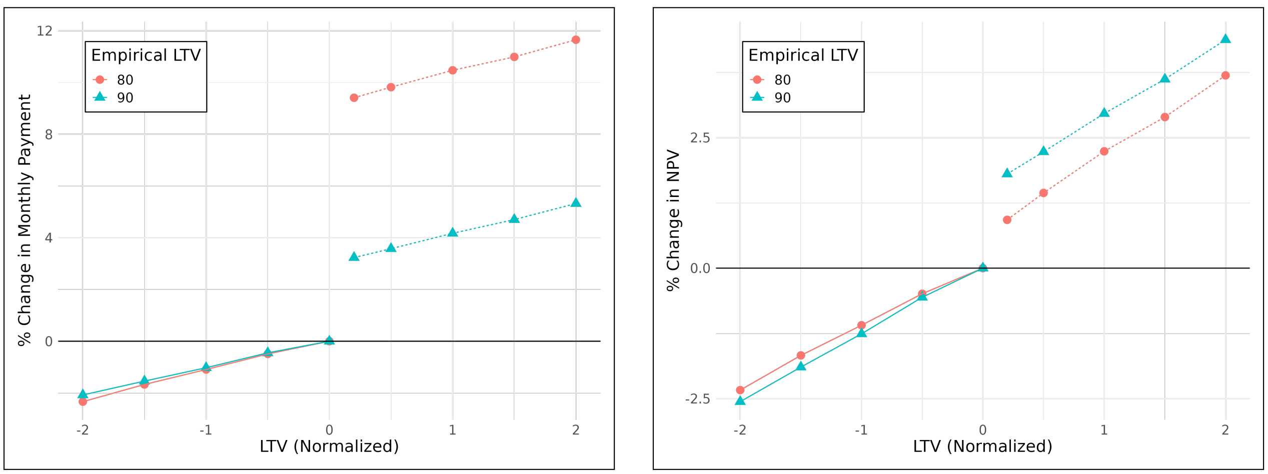


(a) Change in Monthly Payments **(b) Change in NPV of Payments**

Note: The figure shows the effect of mortgage insurance (MI) on the initial monthly payment and the net present value (NPV) of payments. For bunchers located at 80% and 90% LTV, I compute counterfactual loan payment schedules if located at other LTVs, accounting for MI. Panel (a) shows the average expected change in initial monthly payment, normalized. Panel (b) shows the average expected change in NPV of payments.

Figure 4 shows the effect of MI only and does not incorporate other pricing changes at these thresholds. The most important of these are the causal effect of the LTV threshold due to coarse LTV-based pricing policies. I estimate the direct effect of LTV on mortgage rates by regression on observables in Appendix B and find that the effect of crossing the threshold is about 6bps at both the 80% and 90% thresholds.[7] This effect is small relative to the MI rate and not significantly different at the two thresholds. As such, incorporating the effect does not qualitatively change the pattern of discontinuities at the thresholds. Panel (a) of Figure 5 shows that even when incorporating the direct effect of LTV thresholds on rates borrowers still face a much larger notch in initial payments when crossing the 80% threshold than the 90% threshold, while panel (b) shows the same pattern reversal for NPVs with borrowers crossing the 80% threshold facing a smaller notch.

Figure 5: Effect of Mortgage Insurance and Rate Changes on Initial Payment and NPV of Payments



(a) Change in Monthly Payments **(b) Change in NPV of Payments**

Note: The figure shows the effect of mortgage insurance (MI) and changes in rates on the initial monthly payment and the net present value (NPV) of payments. For bunchers located at 80% and 90% LTV I compute counterfactual loan payment schedules if located at other LTVs, accounting for MI. Panel (a) shows the average expected change in initial monthly payment, normalized. Panel (b) shows the average expected change in NPV of payments.

2.3 Testing for Payment Sensitivity with Bunching

The different discontinuities in incentives at the 80% and 90% LTV thresholds present a way to distinguish between the NPV sensitivity and payment sensitivity hypotheses. I set up a simple

[7] See Bartlett et al. (2022) for a description of loan pricing by borrower characteristics in the conforming market. The conforming loan market is very liquid and the GSEs have standardized pricing adjustments that depend on credit score and LTV. As such, controlling for observables accounts for most of the selection across the thresholds and recovers the direct effect of the threshold on LTV.

model of borrower behavior and use the methods from Saez (2010), Chetty et al. (2011) and Kleven and Waseem (2013) to make predictions of bunching in borrower LTV in response to these discontinuities. Comparing the realized bunching in observed LTVs to these predictions can then distinguish between the NPV sensitivity and payment sensitivity hypotheses.

Suppose borrowers are NPV sensitive and trade off the benefit of immediate access to funds against the discounted cost of loan repayments. Represent the optimization problem for borrower i when choosing a loan as:

$$\max_{L} \quad U_i(W_0 - H + LH) - \int_0^T e^{-\delta t} P_t(r, LH) dt \tag{1}$$

$$st. 0 \leq L \leq H \tag{2}$$

where W_0 is initial wealth and H is the fixed home price. L is the LTV choice, so $W_0 - H + LH$ is remaining cash-on-hand after making the down payment. The mortgage has fixed term T and $P_t(\cdot)$ is the potentially time-varying instantaneous flow payment due at time t, which is a function of the interest rate r and the loan amount LH. The agent receives utility $U_i(\cdot)$ from cash-on-hand with $U_i(\cdot)' > 0$ and $U''(\cdot) < 0$, and disutility from the NPV of repayments, captured by the integral term.

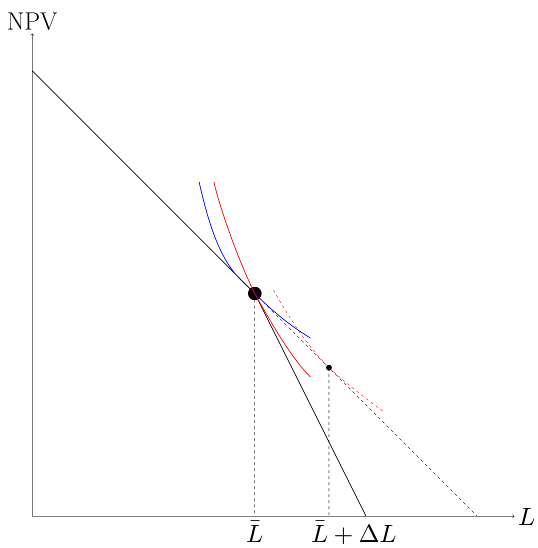
Heterogeneous borrower preferences correspond to different curvature of U_i for each borrower. Each borrower chooses an optimal L to equate the marginal utility from cash-on-hand against the value of future payments.[8] Panel (a) of Figure 6 represents the problem in NPV-L space; in the absence of MI-induced pricing discontinuities, the budget constraint is approximately linear and represented by the solid diagonal line to the left of $\bar{L}$ and the dotted line to the right of $\bar{L}$. Utility is increasing when moving upwards as borrowers benefit from a less negative NPV of repayments; utility increases to the right as higher L directly increases cash-on-hand.

Suppose borrowers' utility is parameterized by a concavity parameter γ such that $U'(\cdot)$ is decreasing in γ. Let γ be distributed in the population by some continuous density $f(\gamma)$. Since W_0 and H are fixed, each value of γ maps to a corresponding optimal LTV L, and the corresponding distribution of initial LTV $g_0(L)$ is smooth. Panel (a) of Figure 6 shows indifference curves for two borrowers of different γ whose optimal choices are $\bar{L}$ and $\bar{L} + \Delta L$ where their indifference curves

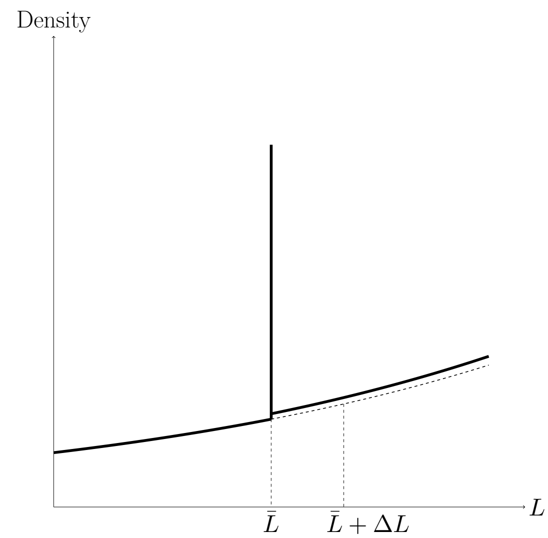
[8] This formulation can easily represent utility from housing consumption. Fix the down payment amount and let H vary, then the value of the house H is a deterministic function of LTV L. Borrowers choose some L that trades off the utility from consuming more housing against higher future repayments.

are tangent to the budget constraint. The corresponding smooth distribution of LTV is shown in panel (b) by the combination of the solid line to the left of $\bar{L}$ and the dotted line to the right of $\bar{L}$.

Figure 6: Effect of Kinks on Observed Density



(a) Kink Budget Set



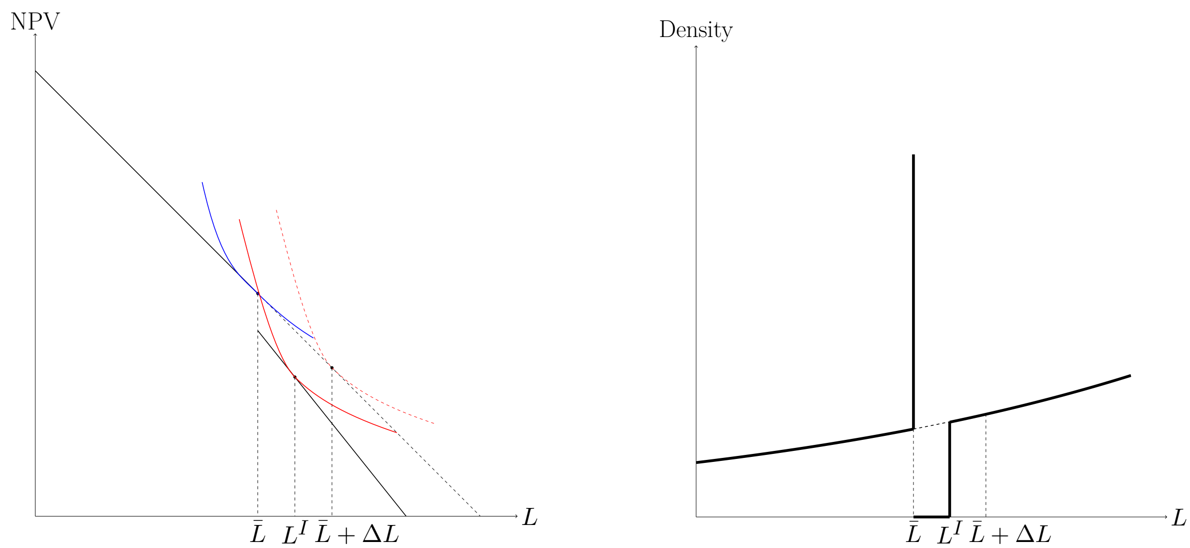
(b) Kink Density

Note: The figure illustrates the theoretical effect of a kink in the budget constraint on the observed distribution of borrower LTVs.

Panel (a) of Figure 6 shows the effect of introducing a kink at LTV threshold $\bar{L}$: the marginal cost of increasing L increases and the budget constraint steepens above $\bar{L}$, rotating from the dotted line to the solid line. As per Saez (2010), the kink induces the marginal buncher initially located at $\bar{L}+\Delta L$ — as well as all borrowers initially in $(\bar{L}, \bar{L}+\Delta L)$ — to move to the kink point, resulting in excess bunching at $\bar{L}$. Borrowers initially located above $\bar{L}+\Delta L$ remain strictly above $\bar{L}$ but re-optimize by reducing their borrowing and shifting to the left. Panel (b) shows the entire density initially to the right of $\bar{L}+\Delta L$ shifts to the left, as represented by the solid line to the right of $\bar{L}$. Critically, borrowers still choose to locate in the region just above $\bar{L}$.

Figure 7 shows the characteristic difference in bunching induced by notches compared to kinks. Panel (a) shows that the notch causes a discontinuous jump down in the budget constraint at the threshold $\bar{L}$. The marginal buncher who was previously at $\bar{L}+\Delta L$ is now indifferent between locating at $\bar{L}$ and L^I. As described in Kleven and Waseem (2013), all borrowers initially located in $[\bar{L}, \bar{L}+\Delta L]$ bunch at the notch, while borrowers initially above $\bar{L}+\Delta L$ reduce their chosen L. The resulting distribution of LTVs is shown in panel (b): there is excess mass at the threshold relative to the no-notch counterfactual distribution and, characteristically, missing mass relative to the no-notch counterfactual in $[\bar{L}, L^I]$ where no borrowers locate.

Figure 7: Effect of Notches on Observed Density



(a) Notch Budget Set (b) Notch Density

Note: The figure illustrates the theoretical effect of a kink in the budget constraint on the observed distribution of borrower LTVs.

