Mortgage Prepayment and Path-Dependent Effects of Monetary Policy*

David Berger† Konstantin Milbradt‡ Fabrice Tourre§ Joseph Vavra¶

October 2018

Abstract

How much ability does the Fed have to stimulate the economy by cutting interest rates? We argue that the presence of substantial household debt in fixed-rate prepayable mortgages means that this question cannot be answered by looking only at how far *current* rates are from zero. Using a household model of mortgage prepayment with endogenous mortgage pricing, wealth distributions and consumption matched to detailed loan-level evidence on the relationship between prepayment and rate incentives, we argue that the ability to stimulate the economy by cutting rates depends not just on the level of current interest rates but also on their previous path: 1) Holding current rates constant, monetary policy is less effective if previous rates were low. 2) Monetary policy "reloads" stimulative power slowly after raising rates. 3) The strength of monetary policy via the mortgage prepayment channel has been amplified by the 30-year secular decline in mortgage rates. All three conclusions imply that even if the Fed raises rates substantially before the next recession arrives, it will likely have less ammunition available for stimulus than in recent recessions.

Keywords: Monetary Policy, Path-Dependence, Refinancing, Mortgage Debt

JEL codes: E50, E21, G21

*We would like to thank Ariza Gusti and Yang Zhang for excellent research assistance. We would also like to thank Erik Hurst, Andreas Fuster, Dan Greenwald, Pascal Noel, Amir Sufi, Amit Seru, Sam Hanson, Gadi Barlevy, Anil Kashyap, Arlene Wong, Greg Kaplan and Adi Sunderam. This research was supported by the Institute for Global Markets and the Fama-Miller Center at the University of Chicago Booth School of Business.

†Northwestern University and NBER; david.berger@northwestern.edu

‡Northwestern University and NBER; milbradt@northwestern.edu

§Copenhagen Business School; ft.fi@cbs.dk

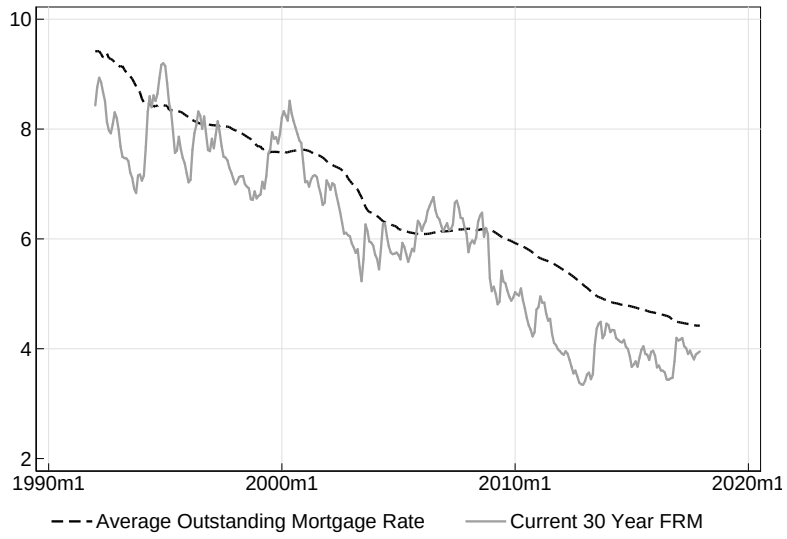
¶University of Chicago and NBER; joseph.vavra@chicagobooth.edu.

1 Introduction

How much room does the Fed have to stimulate the economy by lowering interest rates? At the end of 2015, the Federal Reserve ended its extended period of zero interest rates, and it has steadily increased rates since then. By mid 2018, the Federal Funds target rate had reached 2%, and the FOMC predicted that it would raise rates to above 3% in the near future. Obviously, the higher that interest rates rise, the more they can again be cut. However, in this paper we argue that looking only at *current* rates provides an incomplete view of Fed stimulative power, and that it may take an extended period of time with elevated rates before the Fed regains "ammunition" to stimulate the economy.

In particular, we argue that the presence of vast amounts of US household debt in the form of fixed-rate prepayable mortgages leads to path-dependent consequences of monetary policy and thus stimulus power which depends on both current and past rates. For example, suppose that the current interest rate is cut from 3% to 2%.[1] If rates were previously 3% for a long period of time, then many households have an incentive to refinance their mortgage debt, which can then lead to increases in spending. In contrast, if rates were previously below 2 % for a long period of time, then many households will have already locked in a low rate and will have no incentive to refinance in response to today's rate cut.

Figure 1: Outstanding vs. Current Market Mortgage Rates



The average outstanding mortgage rate is the average interest rate on all fixed-rate first mortgages calculated using BKFS McDash Monthly Performance data. The current 30 year FRM is the monthly average of the Freddie Mac weekly PMMS survey 30 year fixed rate mortgage average: https://fred.stlouisfed.org/series/MORTGAGE30US

Before describing our detailed analysis, we begin by illustrating the basic qualitative importance of path-dependence using simple aggregate time-series relationships. In particular, the solid gray line in Figure 1 shows the current 30-year fixed rate which can be obtained on *new* mortgages at a point in time. The dashed black line shows the average outstanding rate on the stock of mortgages which were originated in prior months.[2] This figure shows that when the current rate (solid) is below the average

[1] For illustrative purposes here we make no explicit distinction between short rates and mortgage rates and do not specify the extent to which Fed policy affects mortgage rates. We make these distinctions precise in our subsequent analysis which endogenizes the link between short rates and mortgage rates and delivers pass-through consistent with empirical estimates.

[2] A household refinancing from the average old rate to the current rate would thus jump from the dashed to the solid line.

outstanding rate (dashed), the outstanding rate converges rapidly towards the current rate, but the reverse is not true: when the current rate is high relative to the old locked in rate, few people refinance and convergence is very slow. This asymmetry leads to a distinct stair-step pattern with the outstanding rate only tracking the current rate when the former is above the latter and thus clearly demonstrating the qualitative presence of path-dependence in the mortgage market.

To provide a more precise and systematic evaluation of this path-dependence channel, we begin with a detailed empirical analysis using loan-level micro data, which we use to motivate a theoretical model of mortgage prepayment featuring endogenous borrowing, lending, consumption and pass-through of short-rates to long-rates. In both our model and in the data, the key feature driving the path-dependent effects of monetary policy is this observation that mortgages with positive "rate gaps" (the difference between the outstanding mortgage rate on a loan m^* and the current market rate m on similar mortgages) are much more likely to refinance. Holding m constant, the past history of rates will affect m^* and thus the response of prepayment rates to current rate changes.

While this is a simple observation, it delivers many insights for the consequences of monetary policy: 1) The strength of monetary stimulus has been substantially amplified by the secular decline in mortgage rates over the last 30 years, and we should anticipate less effective monetary policy in a stable or increasing rate environment. This is because the secular decline in mortgage rates has been a trend force encouraging refinancing and amplifying the strength of monetary policy 2) For similar reasons, in a stochastic but stationary rate environment, monetary policy is less effective after an extended period of low rates like we observed in the aftermath of the Great Recession. This is because if past rates were low, many households will have already locked in low rates, reducing current monetary policy ammunition. 3) It takes a very long time for the Fed to reload its ammunition after raising rates, yet it uses any accumulated ammunition rapidly when lowering rates. This is because households avoid prepaying when current rates are high and rapidly refinance when current rates are low. The remainder of the paper fleshes out these implications using a model of mortgage prepayment fit to a variety of detailed loan and individual level micro data.

Using micro data from Black Knight Financial Services, CoreLogic and Equifax spanning the period 1992-2017, we begin by documenting the relationship between the distribution of loan-level rate incentives and prepayment activity. Pooling across time, we calculate the overall distribution of rate gaps as well as the fraction of loans which prepay for a given rate gap. Overall, we find that there is a strong positive relationship between rate gaps and mortgage prepayment, even after controlling for a variety of other loan characteristics, household fixed effects and time-varying household characteristics.[3] Moreover, and importantly for our theoretical analysis, we find a sharp step in prepayment probabilities at exactly zero: loans with any positive rate incentive are significantly more likely to prepay than loans without such an incentive. This suggests that the fraction of loans with positive rate gaps is a useful summary statistic for the complicated distribution of rate gaps.