Critically for this paper, kinks do not induce missing mass above the threshold while notches do. Under the NPV sensitivity hypothesis, borrowers optimizing over present liquidity and the NPV of future payments would generate excess bunching mass at both the 80% and 90% LTV thresholds, but would only generate missing mass above the 90% threshold. The payment sensitivity hypothesis generates the opposite prediction of more missing mass above the 80% threshold than the 90% threshold. To see this, consider borrowers optimizing over the initial payment who choose:

$$\max_{L} \quad U_i(W_0 - H + LH) - cP_1(r, LH) \tag{3}$$

$$st. 0 \leq L \leq H \tag{4}$$

Payment sensitive borrowers trade off liquidity today against some multiple c of the initial payment P_1, as opposed to the value of the time-varying payment P_t. Section 2.2 shows that borrowers face a bigger notch in P_1 at the 80% LTV threshold than at the 90% LTV threshold.

Using the same bunching logic as above, Appendix C shows that bigger notches induce more missing mass. As such, if borrowers are payment sensitive, there should be more missing mass above the 80% LTV threshold than the 90% LTV threshold.

In summary, the missing mass above the 80% and 90% LTV thresholds can be used to distinguish between NPV sensitivity and payment sensitivity.[9] NPV sensitive borrowers responding to the larger notch in NPV at the 90% threshold would generate more missing mass above 90% LTV. Payment sensitive borrowers responding to the larger notch in initial monthly payment at the 80% threshold would generate the opposite pattern with more missing mass above 80% LTV. I apply this test in Section 3 by defining and calculating the missing mass above the thresholds relative to an estimate of the counterfactual distribution.

2.4 Identification Assumptions

The differences-in-bunching test depends on borrowers around the 80% and 90% LTV thresholds having similar responses to discontinuities of similar size. To illustrate this identification assumption, consider the ideal experiment for testing NPV sensitivity: randomize borrowers into two loan menus, where the first menu features a small notch discontinuity in NPV:LTV at a threshold, while the second menu features a larger notch in NPV at the *same* threshold. Since menus are randomized and the notch is located at the same threshold, adjustment frictions and the distribution of borrower preferences are similar in each treatment arm. Hence, any differences in observed missing mass above the threshold would be due to borrowers responding to the larger notch and increased cost of choosing higher LTV. Observing more missing mass above the threshold, for borrowers facing the menu with a large notch, would support NPV sensitivity; observing the same or less missing mass would favor rejecting the hypothesis.

How could the identification assumption fail? If some borrowers had greater adjustment costs of reducing LTV, they would respond less to a notch of a given size. If frictions were greater for borrowers at the 90% threshold, where there is a bigger notch in NPV, than for borrowers at the 80% threshold, there could be less missing mass above the 90% threshold even if borrowers were NPV sensitive.

[9] I focus on the 80% and 90% thresholds only and not other thresholds for three reasons. First, 80% and 90% are of similar "roundness," which will be differenced out by the comparison of bunching. Second, I exclude the 85% threshold because there is very little count data in that region, which could lead to noisy estimates. Finally, I exclude the 95% threshold because 95% LTV is the maximum LTV for most conforming loans, and there may be a significant extensive margin response in that area.

More generally, the test assumes the difference in size of the NPV notches is a good approximation for the difference in size of unobserved notches in total utility costs. If this approximation fails and the utility cost of exceeding the 90% threshold was smaller than exceeding the 80% threshold, the difference in costs could explain lack of missing mass above 90% instead of failure of NPV sensitivity. As described above, unobserved differences in frictions could violate this assumption. Another possible violation is differences in discount rates: if agents above 90% LTV were more impatient, they would perceive a low utility cost of exceeding the 90% threshold, leading to less bunching and less missing mass above 90%. Similarly, optimism over income growth implies lower disutility from high payments in the future and could cause borrowers to perceive a low cost of exceeding 90%.

Empirically, there is little evidence of heterogeneity in adjustment frictions or preferences that are large enough to invalidate the test. I defer in-depth discussion of these potential violations of the identification assumption to Section 4, after presenting the bunching results. Section 3 presents the bunching results and uses the framework above to show that borrowers act as if they face a bigger notch in their budget constraint at the 80% LTV threshold, exactly where initial payments have the largest notch. Section 4.1 shows that borrowers are similar on observables at both thresholds, and likely have similar preferences. Section 4.2 examines the amount of selection at both thresholds and finds no evidence of differential frictions. Finally, Section 4.3 shows that the absolute level of frictions must be implausibly high to reconcile the bunching results with NPV sensitivity.

3 Bunching Estimation and Results

As described in Section 2, the missing mass above the LTV thresholds relative to the no-MI counterfactual distribution of borrower LTVs can be used to distinguish between the NPV sensitivity and payment sensitivity hypotheses. This section describes the estimation of the counterfactual distribution of borrower LTVs. Section 3.1 sets up the estimation problem and defines bunching estimands of interest; Section 3.2 describes the main bunching estimation procedure; Section 3.3 describes a sensitivity analysis; Section 3.4 presents the results of the bunching estimation and the evidence in favor of payment sensitivity.

3.1 Bunching Counterfactuals and Estimands

The estimation of the bunching counterfactual follows the approach outlined by Kleven and Waseem (2013). In this framework, borrower choices are presumed to be distorted in a window around the policy threshold while remaining undistorted outside this window. The key assumption is that in the absence of the policy the distribution of borrowers would be smooth. The estimation strategy uses data from the undistorted region to recover the counterfactual distribution. Specifically, the econometrician selects an excluded region around the threshold and fits a smooth polynomial of LTV (the running variable) to data outside this region. This polynomial serves as an estimate of the counterfactual distribution of LTV, representing the distribution of LTV in the absence of the mortgage insurance requirement.

Consider data around a LTV threshold T. Begin by binning the data into LTV bins X_i of width 0.1 and let Y_i be the count of loans in bin i, i.e. the number of loans in the data with LTV within ± 0.05pp of X_i%. Denote the set of bins indices by $\mathbb{I}$. I model the counterfactual count of loans in a particular bin in the absence of distortions as a 3rd degree polynomial of LTV:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + \varepsilon_i \tag{5}$$

The MI requirement induces distortions in the bin counts in a region around the LTV threshold. For LTV threshold T, let $X_l < T$ and $X_r > T$ be the left and right endpoints of an excluded window around T. Let $\mathbb{K}$ be the set of bin indices within the excluded window, i.e. $\mathbb{K} = \{i \in \mathbb{I} \quad s.t. \quad i \geq l, k \leq r\}$. The set of *un-excluded* bin indices is the set of bins that are not excluded $\mathbb{J} = \mathbb{I} \backslash \mathbb{K}$.

Using the assumption that bin counts in the un-excluded bins $\mathbb{J}$ are not distorted and that the counterfactual distribution is smooth and well-approximated by the polynomial, I estimate the coefficients of the polynomial by finding $\{\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3\}$ that solve:

$$\min_{\{\beta_0,\beta_1,\beta_2,\beta_3\}} \sum_{i \in \mathbb{J}} \left(Y_i - \beta_0 - \beta_1 X_i - \beta_2 X_i^2 - \beta_3 X_i^3\right)^2 \tag{6}$$

This specification is numerically identical to the procedure in Kleven and Waseem (2013) that uses the whole range of data and dummies for the excluded region; both methods minimize the sum of squared errors for un-excluded bin counts. Estimation with Equation 6 offers additional benefits of computational simplicity and ease of adding more constraints as in Section 3.3 below.

With the estimated $\{\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3\}$, I compute the predicted counterfactual bin counts as $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{\beta}_2 X_i^2 + \hat{\beta}_3 X_i^3$. The excess bunching at the threshold is given by the difference in observed and predicted counts in the left excluded window:

$$\hat{B} = \sum_{i=l}^{T} Y_i - \hat{Y}_i \tag{7}$$

Assuming the counterfactual density is approximately constant in the excluded region above the threshold, $[T, r]$, the average LTV response is given by the estimated amount of bunching divided by the counterfactual density at and to the left of the threshold:

$$\Delta \hat{LTV} = \frac{\hat{B}}{\left(\sum_{i=l}^{T} \hat{Y}_i\right) / (T - X_l)} \tag{8}$$

The excess mass at and to the left of the threshold is given by the difference between the observed count and counterfactual count, scaled by the counterfactual count:

$$ExMass = \frac{\hat{B}}{\sum_{i=l}^{T} \hat{Y}_i} \tag{9}$$

The key difference in predictions between the NPV sensitivity and payment sensitivity models is the amount of missing mass above the LTV thresholds. Similar to the excess mass, the missing mass to the right of the threshold is given by the scaled difference between observed and counterfactual count:

$$MissMass = 1 - \left(\frac{\sum_{i=T+0.1}^{r} \hat{Y}_i - Y_i}{\sum_{i=T+0.1}^{r} \hat{Y}_i} \right) \tag{10}$$

3.2 Estimation and Choice of Excluded Region

Estimated counterfactuals are sensitive to the choice of the excluded region around the threshold. In some settings, bunching to the left of the threshold is stark, allowing the left endpoint of the excluded region to be chosen by visual inspection. The right endpoint is then selected to minimize the difference between the excess count at the threshold and the missing count above the threshold.

Table 2: LTV Data and Excluded Regions for Estimation

Threshold (T)	LTV Data	Left Endpoints	Right Endpoints
80	[76, 84.9]	$\text{L}_T = \{78, 78.2, ...79.6, 79.6\}$	$\text{R}_T = \{81, 81.2, ...84.6, 84.8\}$
90	[86, 94.9]	$\text{L}_T = \{88, 88.2, ...89.6, 89.6\}$	$\text{R}_T = \{91, 91.2, ...94.6, 94.8\}$

Note: The table shows the range of data and the set of possible endpoints for the excluded region used to estimate bunching counterfactuals around each LTV threshold.

In my setting, the point where bunching distortion begins on the left side is not immediately obvious from visual inspection. To mitigate specification errors, I explore a range of different left endpoints for each threshold. Table 2 displays the data and the endpoints for the excluded region used for estimation. For example, the first row shows that for bunching around the 80% threshold I use count data for LTV bins ranging from 76% to 84.9%. For each possible combination of endpoints $(l, r) \in L_{80} \times R_{80}$, I exclude the bins in $[l, r]$ and estimate Equation 6 on the remaining data to obtain the counterfactual prediction and bunching statistics.

The estimations are performed separately for each threshold. For each threshold I then choose the model that minimizes the absolute difference between the excess mass and missing mass. Hence, for threshold T the preferred model has left and right endpoints l, r that minimize the adding-up error between excess and missing mass:

$$\min_{\substack{l \in L_T \\ r \in R_T}} \left| \sum_{i \in [l,T]} \left(Y_i - \hat{Y}_i\right) - \sum_{i \in (T,r]} \left(\hat{Y}_i - Y_i\right) \right| \tag{11}$$

I obtain standard errors for the bunching statistics via bootstrapping. Bootstrap samples are created by resampling from the residuals after estimating Equation 6. For each bootstrap sample, Equation 6 is re-estimated holding l constant but re-choosing r to minimize the adding-up error. The standard errors for bunching statistics are the standard deviation of the bootstrap bunching estimates. Appendix D describes the procedure in detail and shows that it is numerically identical to the method in Chetty et al. (2011).

3.3 Sensitivity Analysis

I conduct a sensitivity analysis of the bunching estimates with respect to the choice of excluded region. Begin by adding a constraint to Equation 6 to generate a new estimating equation:

$$\min_{\{\beta_0,\beta_1,\beta_2,\beta_3\}} \sum_{i\in\mathbb{J}} \left(Y_i - \beta_0 - \beta_1 X_i - \beta_2 X_i^2 - \beta_3 X_i^3\right)^2 \tag{12}$$

$$\text{Subject to} \sum_{i\in J\cup K} \left(\beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3\right) = \sum_{i\in J\cup K} \hat{Y}_i = \sum_{i\in J\cup K} Y_i \tag{13}$$

Like Equation 6, this estimating equation minimizes the sum of squared errors of the polynomial fit for the un-excluded bins $\mathbb{J}$. The constraint in Equation 12 requires the predicted bin counts across the whole range of LTV to equal the empirical count, reflecting the assumption of no extensive margin response to the pricing discontinuity. I perform the constrained optimization in Equation 12 separately for each threshold, using all the possible combinations of endpoints specified in Table 2. This procedure yields 180 counterfactuals and associated bunching statistics for each threshold.

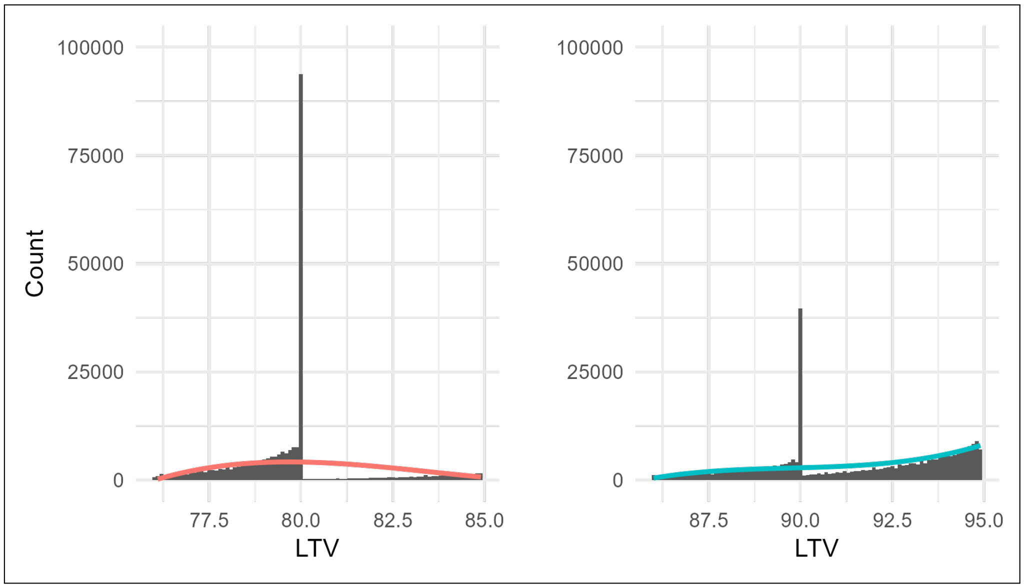
This approach addresses estimation challenges associated with wide excluded regions and limited data. When the excluded region is wide, the estimation procedure in Section 3.2 relies on narrow bands of remaining undistorted data, which can cause bin counts at the extreme ends of the data range to greatly influence the estimated polynomial. In addition, many of the candidate counterfactuals imply implausibly large discrepancies between the excess mass at the threshold and the missing mass above the threshold. Forcing the adding-up constraint to bind exactly and re-estimating across the set of candidate excluded regions leverages the total count information across all the bins and yields more plausible counterfactuals. Furthermore, this method reduces the outsize influence of bin counts at the ends of the LTV range when excluded regions are wide and count data are limited. Consequently, I report a range of bunching statistics across a large set of plausible models and perform a sensitivity analysis, accounting for uncertainty due to model mis-specification.

3.4 Bunching Results

The estimated counterfactuals reveal a bunching pattern inconsistent with the NPV sensitive benchmark, and supportive of the payment sensitivity hypothesis. The main empirical finding is significantly more missing mass of borrowers above the 80% than the 90% LTV threshold. Figure 8 shows the preferred estimates from the procedure described in Section 3.2. The left panel shows significant missing mass above the 80% LTV threshold: the predicted counterfactual bin counts indicated by the red line are substantially higher than the empirical counts shown by the gray bars. This dramatic difference in predicted and observed counts extends substantially above the threshold. In contrast, the right panel shows the existence of less missing mass above the 90% threshold.

The predicted counterfactual count above the threshold exceeds the observed count by less, and the difference between predicted and observed counts shrinks substantially at higher LTVs.

Figure 8: Estimated Counterfactual LTV Distributions



Note: The figure shows the counterfactual distributions of borrower LTV, estimated separately around the 80% and 90% LTV thresholds using Equation 6.

Table 3 reports the bunching statistics from the preferred model at each threshold and quantifies the differences in bunching and missing mass. The number of excess borrowers bunching at or just below the 80% threshold is 120,000, compared to 42,000 at the 90% threshold. These excess bunching counts represent 2.3 and 3.0 times the predicted counterfactual count respectively. Using the approximation from Kleven and Waseem (2013) presented in Equation 8, I estimate the average reduction in LTV at the 80% threshold to be 3.2 LTV points, compared to 1.2 LTV points for bunchers at the 90% threshold. However, the estimates of excess mass and LTV response are noisy due to uncertainty about the true counterfactual distribution and potential model mis-specification.

The estimates of missing mass are more precise and demonstrate a stark difference between borrowers at the two thresholds. In the excluded region above the 80% threshold, 0.816 of the expected counterfactual borrowers are missing, compared to just 0.322 above the 90% threshold, consistent with bunching at the 80% threshold being much more notch-like than at the 90% threshold. The missing mass is precisely estimated because it depends only on the difference between the actual and counterfactual counts in the excluded window above the threshold, and is relatively stable across various plausible counterfactuals. In contrast, excess mass and LTV response are functions of the bunching count as well as the counterfactual density at and to the left of the threshold, and are thus more sensitive to small changes in the counterfactual polynomial and the width of the left bunching window.

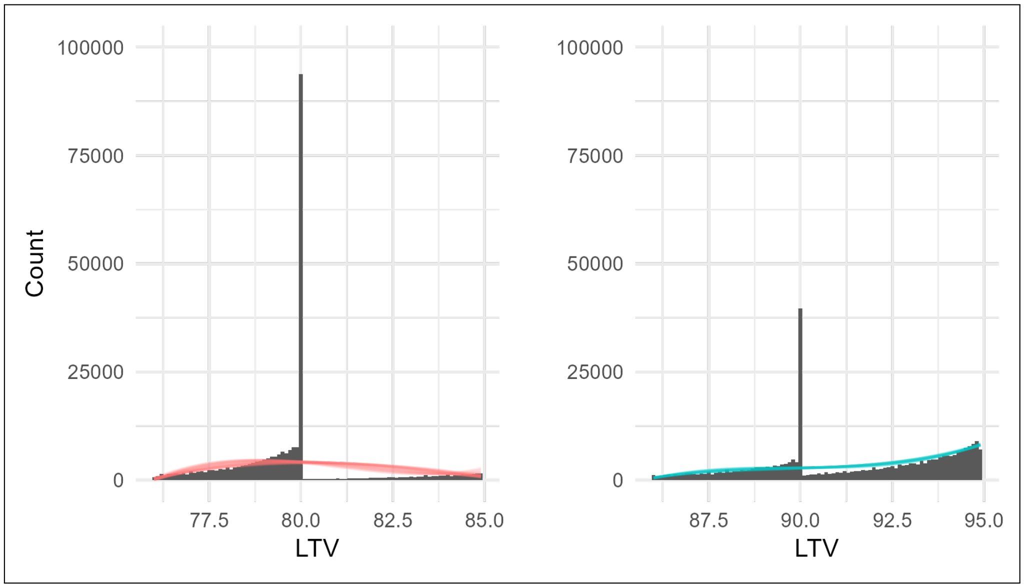
Table 3: Estimation Results

	80% LTV	90% LTV
Bunching Count	119151	42337
	(5905)	(1335)
Excess Mass	2.296	3.038
	(0.579)	(0.352)
LTV Response	3.214	1.215
	(0.811)	(0.141)
Missing Mass Ratio	0.816	0.322
	(0.044)	(0.049)

Note: The table reports the bunching statistics from the estimated counterfactual LTV distributions, obtained from the procedure in Section 3.2. As defined in Equations 7 - 10, Bunching Count is the difference between actual and predicted counts in LTV bins in the left excluded region up to the bunch point; Excess Mass is the ratio of the bunching count to the predicted density in the left excluded region; LTV Response is the average reduction in LTV by bunchers to reach the threshold, and Missing Mass Ratio is the difference between predicted and actual counts in the right excluded region above the threshold, scaled by the predicted count. Standard errors in parentheses are obtained via bootstrap.

The results from the sensitivity analysis in Section 3.3 show that the pattern of more missing mass above 80% LTV is very robust. Figure 9 overlays the estimated counterfactuals from the constrained optimization in Equation 12 against the empirical count of LTVs, with the left panel showing all of the estimated counterfactuals around 80% LTV and the right panel showing all of the estimated counterfactuals around 90% LTV. All of these counterfactuals are constrained such that the total predicted count is equal to the empirical count across the respective LTV range. The left panel shows that, across all the estimated counterfactuals, there is significant missing mass above the 80% threshold, while the right panel shows less missing mass above the 90% threshold. Table 4 reports the quantiles of the estimated missing mass from the sensitivity analysis. The median estimates of 0.88 and 0.34 at the 80% and 90% thresholds are very close to the preferred estimates from the unconstrained estimation of 0.82 and 0.32. In addition, the difference in missing mass at the thresholds is very robust: the smallest estimate of missing mass above the 80% threshold is 0.79, which far exceeds the largest estimate of 0.48 missing mass above the 90% threshold.

Figure 9: Estimated Counterfactual LTV Distributions



Note: The figure shows the estimated counterfactuals from the sensitivity analysis in Section 3.3. At each threshold, for each candidate excluded region, the bunching counterfactual is estimated using Equation 12, with the constraint that the predicted count equals the empirical count. The figure plots all the estimated counterfactuals against the empirical LTV counts, with the range around 80% LTV on the left and the range around 90% LTV on the right.

Table 4: Sensitivity Analysis, Quantiles of Estimated Missing Mass at Thresholds

Percentile	Missing Mass, 80% LTV	Missing Mass, 90% LTV
0%	0.79	0.21
25%	0.85	0.30
50%	0.88	0.34
75%	0.90	0.39
100%	0.92	0.48

Note: The table reports quantiles of the estimated missing mass from the sensitivity analysis in Section 3.3. At each threshold, for each candidate excluded region, the bunching counterfactual is estimated using Equation 12, with the constraint that the predicted count equals the empirical count. For each bunching counterfactual, the estimate of missing mass in the excluded region above the threshold is computed as per Equation 10. The table reports quantiles of the estimated missing mass for the 80% and 90% thresholds.

4 Identification Threats: Preference Heterogeneity or Frictions

The bunching result provides strong evidence in support of the payment sensitivity hypothesis: borrowers bunch with the most missing mass above the 80% LTV threshold. As shown in Section 2.2, borrowers face larger notches in initial monthly payment at the 80% threshold than the 90% threshold, while the opposite pattern holds for the NPV of payments. Since greater missing mass above a threshold is indicative of a larger notch discontinuity in the budget constraint, the bunching evidence suggests that borrowers are bunching in response to changes in the initial monthly payment as opposed to the NPV of payments.

This section discusses potential violations of the identification assumption that would rationalize the bunching result with the NPV sensitivity benchmark, without the need for payment sensitivity. The key assumption underlying the bunching result is that borrowers near each threshold respond similarly to similar-sized notches. If this were true, more missing mass above the 80% threshold necessarily implies a larger notch in the budget constraint at the 80% threshold, and that initial payments are the main consideration in borrowers' optimization problem. Whereas, if borrowers near 80% responded more to a notch of a given size than borrowers near 90%, the missing mass above 80% could simply reflect NPV sensitive borrowers with heterogeneous responsiveness, rather than payment sensitivity.

Broadly, heterogeneous responses could be due to differences in preferences or differences in adjustment costs. If borrowers near the 90% threshold had more preference for and higher marginal utility of cash-on-hand, they would respond less to a notch of a given size than borrowers with lower marginal utility of cash-on-hand. Similarly, if borrowers near the 90% threshold faced larger frictions in reducing their LTV, they would respond less to a notch of given size. Section 4.1 shows that borrowers near both thresholds are similar on observable measures of income and access to credit, suggesting that they have similar preferences over cash-on-hand. Section 4.2 compares selection into bunching at the thresholds and shows that adjustment frictions are similar for both sets of borrowers. Finally, Section 4.3 shows that, given the cost of locating above thresholds, absolute levels of adjustment costs must be implausibly high to rationalize the bunching results with NPV sensitivity.

4.1 Borrower Characteristics Near Thresholds

Borrowers near both thresholds have similarly high income and ample access to credit. Table 5 displays averages of some characteristics for borrowers exactly at the 80% and 90% LTV thresholds, and for borrowers in the 3-point range above each threshold. Borrowers bunching at the 80% and 90% thresholds have average incomes of $109,000 and $102,000, respectively, and nearly identical credit scores of 735 and 733. They have ample access to credit, with $22,000 and $18,000 of unused credit card limits. Borrowers who do not bunch tend to have lower income and slightly less access to credit, but are similar at each threshold. Borrowers in the $(81-83]$% and $(91-93]$% LTV ranges have average incomes of $89,000 and $87,000, respectively, and credit scores of 729 and 730. They have $16,000 and $15,000 of unused credit card lines.

These summary statistics indicate that borrowers near the 80% and 90% thresholds are very similar. While there is obvious selection into bunching, with higher income borrowers more likely to locate at a threshold, the overall group of borrowers at or near 80% LTV is similar to the group at or near 90% LTV. This suggests that their preferences for cash-on-hand are similar, making it implausible for borrowers near 80% facing a small notch in NPV to be more responsive than borrowers near 90% facing a notch of twice the size. Hence, the bunching result of more missing mass above the 80% threshold is difficult to reconcile with NPV sensitivity, and is more indicative of payment sensitivity.

Table 5: Borrower Characteristics at and above LTV Thresholds

	All	80% LTV	81-83% LTV	90% LTV	91-93% LTV
Income ($)	99090	109300	89100	102000	86510
Credit Score at Origination	732	735	729	733	730
Card Balances ($)	5999	5902	5765	6161	5955
Card Limits ($)	24260	27540	22030	24290	20850
Available Card Limit ($)	18260	21640	16260	18130	14900

Note: The table shows the averages of borrower characteristics for borrowers located exactly at or in the three-point range above the 80% and 90% LTV thresholds. All variables are from the matched HMDA-McDash-CRISM dataset. Income and Credit Score at Origination are reported at the time of origination. Card Balances and Card Limits are from the quarter before origination. Available Card Limit is constructed from balances and limits.

4.2 Relative Frictions: Evidence from Selection

If adjustment frictions were higher for borrowers around 90%, the missing mass above 80% might reflect ease of reducing LTV rather than payment sensitivity. However, the relative amount of selection into bunching at the thresholds provides evidence against differential frictions.