Turning to time-series evidence, we find this is the case: the fraction of loans with positive rate

[3]Most of our results focus on total prepayment since the distribution of rate gaps is determined by all prepayment and not just refinancing. However, one would expect that $frac > 0$ is particularly important for rate refinancing (as opposed to cash-out refinancing or prepayment due to moving houses). While prepayment cannot be decomposed using data from individual loans, from 2005-2017, we can link loans to households using Equifax CRISM data. This allows us to measure both which loans prepay and associate a prepaying loan with the (potential) new loan which is originated and so distinguish prepayment types. We find the sharpest effects of $frac > 0$ for rate refinancing.

gaps ($frac > 0$) in a given month both changes dramatically across time and strongly predicts the fraction of loans prepaying in that month.[4] Importantly, we show that $frac > 0$ is a stronger predictor of prepayment than any other threshold, such as the fraction of loans with at least a 50 basis point rate gap or the fraction of loans with at least a 100 basis point rate gap.[5] More surprisingly, we find that $frac > 0$ conveys almost all of the information contained in the *entire* shape of the gap distribution: including fully non-parametric controls for the shape of the rate distribution at a point in time has very little predictive power for prepayment after controlling for $frac > 0$. When turning to the theoretical implications of our empirical evidence, we rely on this result to substantially simplify our analysis.

While most of our empirical work focuses on aggregate prepayment activity, we really care about households' mortgage payments and spending behavior, not on prepayment per se. However, we find that average outstanding rates mirror prepayment behavior: when $frac > 0$ is large, more loans prepay and the average outstanding rate drops more rapidly. Moreover, as predicted by our theoretical analysis, we find that when $frac > 0$ is large, there is greater pass-through of current mortgage rate changes into the average outstanding mortgage rate. Finally, in order to explore implications for spending, we turn to regional analysis using local auto sales data from Polk. We find that regions with greater $frac > 0$ have 1) greater prepayment activity and 2) prepayment activity and auto sales which are more responsive to interest rate changes (even after controlling for both region and month fixed effects).[6]

What do these strong empirical relationships between mortgage rate gaps and prepayment imply for monetary policy? In order to explore this question, we turn to a theoretical model that can be used to assess a variety of effects and counterfactuals which cannot be measured directly in our data. In particular, we embed a simple model of mortgage prepayment into an incomplete markets model with endogenous mortgage pricing. We intentionally focus on a simple model of mortgages which includes only the minimal elements necessary to generate path-dependence. This isolates the basic mechanism which generates path-dependent monetary policy from a host of other features of mortgage contracts and housing which are less essential for this result.

Most importantly, we capture the "state-dependent" relationship between prepayment and rate gaps in a simple manner by assuming that households follow a "Calvo-style" refinancing process: they can only refinance at Poisson arrival times, and will do so if the rate on their mortgage is above the current mortgage market rate.[7] This random process proxies for a variety of pecuniary and non-pecuniary costs of refinancing and it generates a simple random "step-hazard" of prepayment in which mortgages with positive rate gaps are prepaid at a constant but higher rate than mortgages with negative rate gaps. As we show formally, this is the simplest model of infrequent refinancing which still allows for refinancing decisions that depend on rate gaps, but our empirical results show that it nevertheless

[4] One might rightfully be concerned that $frac > 0$ is endogenous and that this relationship may not be causal, however we show that results are similar when instrumenting for $frac > 0$ using lagged high-frequency monetary policy shocks.

[5] We might have expected that a positive threshold (such as 50 or 100 basis points) for our sufficient statistic would have been a better predictor of prepayments due to the small refinancing costs related to appraisal, attorney documentation review and title insurance, but this is not the case in the data. Note also that our time-series statistical statement does not imply that loans with gaps between 0 and 50 basis points are more likely to adjust than loans with gaps greater than 50 basis points. Loans with larger gaps are (mildly) more likely to adjust.