If borrowers around the 90% threshold faced higher frictions, there should be less responsiveness to the discontinuous incentive schedule and less selection across the 90% threshold than across the 80% threshold. This would result in the borrowers who remain above the 90% threshold being less negatively selected compared to borrowers who remain above the 80% threshold. I test for differential selection using borrowers located in the 80-83% and 90-93% LTV ranges. Associate borrowers in $(80\% - 83\%]$ LTV with the 80% threshold and borrowers in $(90\% - 93\%]$ LTV with the 90% threshold and estimate the following equation:

$$Y_i = \beta_0 + \beta_1 LTV90_i + \beta_2 ABOVE_i + \beta_3(LTV90_i \times ABOVE_i) + \epsilon_i \quad (14)$$

where Y_i represents an outcome, $LTV90_i$ is an indicator variable for being at or above the 90% LTV threshold, and $ABOVE_i$ is an indicator for being above one's associated threshold. Then, β_0 is the effect of being at or near 80%, β_1 is the difference between borrowers near 80% and near 90%. β_2 is the effect of exceeding the 80% threshold, and $\beta_2 + \beta_3$ is the effect of exceeding the 90% threshold.

Given that LTV thresholds are unlikely to have a causal effect on borrower characteristics such as income or credit score measured at or prior to application, the β_2 and β_3 coefficients represent selection effects. Specifically, β_2 is the amount of selection occurring across the 80% threshold, while β_3 is the additional selection occurring at the 90% threshold.

Table 6: Selection at Thresholds

Dependent Variables:	IHS(Income)	Credit Score	IHS(Card Balances)	IHS(Card Limits)
Model:	(1)	(2)	(3)	(4)
Variables				
Constant	12.1***	734.1***	7.57***	10.3***
	(0.002)	(0.118)	(0.010)	(0.004)
LTV 90	-0.022***	-1.69***	-0.005	-0.148***
	(0.004)	(0.217)	(0.019)	(0.008)
ABOVE	-0.166***	-4.69***	0.024	-0.316***
	(0.038)	(1.66)	(0.137)	(0.059)
LTV 90 × ABOVE	-0.024	0.548	-0.191	0.051
	(0.041)	(1.87)	(0.157)	(0.067)
Fit statistics				
Observations	126,632	126,632	116,114	116,114
R^2	0.00176	0.00085	4.8×10^{-5}	0.00453
Adjusted R^2	0.00174	0.00082	2.22×10^{-5}	0.00450

Heteroskedasticity-robust standard-errors in parentheses
Signif. Codes: ***: 0.01, **: 0.05, *: 0.1

Note: The table shows the regression results from estimating Equation 14. All outcomes except credit score are the inverse hyperbolic sine of the variable, and coefficients can be interpreted as fractional changes. Income and Credit Score are measured at the time of origination. Card Balances and Card Limits are the balance and credit limit on credit cards one quarter prior to loan origination.

The results from estimating Equation 14 show that borrowers near the 90% threshold do not face significantly higher frictions than borrowers near the 80% threshold. Table 6 presents the regression results, using credit scores and the inverse hyperbolic sine of income, credit card balances, and card limits as outcome variables to proxy for borrower liquidity and quality. The coefficients on "LTV 90" show that borrowers near the 80% and 90% threshold exhibit stark differences, and the coefficients on "ABOVE" show there is significant selection occurring at the 80% threshold. Notably, the coefficients on the interaction term "LTV 90 X ABOVE" are not significantly different from zero, indicating little additional selection at the 90% threshold on income, credit score, card balances, or card limits. The lack of substantially different selection at the thresholds suggests that differential frictions do not drive the bunching pattern.

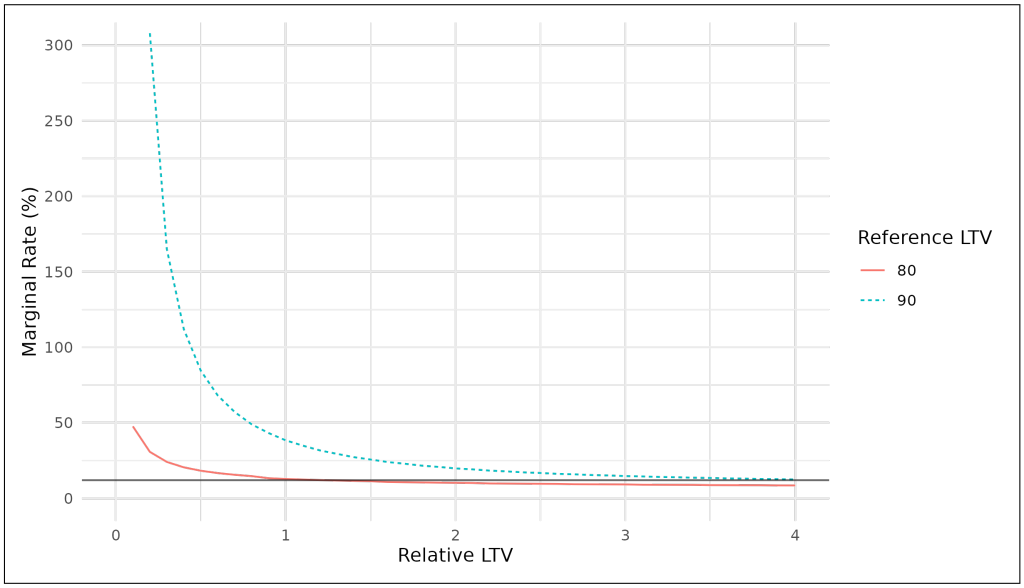
4.3 Absolute Frictions

Rationalizing the bunching result with friction-augmented NPV sensitivity requires not only differential frictions at the thresholds, but also implausibly high absolute levels of adjustment costs. In a frictionless model, there should be *zero* mass above a notch in the dominated region; Kleven and Waseem (2013) show that the remaining mass (i.e. the lack of missing mass) is indicative of adjustment costs or frictions. Hence, if any borrowers locate in the region above a threshold, they must face adjustment costs that exceed the loss from remaining above the threshold.

The cost of remaining above the LTV thresholds is substantial; frictions must be even larger to explain the lack of bunching. Figure 10 reports the cost of remaining above a threshold using the effective marginal interest rate for a typical borrower. The x-axis represents the incremental borrowing in LTV points relative to the threshold, while the y-axis represents the effective interest rate on the incremental borrowing. Exceeding the threshold causes a jump in the average interest rate and MI rate on the loan. This results in significantly higher total payments and high effective rates, especially for small amounts of incremental borrowing. For example, a borrower locating at 91% instead of 90% LTV makes additional payments such that their effective interest rate on one LTV point of borrowing (approximately $2,500) is 40% a year.

It is implausible that adjustment costs exceed the costs of locating above the thresholds. The borrower locating at 91% LTV needs to increase their down payment by one LTV point to reach the threshold; the cost of doing so is unlikely to exceed the implicit borrowing cost of 40%/year. As shown in Table 5, borrowers who locate above thresholds and fail to bunch have high income. Importantly, they have ample access to liquidity and borrowing, including more than $14,000 of

Figure 10: Marginal Rates



Note: The figure shows the effective interest rate on incremental borrowing for a representative borrower located at the 80% and 90% LTV thresholds. The horizontal solid line is the average commercial bank interest rate on credit card plans, from the Board of Governors of the Federal Reserve System (2024).

unused credit card limits.[10] The horizontal solid line in Figure 10 represents the average credit card rate of 12% in this period (Board of Governors of the Federal Reserve System, 2024) - many borrowers who locate above the 90% threshold would be better off if they made a larger down payment and accumulated credit card debt for several months instead of locating above 90% LTV.

In summary, the effective NPV cost of locating above the threshold is too high for frictions to explain NPV sensitive borrowers locating above 90% LTV. Furthermore, explaining the bunching result with friction-augmented NPV sensitivity requires there to be greater frictions around the 90% threshold than the 80% threshold. However, the amount of selection at the thresholds indicates frictions are of similar size at the thresholds. Finally, rationalizing bunching with NPV sensitivity and preference heterogeneity requires borrowers near 90% to have a much greater preference for cash-on-hand; this is unlikely given the similar incomes and access to credit for borrowers around both thresholds. In total, the bunching pattern of more missing mass above the 80% threshold is difficult to square with NPV sensitivity without extreme assumptions about unobserved prefer-

[10]Panel (a) of Appendix Figure A2 shows that even the 25th percentile of borrowers in the sample have higher incomes than the US household median of $50,000. Panel (b) of Appendix Figure A2 shows that the average unused limit on credit cards is above $14,000, with most households having access to at least several thousand dollars of credit.

ence heterogeneity or differential frictions. Instead, payment sensitivity provides a parsimonious explanation for observed borrower choices.

5 Mechanisms

What drives payment sensitivity and the observed bunching pattern? Forward-looking optimizing agents with access to some borrowing, represented by the NPV sensitivity benchmark, cannot generate the observed bunching pattern with more missing mass above the 80% LTV threshold. As discussed in Section 4, augmenting the benchmark model with adjustment frictions or preference heterogeneity cannot reconcile the bunching result with the benchmark model.

This section develops a dynamic consumption-savings model to test other mechanisms that could generate the observed bunching pattern. The model incorporates liquidity constraints, preference heterogeneity, and quasi-hyperbolic discounting. Using the model, I find that even extreme liquidity constraints cannot generate the empirical bunching pattern, nor can standard formulations of present-bias. Instead, the observed bunching and payment sensitivity at the LTV thresholds can be generated only by very particular preferences: borrowers must highly weight utility in the initial and some temporally near periods, but the weight on utility on more distant periods must fall dramatically, and within a narrow window. These preferences can be operationalized by modified $\beta-\delta$ quasi-hyperbolic discounting, where the multiplicative factor β affects only periods after some specified period t_1, instead of all periods after the initial period.

5.1 Consumption-Savings Model

The model has agents making decisions over a finite horizon T. Agents are endowed with initial assets a_0 and have fixed, deterministic income y. Agents have constant relative risk aversion utility with heterogeneous risk aversion parameter γ. In each period t agents choose consumption c_t. In the initial period, they have an additional choice variable L that represents the LTV choice when purchasing a house of fixed value H. There is no other borrowing besides the initial mortgage choice.

The agent's problem is:

$$\max_{c_t, L} \quad U(c_0) + \beta \left[\sum_{t=1}^{T} \delta^t U(c_t) \right] \tag{15}$$

$$U(c) = \begin{cases} \frac{c^{1-\gamma}-1}{1-\gamma} & \text{if } \gamma \neq 1 \\ ln(c) & \text{if } \gamma = 1 \end{cases} \tag{16}$$

The budget constraints are:

$$a_1 = y + (1 + r_{\text{return}}) \cdot \left(a_0 - c_0 - \overbrace{(1 - L) \cdot H}^{\text{Down payment}} \right) \tag{17}$$

$$a_{t+1} = y + (1 + r_{\text{return}}) \cdot \left[a_t - c_t - \overbrace{P_t(L, r, H)}^{\text{Recurring payment}} \right] \tag{18}$$

$$a_t \geq 0 \quad \forall t \tag{19}$$

where a_t represents assets at time t, r_{return} is the interest rate on saving, y is income, and $P_t(L, r, H)$ is the mortgage payment at time t, where r is the interest rate on the loan. Payments are a function of LTV L, allowing for MI policy at thresholds. The exponential discount parameter is δ, and β is a multiplicative factor that allows for quasi-hyperbolic discounting.

The model is solved by dynamic programming, with two state variables a_t and P_t and two choice variables c_t and L. L is chosen only in the first period and affects the entire stream of expected P_t. The model is solved for a range of agents chosen to represent a typical loan, with parameter values as shown in Table 7. The values of MI_{80}, MI_{85}, and MI_{90} represent the annual MI rate above the 80%, 85%, and 90% LTV thresholds, respectively. For computational simplicity, the model uses $T = 30$ years instead of $T = 360$ months. To account for MI terminating part-way through a year, loan payments are calculated based on a 360-month schedule, and then averaged to yield the annual periodic payment P_t.

5.1.1 Model Results: Liquidity Constraints

The observed bunching pattern in Section 3 is inconsistent with standard, forward-looking models of optimizing behavior. Neither complete markets with no liquidity constraints, nor the other extreme of no outside borrowing, can rationalize the bunching results with forward-looking agents that evaluate the full stream of repayments. The NPV sensitivity benchmark is a sufficient statistic for the total costs of borrowing under complete markets and a useful predictor of optimizing behavior when agents have some ability to move consumption across time. However, it is inconsistent with

Table 7: Model Parameters

Parameter	**Value(s)**
Home Price	$400,000
T	30
δ (exponential discounting)	0.95
β (quasi-hyperbolic discounting)	1.0
r_{return}	3%
r	4.2%
γ	$\{0.5, 0.75, 1, 2, 4\}$
Income	$[50000, 100000]$
Initial assets	$[60000, 160000]$
$MI_{80}, MI_{85}, MI_{90}$	$0.5\%, 0.61\%, 0.78\%$

Note: The table shows the parameters used in model estimation. The model is solved for all possible combinations of parameters shown in the table. The risk aversion parameter γ takes one of the five unique values shown. Possible values for income and initial assets are multiples of $5,000 within the shown ranges.

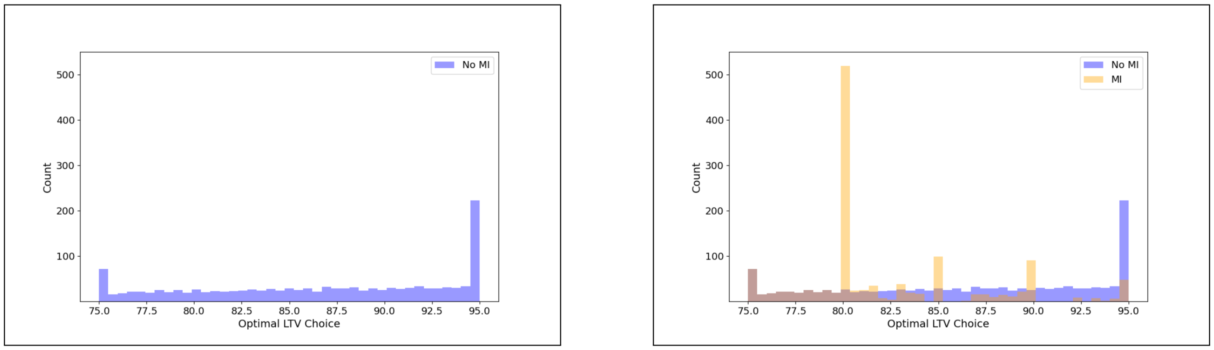
the empirical bunching pattern. Meanwhile, the model described above demonstrates that extreme liquidity constraints cannot explain the bunching pattern either.

Figure 11 shows the results from solving the model with extreme liquidity constraints - the only source of credit is the mortgage itself. Panel (a) shows the distribution of optimal LTV choices in the baseline case with no MI. Some agents choose corner solutions at the exogenously imposed borrowing limits of 75% and 95% LTV. Otherwise, the distribution of optimal LTV choices is smooth through the thresholds. Panel (b) shows the distribution of optimal choices after the introduction of the MI policy. The MI policy generates bunching at the thresholds, but is unable to reproduce the empirical bunching result of more missing mass above the 80% threshold. Instead, 0.96 of the expected mass in the 90-93% LTV window is missing relative to baseline, compared to just 0.04 missing mass in the 80-83% LTV window.

In general, it is difficult to generate the observed bunching pattern with liquidity constraints alone if agents are forward looking. The discontinuous increase in total future borrowing costs is much larger at the 90% than the 80% LTV threshold. Forward-looking agents would experience larger disutility when exceeding the 90% threshold, and thus bunch with more missing mass above the 90% threshold. In order to rationalize the empirical pattern of less missing mass above the

90% threshold, agents near the 90% threshold would need to have much larger adjustment costs or liquidity constraints at the 90% threshold. However, as shown in Section 4, these frictions would have to be implausibly large in both absolute and relative terms.

Figure 11: Distributions of Optimal LTV Choices



(a) Baseline **(b) With Mortgage Insurance**

Note: The figure shows the distribution of optimal initial loan-to-value ratio (LTV) choices for agents solving equation 15 with model parameters in Table 7 and exogenous minimum and maximum borrowing limits of 75%/95% LTV. Panel (a) shows the distribution of optimal LTV choice in the absence of pricing discontinuities at LTV thresholds. Panel (b) adds the distribution of optimal LTV when agents optimize in the presence of mortgage insurance (MI) requirements at the 80%, 85%, and 90% LTV thresholds.

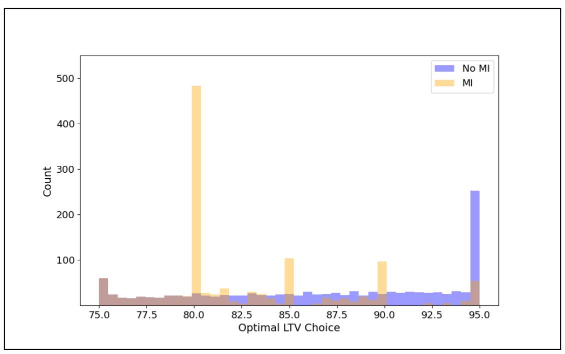
5.1.2 Model Results: Present-Bias

If agents were less forward-looking and more impatient, they might generate the observed bunching pattern in response to MI policies. However, the model shows that simple present-bias alone cannot explain the results, and that preferences must take a very particular form to explain observed bunching. Agents in the model optimize over the discounted flow utility from consumption over time. As an illustrative example, consider the simple model of present-bias from Laibson (1997) in the form of quasi-hyperbolic discounting, under which utility in all but the initial choice period is discounted by the multiplicative factor β.

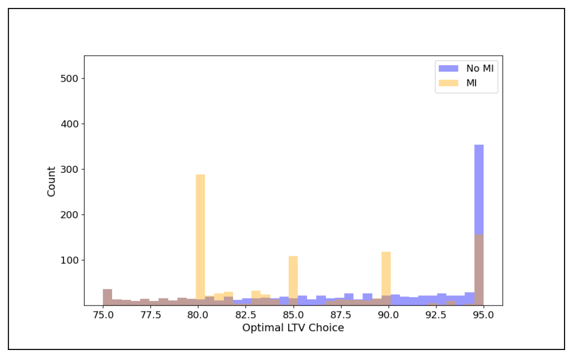
This formulation of present-bias, where all future periods are discounted by $\beta\delta^t$ instead of δ^t as in the exponential discounting model, cannot qualitatively change the observed bunching pattern. MI imposes costs in the form of higher future payments, but has no direct effect on consumption in the initial period. Hence, the problem for present-biased agents is equivalent to the problem for patient agents, with the disutility from future payments scaled by β. The relative cost of exceeding the 80% threshold thus remains smaller than exceeding the 90% threshold and predicts less missing mass above 80%, contrary to the observed bunching pattern.

Figure 12 shows that both moderate and substantial present-bias fail to generate the empirical bunching pattern. Panel (a) shows the optimal LTV choices for agents with $\beta = 0.9$, with the purple bars representing the distribution with no MI and the orange bars showing the distribution with MI. Present-bias causes more borrowers to locate at the 95% LTV corner solution compared to patient agents with $\beta = 1$. However, the pattern of missing mass is unchanged from that of patient agents: the missing mass relative to the purple, no-MI counterfactual above the 80% threshold is 0.02, and the missing mass above the 90% threshold is 0.97. Panel (b) similarly shows the no-MI and with-MI distributions for agents with $\beta = 0.6$. In this case, even more agents choose to locate at the corner solution of 95%. In addition, the pattern of missing mass is even starker: there is 0.27 *excess* mass relative to the counterfactual above the 80% threshold, while there is still 0.95 missing mass above the 90% threshold.

Figure 12: Distributions of Optimal LTV Choices



(a) $\beta = 0.9$



(b) $\beta = 0.6$

Note: The figure shows the distribution of optimal initial loan-to-value ratio (LTV) choices for agents solving Equation 15 with exogenous minimum and maximum borrowing limits of 75%/95% LTV. Model parameters are as in Table 7 except for the β parameter. Panel (a) shows the distribution of optimal LTV choice without and with the MI requirements for agents with $\beta = 0.9$. Panel (b) presents results from the same exercise with $\beta = 0.6$.

5.2 Necessary Conditions on Preferences

While simple present-bias cannot generate the empirical bunching pattern, some very specific preferences can do so. If borrowers have high utility weight on the initial period and some early repayment periods, and these weights decline rapidly at precise times, they would react to MI by bunching in the observed manner. In this case, borrowers have preferences such that the discounted utility cost of exceeding the 90% LTV threshold is no larger than the cost of exceeding the 80% threshold. Suppose borrowers have disutility over future payments with the following discounting structure:

$$-\left[\sum_{t=1}^{t_1}\delta^t P_t\right]-\beta\left[\sum_{t=t_1+1}^{T}\delta^t P_t\right] \tag{20}$$

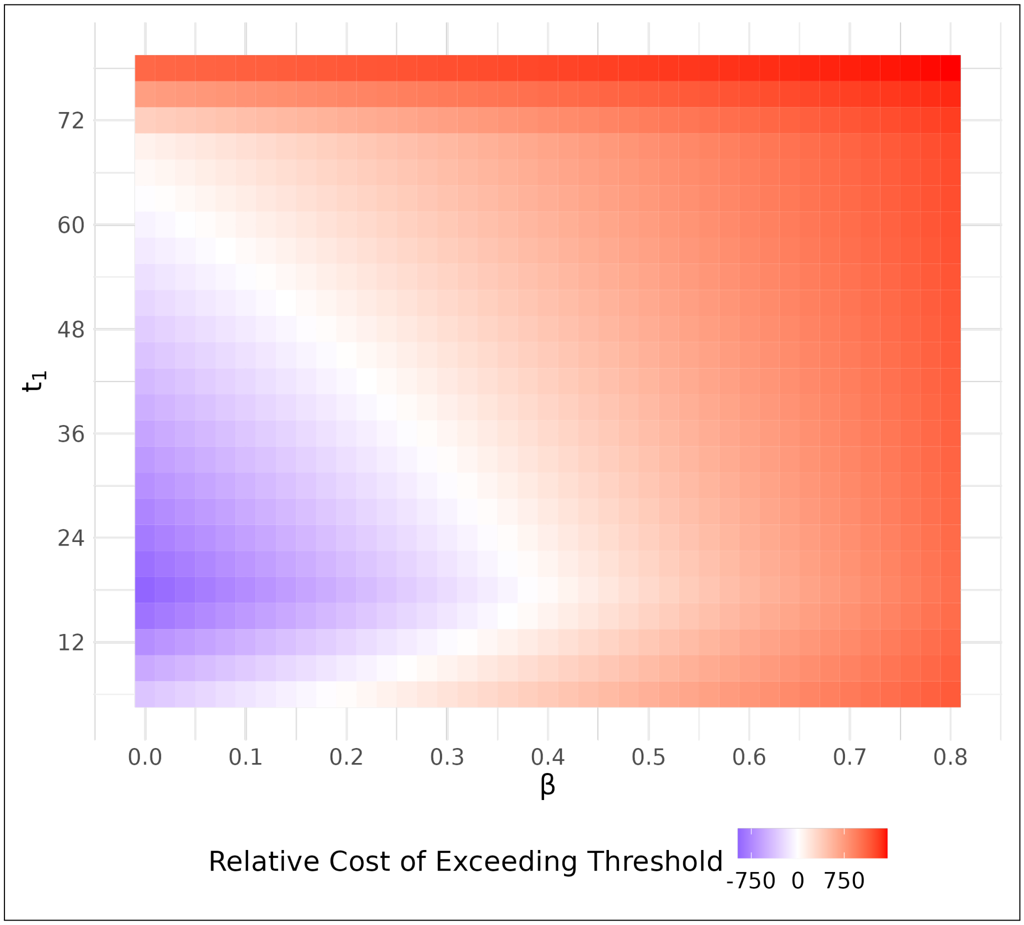
This formulation uses the simplifying assumption that the discounted flow consumption loss from making payments is a sufficient measure of borrower disutility in a period. Borrowers discount periods after the initial period at the standard exponential discount rate δ^t. At some more distant time t_1, hyperbolic discounting "turns on" and periods after t_1 are discounted by a factor of $\beta\delta^t$.

Compare two representative borrowers located around the 80% and 90% LTV thresholds using the discounting scheme above. The MI rate step-ups are 50bps and 11bps respectively at the 80 and 90% thresholds, and the loan rate step-up is 6bps at both thresholds. For each borrower and each possible set of values (β, t_1) compute the increase in the discounted value of payments when locating at 82% vs 80% LTV, and 92% vs 90% LTV.

To rationalize the empirical bunching pattern, the discounted cost of exceeding 90% LTV must be no bigger than the cost of exceeding 80% LTV. This is true only when β is low and, critically, "turns on" at very precise times. Figure 13 compares the cost of exceeding both thresholds for various values of β and t_1. Areas in purple indicate a combination of (β, t_1) under which the discounted cost of raising LTV by two points at the 90% threshold is less than the cost of doing the same at the 80% threshold. As the figure shows, β must generally be below 0.5 for this to be true. In addition, the value of t_1 must also fall within a narrow range depending on the value of β. This result is driven by two factors: β has to "turn on" before the automatic MI termination date for the borrower at or above 90%, as it must lower the utility cost of some MI payments for the 90+% LTV borrower. However, β cannot "turn on" too early, as doing so would lower the cost of exceeding the 80% threshold too much.

Borrowers must have highly non-standard preferences for the discounted cost of exceeding 90% LTV to be less than the cost of exceeding 80% LTV and to explain the relative lack of missing mass above 90% LTV. For example, Figure 13 shows that, at $\beta = 0.3$, t_1 must be within 12 and 36 months (in the purple and white areas). This implies that the effect of the additional MI payment on per-period flow utility must drop by 70% sometime between month 12 and 36. Suppose an agent has CRRA utility and consumes c_0 in the periods before t_1. The MI payment reduces their

Figure 13: Relative Cost of Exceeding 80% and 90% LTV Thresholds



Note: The figure shows the difference in costs of exceeding the 80% and 90% LTV thresholds for a representative borrower. The cost of exceeding a threshold is the increase in the value of payments owed when locating two LTV points above the threshold. Payments are discounted by an exponential discount factor of δ^t with $\delta = 0.95$, and additionally by a multiplicative factor of β for periods $t > t_1$. The figure plots (cost of exceeding 90%) - (cost of exceeding 80%); purple regions indicate that exceeding the 90% threshold is more costly. The property value is $250,000 and interest rate is 4.2%. Additional mortgage insurance when exceeding 80/90% LTV is 50/11bps.

consumption by some amount X, and reduces utility by $U'(c_0)X$. Then after t_1 their consumption must have grown to c_1 such that the effect of making the same payment X reduces utility by only 0.3 times as much. This implies $0.3U'(c_0)X = U'(c_1)X$, or $\frac{c_1}{c_0} = 0.3^{-1/\gamma}$. For risk aversion of $\gamma = 2$, consumption c_1 must be 82.5% higher than c_0, and that this growth in consumption must occur around t_1 and not at any other time.