[6] Specifications with month fixed effects also help alleviate concerns about endogeneity of monetary policy.

[7] We also abstract from housing choice, equity extraction and default by endowing all households with a simple non-defaultable console mortgage with constant balance. We later argue that these simplifications are unimportant for our conclusions.

has good explanatory power for actual prepayment patterns.[8] This empirically realistic dependence of individual refinancing decisions on individual rate gaps is then the crucial feature which generates aggregate path-dependence of monetary policy.

Importantly, in addition to delivering straightforward intuition, this simple form of state-dependence buys us substantial analytical tractability and allows us to break our model into two blocks: a mortgage refinancing component and a consumption-savings component. Our setup implies that household refinancing decisions are orthogonal to consumption-savings decisions, which means that the mortgage refinancing block of the model can be analyzed on its own and does not depend on households' preferences, labor income characteristics, borrowing constraints or wealth. The mortgage block of the model pins down the equilibrium mortgage interest rate given exogenous short term interest rates as well as all mortgage related outcomes like prepayment and rate gaps. This allows us to explore the transmission of conventional monetary policy into mortgage rates and resulting mortgage outcomes without specifying the consumption block of the model.

However, we ultimately care about the transmission of monetary policy to spending rather than mortgage market outcomes. The mortgage block of our model delivers a redistribution of disposable income in response to changes in interest rates, which depends on the distribution of outstanding mortgage rates: when interest rates decline, wealthy net-lender households face lower returns on their net wealth while poor net-borrower households free up money if they can refinance into a lower mortgage rate.[9] The consumption block of the model translates these disposable income shocks into consumption. Since the consumption block is setup up as an intentionally standard incomplete markets problem, the transmission mechanism from this redistribution of disposable income is straightforward: the joint distribution of wealth and mortgage coupons directly determines the response of aggregate spending to current rate shocks.[10] Rate histories affect current responses by altering this distribution. Our framework, which integrates the simplest model of state-dependent refinancing into an off-the shelf incomplete markets model, then delivers substantial aggregate path-dependence.[11]

What are these effects? First, our model delivers many implications for mortgage market outcomes. These results can all be explained by the fact that under our step-hazard setup, $frac > 0$ is a sufficient-statistic for the response of mortgage coupons to interest rate shocks. More specifically, using an application of results in Caballero and Engel (2007) which characterizes impulse responses in models with state-dependence, we show that the initial response of average mortgage coupons to a change in mortgage rates depends only on $frac > 0$ and not on any other features of the gap distribution. This result is driven crucially by the fact that our model features a step-hazard with a single jump at zero, and this is the formal sense in which our mortgage prepayment model is the simplest possible

[8] Enriching the model to include a fixed cost of refinancing would not change any of the basic economic forces or our conclusions for path-dependence but would substantially complicate the model solution and analysis.

[9] The model thus features a channel through which lenders are hurt by rate declines, which is an important feature emphasized by Greenwald (2017).

[10] We do not endogenize labor income or the relationship between aggregate spending and output, which will depend on the strength of nominal rigidities. Thus, our model shows that fixed-rate mortgages lead to equilibrium nominal spending responses to monetary policy which are path-dependent, but does not translate these time-varying spending effects to aggregate production. See Greenwald (2017) and Greenwald, Landvoigt and Van Nieuwerburgh (2018) for representative agent GE models with production where mortgage rate driven spending effects translate to important real GDP effects.

[11] We view this as an illustration that the prepayment channel can have important path-dependent effects in a standard model which explicitly isolates this novel mechanism from other complicating but well-understood forces. However, a more quantitatively precise policy evaluation would clearly require a refined analysis with richer consumption dynamics integrated into a full-fledged DSGE setup, which we leave for future work.

model with state-dependence: all other models with state-dependent prepayment require additional information on the distribution of rate gaps which is unnecessary in our setup.