Rationalizing the relative lack of bunching mass above 90% based on preferences or changes in marginal utility must explain how the disutility of exceeding 80% LTV is at least equal to disutility from exceeding 90%. Exceeding the 80% threshold entails a larger increase in nominal MI payments, but for a shorter duration and with smaller total nominal value. The example above requires the disutility from payments to drop sharply at some period. The instantaneous change is not strictly required; however, any potential explanation must generate high disutility costs from making MI payments in some early periods that eventually decline. In addition, the decline cannot occur too early in the payment schedule.

5.3 Summary: Payment-Sensitivity Requires Non-Standard Preferences

It is difficult to rationalize observed bunching with borrowers optimizing over the NPV of payments. Explanations that rely on frictions or down-payment constraints have to explain borrowers choosing to implicitly borrow at extremely high effective rates, often exceeding the cost of unsecured credit card borrowing, even while many borrowers have high income and ample access to liquidity. In addition, these frictions must be differentially higher for borrowers locating above the 90% threshold than borrowers locating above the 80% threshold, for which there is no strong evidence.

For the observed bunching to be explained by preferences, it must be the case that the fewer, earlier, larger MI payments with smaller total nominal value when exceeding 80% LTV generate greater disutility than more payments of smaller size, over a much longer period and with higher total nominal value when exceeding 90% LTV. Agents must have high disutility from payments in some early periods that declines sizably over time, but not immediately. This condition rules out certain preferences and biases as potential drivers of payment sensitivity. For example, simple present-bias where all future periods are discounted by β would not generate the required pattern.

The bunching evidence strongly suggests payment sensitivity: borrowers act as if they optimize over immediate liquidity and the size of the initial or early monthly payments. This sensitivity to nominal recurring payments is consistent with multiple explanations. Reference-dependence could generate the required pattern of disutility from increased payments that declines over time. Complexity aversion could drive borrowers to optimize over nominal recurring payments as opposed to grappling with the full schedule of payments that can vary over time. Inattention to changes in the payment schedule could cause borrowers to over-estimate the cost of exceeding 80% LTV. Mental accounting could generate payment sensitivity as borrowers target a specific payment amount for shelter costs. It is beyond the scope of this paper to fully discern between the many possible mechanisms that drive payment sensitivity; nonetheless, these mechanisms cause agents to respond much more strongly to initial monthly payments than to total loan costs.

6 Implications for Regulation and Monetary Policy

Borrowers' sensitivity to recurring payments has significant implications for the design of financial products, financial regulation, and the transmission of monetary policy. If some borrowers are sensitive only to the recurring payment and not the total value of payments, lenders can exploit this behavior by shrouding total loan costs. For example, adjustable-rate mortgages or interest-only mortgages shrink early recurring payments by loading larger payments in the future, while extending the duration of an amortizing loan lowers the recurring payment by spreading payments across more periods. Payment sensitive types might over-borrow when offered these products, potentially leading to an inefficient equilibrium of the type described by Gabaix and Laibson (2006).

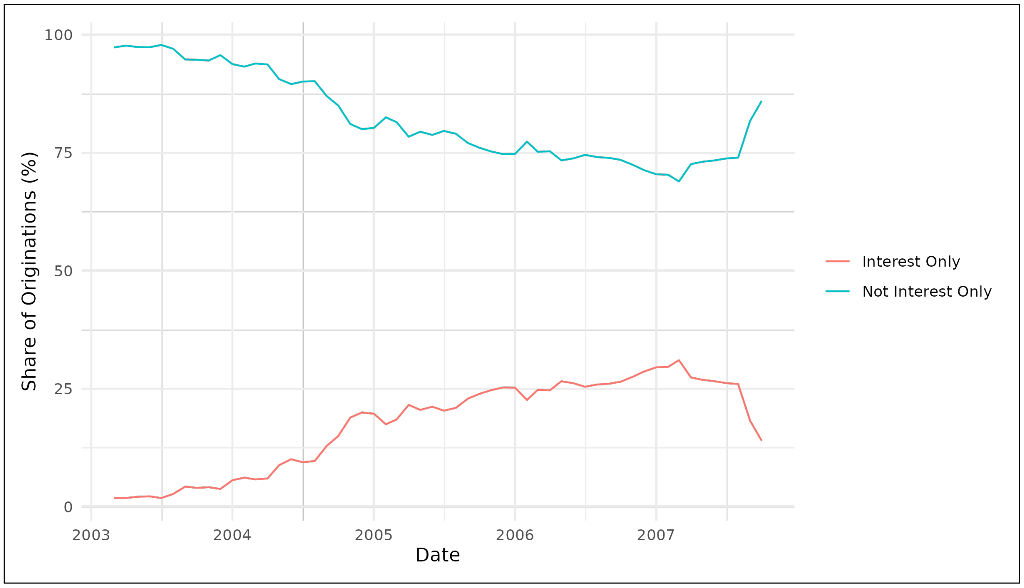
I illustrate the potential magnitude of these effects by examining the US mortgage market prior to the Great Recession. US housing prices rose rapidly in the early 2000s, with the most significant growth occurring between 2004 and 2007. The S&P CoreLogic Case-Shiller U.S. National Home Price Index increased by approximately 30% during this period (CoreLogic, 2024), while mortgage origination volumes grew rapidly during this period from \$1.14 trillion in 2000 to a high of \$3.03 trillion in 2005 (Mortgage Bankers Association, 2023). Much of this growth occurred after the Federal Reserve initiated a rate-hiking cycle in mid 2004.

Interest-only (IO) mortgages grew in popularity during this period of rising home prices and interest rates. Figure 14 shows that prior to 2004 IO loans were mostly non-existent. However, as the rate-hiking cycle began in 2004, the share of IO mortgage originations rose, reaching more than 25% by the time the rate-hiking cycle ended in 2006 Q3. A number of papers have found that the growth of non-traditional loans such as IO mortgages during this period could have contributed to a speculative bubble in housing prices by increasing borrowing (Amromin et al. (2018), Barlevy and Fisher (2021), Dokko et al. (2024)).

Payment sensitivity provides a possible explanation for *how* non-traditional loans expand borrowing: IO mortgages are non-amortizing and have lower monthly payments for a given loan amount compared to a more typical amortizing loan or, equivalently, permit additional borrowing while keeping the monthly payment unchanged. IO loans are thus particularly attractive to payment sensitive borrowers. In the extreme case where a borrower is responsive solely to the recurring payment, an IO loan with the same payment as an amortizing loan allows much more borrowing at the same apparent cost. As such, expansion of IO loans may have enabled excessive borrowing during this period of rapidly growing house prices.

I approximate the size of additional borrowing from IO loans and payment-targeting. Data from McDash show that over 97% of IO loans had 30-year terms and the median IO rate was around

Figure 14: Share of Interest Only Mortgage Originations



Note: The figure shows the share of interest-only loan originations. Data are from ICE, McDash and include all mortgages in the data for 1-4 unit properties.

6.3%. A 30-year, fixed-rate fully amortizing loan with the same payment and same interest rate as an IO loan would entail a 15.2% lower initial borrowing amount. In 2005, 20% of the \$3 trillion in mortgage originations were IO loans. Put together, this implies that substitution from amortizing to IO loans allowed payment sensitive borrowers to take on 20% $\times$ \$3 trillion $\times$ 15.2% = \$91.2 billion per year in additional mortgage debt, representing 3.04% of the entire mortgage origination market. This calculation does not account for the fact that IO mortgages typically had lower rates that amortizing loans, which would have permitted even more additional borrowing.

Payment sensitivity has implications for a wide range of financial regulation and monetary policy, including but not limited to mortgage finance. Financial products that shift or spread payments across time and obscure total loan costs are common: Argyle et al. (2020) document evidence of increased auto-loan borrowing due to the extension of loan duration, while Fannie Mae took steps towards buying and securitizing 40-year mortgages prior to the Great Recession in a bid to capture more marginal borrowers (Newswires, 2005). While regulations such as the Qualified Mortgage and Ability-to-Repay rules under the Dodd-Frank Act have taken steps to mitigate potentially harmful lending practices, regulators may need stronger rules regarding disclosure of total costs to prevent exploitation of payment sensitivity, especially with the growth of non-traditional loan products such as buy-now-pay-later. Payment sensitivity also suggests a channel that can attenuate monetary policy: lenders can break the mechanical link between rising interest rates and the perceived cost of borrowing, potentially offsetting the effects of monetary tightening. Central banks may need to pay attention to these effects when modelling the effect of policy on consumer borrowing.

7 Conclusion

This paper presents evidence of recurring payment sensitivity in household borrowing, particularly in the context of mortgage choices. Borrowers respond to discontinuities in mortgage insurance requirements by bunching in their loan-to-value ratio (LTV) choice. The bunching pattern exhibits the greatest missing mass above LTV thresholds with the largest notches in initial monthly payment, and less missing mass above large notches in the net present value (NPV) of loan payments. The results are consistent with payment sensitive borrowers responding to discontinuous notches in recurring payments with respect to LTV, and inconsistent with NPV sensitive borrowers responding to notches in the NPV of loan payments with respect to LTV. A dynamic consumption-savings model shows that liquidity constraints or standard present-bias are unable to rationalize the bunching result with NPV sensitivity. Furthermore, preference-based explanations for observed bunching require borrowers to place high utility weight on some early periods, which then declines sharply several years in the future.

Payment sensitivity has important implications for understanding consumer financial behavior and the broader macroeconomic effects of financial product and loan design. If borrowers prioritize minimizing recurring payments, lenders can shroud borrowing costs by adjusting loan features like duration or interest rates without significantly altering monthly payments, which could in turn lead to excessive borrowing. Additionally, the focus on recurring payments suggests that traditional metrics for assessing consumer debt burden may underestimate the actual financial strain on borrowers, particularly if loan products are designed to minimize initial payments while increasing lifetime costs. Policymakers should consider these findings when conducting monetary policy and when evaluating consumer protection and lending regulations.

The paper has several limitations that suggest avenues for future work. First, the sample is by definition one of homeowners, who tend to be wealthier and higher-income than the general population. It would be useful to study if payment sensitivity is also present in choices over more everyday financial decisions than mortgages, and in lower-income or more credit-constrained populations. Second, while the paper tests some mechanisms that could generate the as-if pattern of payment sensitivity, it does not establish a single psychological or economic explanation for the behavior. Observed payment sensitivity in mortgage choice is consistent with several behavioral theories, including but not limited to complexity aversion, reference-dependence, or (rational) inattention. Further research could explore which behavioral drivers underpin payment sensitivity. Finally, this work could be expanded with a richer, fully calibrated model of payment sensitivity to quantitatively evaluate the effects of policies such as restrictions on types of loan products.

References

Agarwal, S., I. Ben-David, and V. Yao (2017). Systematic mistakes in the mortgage market and lack of financial sophistication. Journal of Financial Economics 123(1), 42–58.

Amromin, G., J. Huang, C. Sialm, and E. Zhong (2018). Complex Mortgages*. Review of Finance 22(6), 1975–2007.

Argyle, B. S., T. D. Nadauld, and C. J. Palmer (2020). Monthly payment targeting and the demand for maturity. The Review of Financial Studies 33(11), 5416–5462.

Bäckman, C., P. Moran, and P. van Santen (2024). Mortgage Design, Repayment Schedules, and Household Borrowing.

Barlevy, G. and J. D. M. Fisher (2021). Why were interest-only mortgages so popular during the U.S. housing boom? Review of Economic Dynamics 41, 205–224.

Bartlett, R., A. Morse, R. Stanton, and N. Wallace (2022). Consumer-lending discrimination in the fintech era. Journal of Financial Economics 143(1), 30–56.

Beshears, J., J. J. Choi, D. Laibson, and B. C. Madrian (2018). Behavioral Household Finance. Handbook of Behavioral Economics: Foundations and Applications 1 (Vol. 1).

Best, M. C., J. S. Cloyne, E. Ilzetzki, and H. J. Kleven (2020). Estimating the elasticity of intertemporal substitution using mortgage notches. The Review of Economic Studies 87(2), 656–690.

Bhutta, N. and B. J. Keys (2022). Moral hazard during the housing boom: Evidence from private mortgage insurance. The Review of Financial Studies 35(2), 771–813.

Board of Governors of the Federal Reserve System (2024). Commercial Bank Interest Rate on Credit Card Plans, All Accounts.

Board of Governors of the Federal Reserve System (U.S.) (2024). Survey of consumer finances. https://www.federalreserve.gov/econres/scfindex.htm.

Campbell, J. Y. (2006). Household Finance. The Journal of Finance 61(4), 1553–1604.

Chetty, R., J. N. Friedman, T. Olsen, and L. Pistaferri (2011). Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records *. The Quarterly Journal of Economics 126(2), 749–804.

CoreLogic (2024). S&P CoreLogic Case-Shiller U.S. National Home Price Index.

DeFusco, A. A. and A. Paciorek (2017). The Interest Rate Elasticity of Mortgage Demand: Evidence from Bunching at the Conforming Loan Limit. American Economic Journal: Economic Policy 9(1), 210–240.

Department of Housing and Urban Development (2014). Rental burdens, rethinking affordability measures. https://web.archive.org/web/20241101055019/https://www.huduser.gov/portal/pdredge/pdr_edge_featd_article_092214.html.

Dokko, J., B. J. Keys, and L. Relihan (2024). Affordability, Financial Innovation, and the Start of the Housing Boom.

Federal Reserve Bank of New York (2023). Household Debt and Credit Report. Technical Report 2022:Q4.

Follain, J. R. and R. M. Dunsky (1997). The demand for mortgage debt and the income tax. Journal of Housing Research, 155–199.

Fuster, A. and B. Zafar (2021). The Sensitivity of Housing Demand to Financing Conditions: Evidence from a Survey. American Economic Journal: Economic Policy 13(1), 231–265.

Gabaix, X. and D. Laibson (2006). Shrouded attributes, consumer myopia, and information suppression in competitive markets. The Quarterly Journal of Economics 121(2), 505–540.

Gathergood, J., N. Mahoney, N. Stewart, and J. Weber (2019). How Do Individuals Repay Their Debt? The Balance-Matching Heuristic. American Economic Review 109(3), 844–875.

Hanson, A. (2020). Taxes and Borrower Behavior: Evidence from the Mortgage Interest Deductibility Limit. Journal of Urban Economics 118, 103256.

Hastings, J. S. and J. M. Shapiro (2013). Fungibility and Consumer Choice: Evidence from Commodity Price Shocks*. The Quarterly Journal of Economics 128(4), 1449–1498.

Jappelli, T. and L. Pistaferri (2007). Do people respond to tax incentives? An analysis of the Italian reform of the deductibility of home mortgage interests. European Economic Review 51(2), 247–271.

Juster, F. T. (1964). Consumer Sensitivity to the Price of Credit. The Journal of Finance 19(2), 222–233.

Keys, B. J., D. G. Pope, and J. C. Pope (2016). Failure to refinance. Journal of Financial Economics 122(3), 482–499.

Kleven, H. J. (2016). Bunching. Annual Review of Economics 8(1), 435–464.

Kleven, H. J. and M. Waseem (2013). Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan *. The Quarterly Journal of Economics 128(2), 669–723.

Laibson, D. (1997). Golden Eggs and Hyperbolic Discounting*. The Quarterly Journal of Economics 112(2), 443–478.

Mortgage Bankers Association (2023). Mortgage origination estimates.

Newswires, D. J. (2005). Fannie Mae Plans to Increase Buying of 40-Year Mortgages. Wall Street Journal.

Saez, E. (2010). Do Taxpayers Bunch at Kink Points? American Economic Journal: Economic Policy 2(3), 180–212.

Stango, V. and J. Zinman (2009). Exponential growth bias and household finance. The Journal of Finance 64(6), 2807–2849.

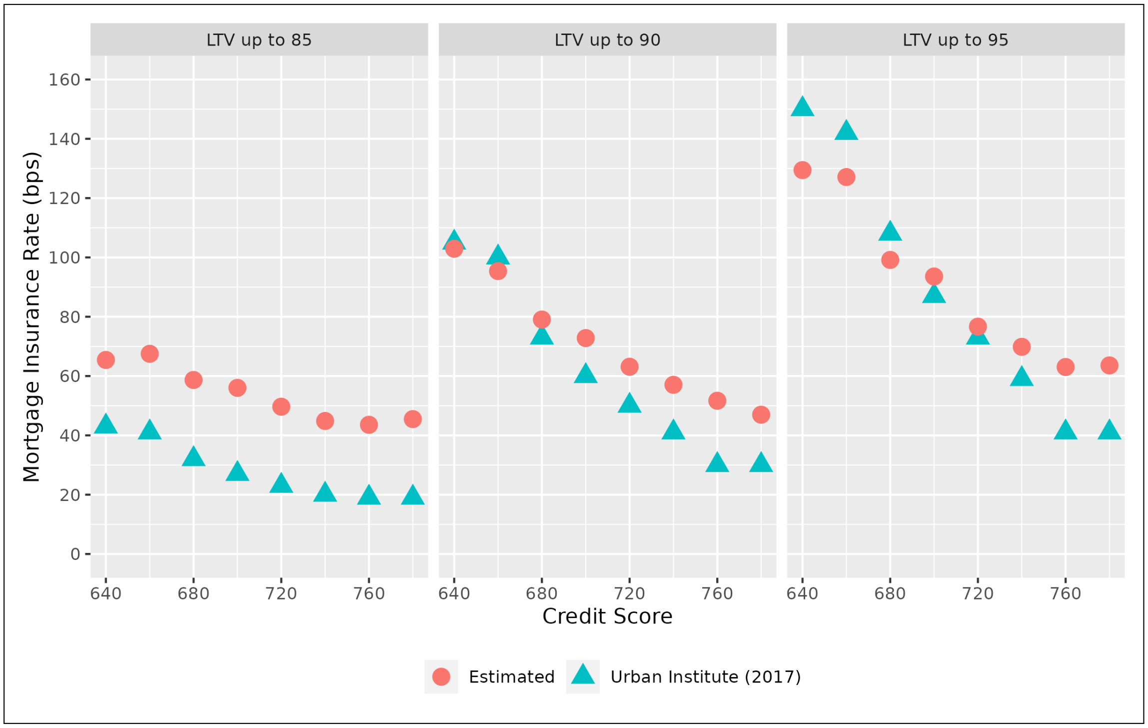
Stango, V. and J. Zinman (2023). We Are All Behavioural, More, or Less: A Taxonomy of Consumer Decision-Making. The Review of Economic Studies 90(3), 1470–1498.

Urban Institute (2017). Mortgage Insurance Data at a Glance - 2017 | Urban Institute. Technical report.

Urban Institute (2024). Housing Finance: At a Glance Monthly Chartbook, March 2024 | Urban Institute. Technical report.

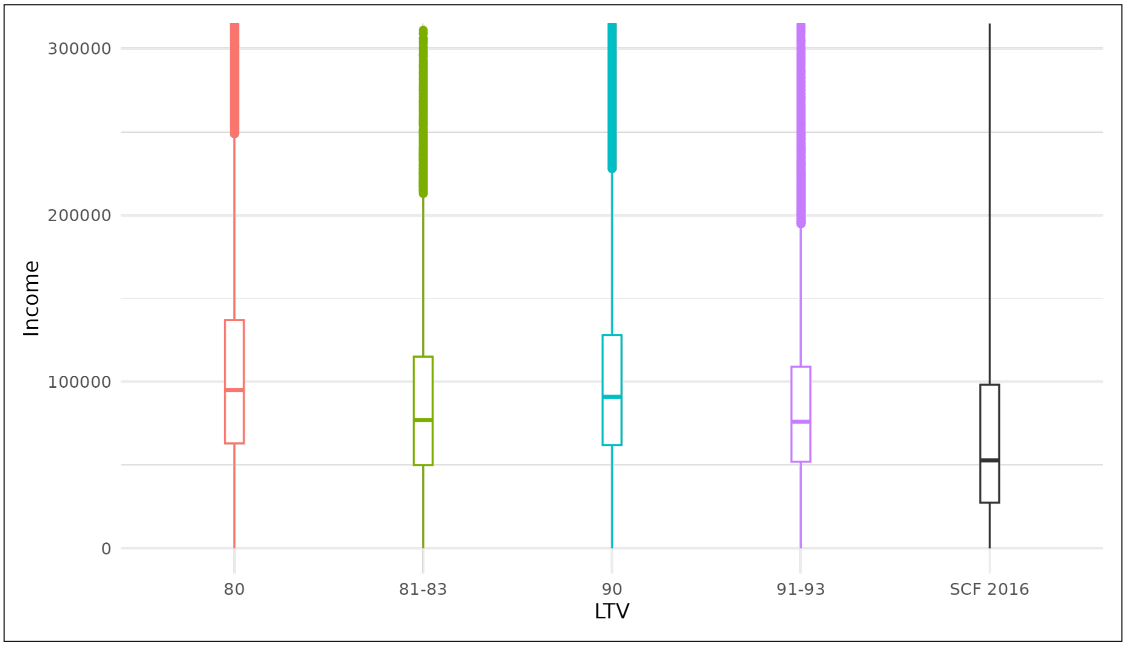
Appendix A Supplementary Figures

Figure A1: Estimated Mortgage Insurance Premia

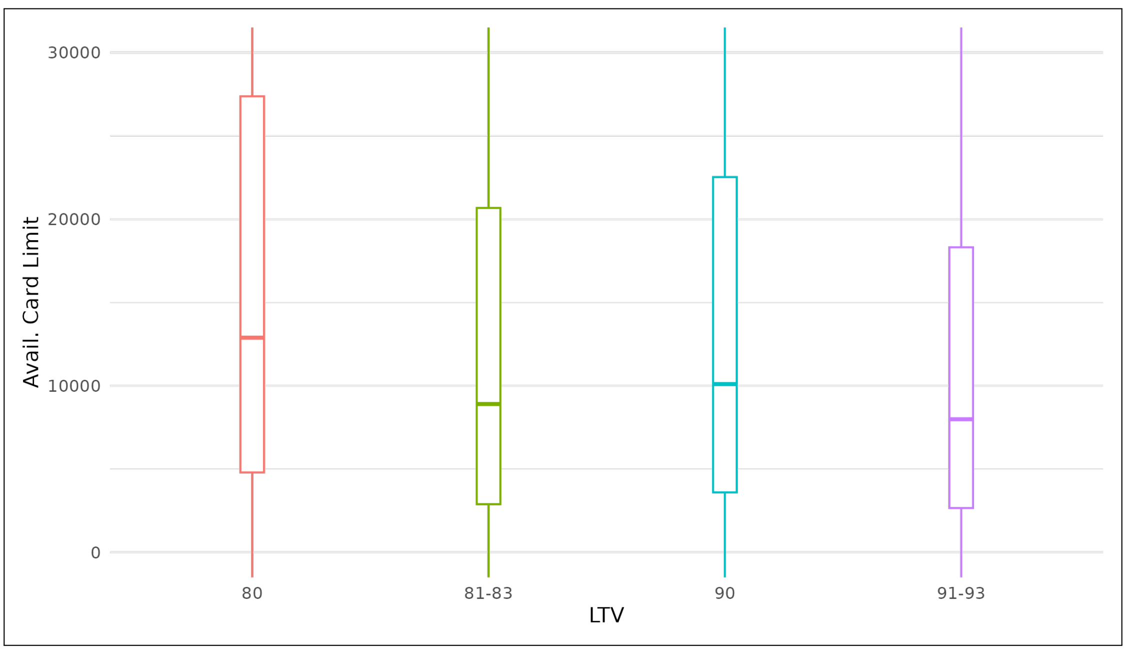


Note: The figure shows the estimated average mortgage insurance premium faced by borrowers in various 5-point LTV bins and 20-point credit score bins. The red circles are imputed from the Taxes and Insurance Payments in the McDash data, while the blue triangles report estimates from the Urban Institute (2017).

Figure A2: Distributions of Income and Available Credit Card Limits at LTV Thresholds



(a) Income



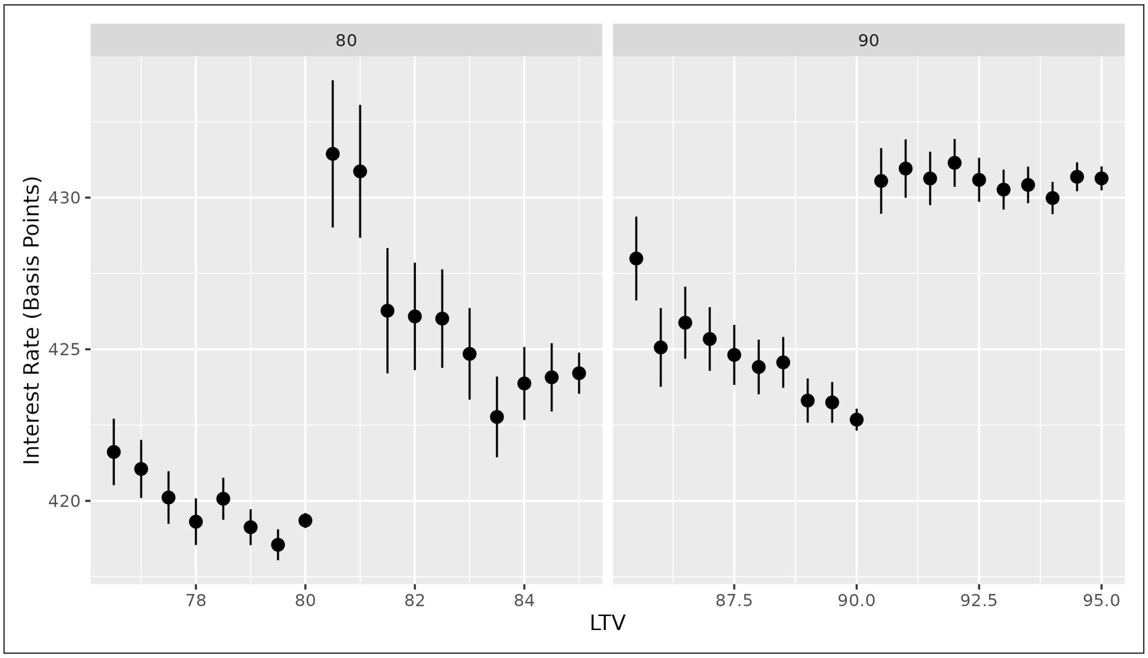
(b) Available Card Limits

Note: Panel (a) shows the distribution of borrower income at or above LTV thresholds, as well as the distribution of income in the 2016 Survey of Consumer Finances from the Board of Governors of the Federal Reserve System. Panel (b) shows the distribution of available credit card limits, defined as the difference between credit limits and credit balances. Results for '80' and '90' represent borrowers with LTV in $(79-80]$ and $(89-90]$, respectively, while results for '81-83' and '91-93' represent borrowers located above the threshold up to and including three LTV points above the threshold.

Appendix B Effect of LTV Thresholds on Rates

Mortgage Insurance (MI) is the main driver of differences in initial payments and NPV of payments across the LTV thresholds. However, loan interest rates also change at the same LTV thresholds due to pricing policies from the GSEs and lenders. Figure B3 shows that the spread in rates when crossing both the 80% and 90% LTV thresholds is about 11bps.

Figure B3: Interest Rate Spread



Note: The figure shows the average interest rate for borrowers in various LTV bins.

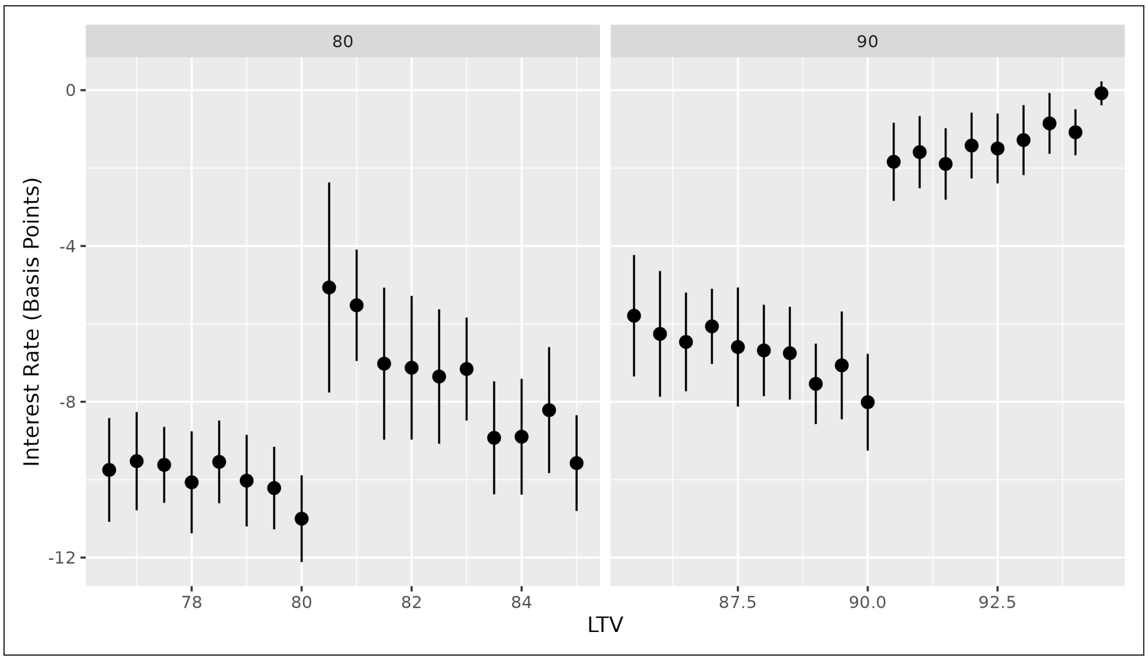
The spread could be due to the causal effect of exceeding an LTV threshold, or due to selection effects whereby high-quality borrowers move to the threshold, leaving lower-quality borrowers above the threshold. To isolate the causal effect from selection effects, I perform a regression of interest rate on borrower characteristics separately at each threshold:

$$\text{RATE}_{it} = \beta_1 \text{LTV}_i + \sum_{k=1}^{3} \beta_{2k} \text{DTI}_i^k + \beta_5 \text{SCORE}_i + \beta_6 \text{ZIP}_i + \beta_7 \text{QTR}_t + \varepsilon_i \tag{21}$$

where LTV is a vector of dummies for half-point LTV bins, DTI is debt-to-income ratio, SCORE is a vector of dummies for 20-point bins of credit score, ZIP and QTR are fixed effects for the ZIP code of the property and the quarter of origination. The conforming loan market is very liquid and the GSEs have standardized pricing adjustments that depend on credit score and LTV. As such, it is likely that the controls as specified account for most of the selection across the thresholds, and that the β_1 coefficients mostly represent the causal effect of the threshold on LTV.

Figure B4 shows the β_1 coefficients for each threshold. The regression-adjusted spread at each threshold is about 6bps, and there appears to be no significant difference in this spread at the 80% and 90% thresholds.

Figure B4: Regression-Adjusted Interest Rate Spread



Note: The figure shows the regression-adjusted effect of LTV on interest rate for borrowers in various LTV bins.

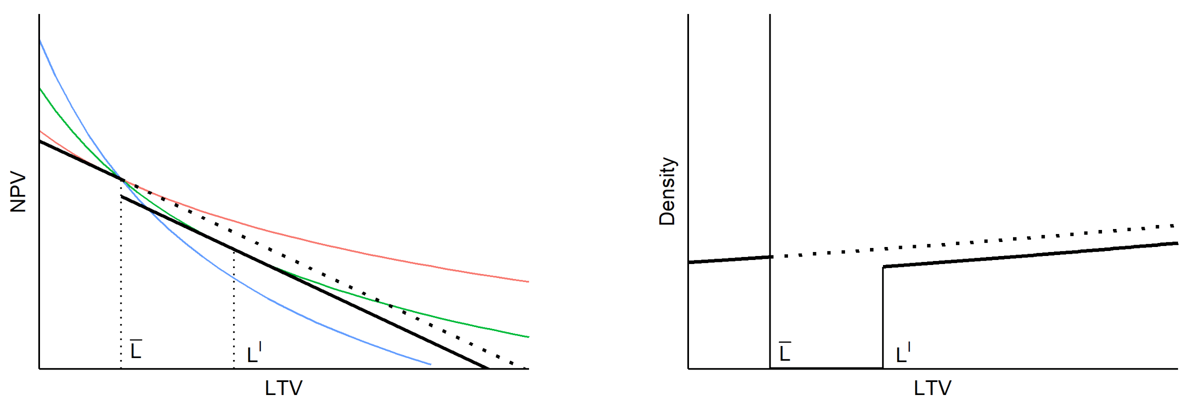
Appendix C Bunching and Notch Size

Mortgage Insurance induces a kink in the net present value (NPV) of payments at the 80% threshold and a notch at the 90% threshold. However, taking into account the effect of interest rate changes at LTV thresholds, there is a small notch in NPVs at the 80% threshold and a larger notch at the 90% threshold. As such, it is useful to develop predictions of the amount of missing mass as a function of the size of the notch in the budget constraint.

Figure C5 shows the baseline effect of a notch. Prior to the introduction of the notch, the smooth continuum of bunchers represented by indifference curves of various colors generates a smooth distribution of LTV choices. The red indifference curve is tangent to the budget constraint at $\bar{L}$. When the notch is introduced, the green buncher is the marginal buncher - she is indifferent between $\bar{L}$ and L^I. The notch induces missing mass in the region $[\bar{L}, L^I]$

Now consider the introduction of a larger notch at $\bar{L}$ as show in Figure C6. Now, the blue buncher is the new marginal buncher. The new indifference point L^I is farther to the right, and the area of missing mass is larger. At the *same* threshold, the larger notch size always induces a larger area of missing mass. When comparing notches of different sizes at different thresholds, the result holds as long as the indifference curves are well-behaved and the difference in convexity of preferences is not too great.

Figure C5: Notch-Induced Bunching

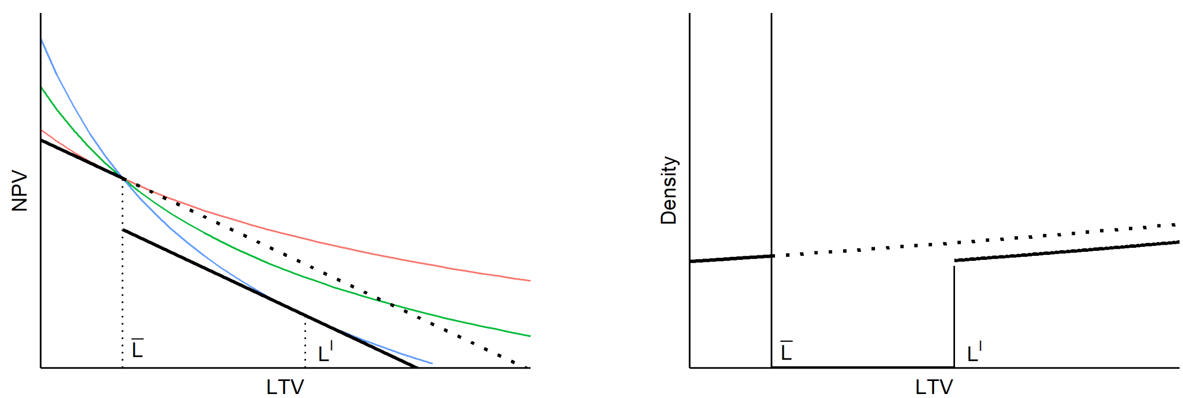


(a) Notch Budget Set

(b) Notch Density

Note: The figure shows the indifference curves for various borrowers and the missing mass induced by a small notch in the budget set.

Figure C6: Notch-Induced Bunching



(a) Notch Budget Set

(b) Notch Density

Note: The figure shows the indifference curves for various borrowers and the missing mass induced by a larger notch in the budget set.

Appendix D Bootstrap Standard Errors

This section describes the procedure for obtaining bootstrap standard errors of bunching estimates, and shows that the method is numerically identical to the method in Chetty et al. (2011).

D.1 Bootstrap in Chetty et al. (2011)

The estimation procedure in Chetty et al. (2011) estimates the counterfactual distribution of the outcome variable Y in the absence of the discontinuity by modelling Y as a polynomial of the running variable X. The regression equation is:

$$Y_i = \alpha + \beta_1 X_i + \beta_2 X_i^2 + \cdots + \beta_p X_i^p + \sum_k \delta_k D_k + \epsilon_i, \tag{22}$$

where Y_i is the count of observations in bin i, X_i is the value of running variable in bin i, D_k are indicator variables for each bin k within the excluded region, and ϵ_i is an error term. The coefficients δ_k allow the model to fit bin counts perfectly in the excluded region.

Equation 22 is estimated on the observed data to obtain the counterfactual distribution and the bunching statistics. The counterfactual polynomial used in computing bunching statistics $\widehat{Y_i^{\text{cf}}}$ are the predictions without the dummy terms:

$$Y_i^{\text{cf}} = \alpha + \hat{\beta}_1 X_i + \hat{\beta}_2 X_i^2 + \cdots + \hat{\beta}_p X_i^p \tag{23}$$

while the full predicted values are given by:

$$\hat{Y}_i = \alpha + \hat{\beta}_1 X_i + \hat{\beta}_2 X_i^2 + \cdots + \hat{\beta}_p X_i^p + \sum_k \hat{\delta}_k D_k \tag{24}$$

Note that $\hat{Y}_i = Y_i$ for observations within the excluded region, as the dummies allow for a perfect fit.

B bootstrap samples are then constructed by adding the $\hat{Y}_i$ to a sample (with replacement) of residuals from Equation 22. For each bootstrap sample $\mathbf{Y}_{\mathbf{b}}^{\text{boot}}$, Equation 22 is re-estimated to obtain a new counterfactual. This procedure yields B estimates of each bunching statistic. Finally, the standard errors of the bunching estimates are computed as the standard deviations of the bootstrapped bunching statistics.

D.2 Bootstrap in this paper

I define the excluded bins as $\mathbb{K}$ and the un-excluded bins as $\mathbb{J}$ and estimate the bunching counterfactual and compute bunching statistics using the following equation:

$$\min_{\{\beta_0,\beta_1,\beta_2,\beta_3\}} \sum_{i \in \mathbb{J}} \left(Y_i - \beta_0 - \beta_1 X_i - \beta_2 X_i^2 - \beta_3 X_i^3\right)^2 \tag{25}$$

This method yields identical estimates of the β terms and bunching statistics as the method in Chetty et al. (2011), since both procedures minimize the sum squared error for bin counts in the un-excluded region. The residuals for bins in the un-excluded region are also identical. Equation 25 yields predicted values for all bins:

$$\hat{Y}_i = Y_i^{\text{cf}} = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{\beta}_2 X_i^2 + \hat{\beta}_3 X_i^3 \tag{26}$$

To construct the bootstrap samples, first define:

$$Y_i^{\text{boot, predict}} = \begin{cases} \hat{Y}_i, & \text{if } i \in \mathbb{J} \\ Y_i, & \text{if } i \in \mathbb{K} \end{cases} \tag{27}$$

These $Y_i^{\text{boot, predict}}$ are identical to the $\hat{Y}_i$ predicted values from Chetty et al. (2011) in Equation 24. They take value Y_i for i in the excluded region, and value Y_i^{cf} for i in the un-excluded region.

To construct the bootstrap sample, I sample with replacement from the residuals of the initial estimation of Equation 25. Add the $Y_i^{\text{boot, predict}}$ to these residuals to obtain a bootstrap sample $\mathbf{Y^b}$. Finally, I re-estimate Equation 25 on the B bootstrap samples to obtain B bunching statistics. The standard errors of the bunching statistics are again given by the standard deviation of the B bootstrap bunching statistics.