The strong theoretical relationship between $frac > 0$ and responses to rate changes together with the dynamics of $frac > 0$ implied by our prepayment model naturally deliver all the implications discussed at the start of the paper: 1) $frac > 0$ is large when rates are trending down, which increases responsiveness to monetary policy. 2) $frac > 0$ is small if previous rates were lower than today, which decreases responsiveness to monetary policy. 3) Because households respond more rapidly to positive than negative gaps: if rates permanently increase, $frac > 0$ initially declines and then only very slowly returns to a long-run stationary value while if rates permanently decrease, $frac > 0$ initially increases but then more rapidly converges back to a long-run average. This means that monetary policy uses its ammunition up when lowering rates more rapidly than it recovers ammunition when raising rates. Finally, the presence of endogenous consumption and savings allows us to translate these redistributional effects of rate changes into aggregate spending responses. Overall, we find that the prepayment channel has substantial effects on aggregate spending and that when monetary policy is more effective at stimulating prepayment, it is then also more effective at stimulating aggregate spending.

There is by now both a large empirical and theoretical literature arguing for an important role for mortgage rates in the transmission of monetary policy.[12] Our point that time-varying refinancing incentives lead to time-varying effects of monetary policy is similar to insights in Beraja, Fuster, Hurst and Vavra (2018). They focus on variation in refinancing incentives arising from variation in house prices and resulting home equity while we focus on incentives arising from time-variation in interest rates. This distinction matters in a substantive, policy-relevant way because interest rates and resulting rate incentives respond very directly and almost immediately to changes in monetary policy while house prices are only indirectly and more slowly affected by monetary policy. For example, Wong (2018) documents strong effects of Fed Funds shocks on mortgage rates using high frequency analysis around FOMC announcement dates.[13]

This means that the current distribution of rate gaps and effectiveness of monetary policy is very directly influenced by the past history of monetary policy. It is this direct interaction between the past path of monetary policy and its current effectiveness that we label path-dependence and which distinguishes it from many other sources of state-dependence which have no direct dependence on past policy actions.[14] Path-dependence necessarily implies state-dependence, but the converse is not true. It is the intertemporal feedback between today's actions and tomorrow's rate gaps and policy effectiveness that distinguishes our results from prior studies of state-dependence.[15]

[12] See Di Maggio, Kermani, Keys, Piskorski, Ramcharan, Seru and Yao (2017), Agarwal, Amromin, Chomsisengphet, Landvoigt, Piskorski, Seru and Yao (2017), Greenwald (2017), Wong (2018), Beraja, Fuster, Hurst and Vavra (2018), Hedlund, Karahan, Mitman and Ozkan (2017), Guren, Krishnamurthy and McQuade (2018) and Abel and Fuster (2018).

[13] She finds that current policy shocks have a pass-through of 0.4 into mortgage rates. Gertler and Karadi (2015) explore rate surprises at a variety of short-horizons and find pass-through of 0.17-0.48 into mortgage rates. The high-frequency identification literature further explores real vs. nominal pass-through, the role of changes in expected current rates vs. risk premia and whether transmission occurs through changes in current rates or information effects (cf. Nakamura and Steinsson (2018)). These distinctions are not important for our analysis: our mechanism only requires the simpler observation that conventional Fed policy influences nominal mortgage rates.

[14] See e.g. Vavra (2014), Berger and Vavra (2015), Winberry (2016), Berger and Vavra (2018*a*), and Berger and Vavra (2018*b*).

[15] To be clear, we refer to path-dependence under the assumption that only aggregate macro variables are tracked as states. If we condition on $frac > 0$ and the cross-sectional distribution of savings, then past rates have no additional impact on current effectiveness, in which case our model exhibits state but not path-dependence. However, the link between current rate changes and future $frac > 0$ means today's policy actions will influence tomorrow's policy effectiveness.

Monetary policy transmission in our model is closely related to the interest rate exposure channel explored in Auclert (2017). The difference is that interest rate exposure in our framework is a function of prepayment, which in turn depends on the outstanding distribution of rate gaps. Since the gap distribution depends on past interest rates, this makes interest rate exposure and effects of monetary policy path-dependent.

Our paper is most similar to concurrent work in Eichenbaum, Rebelo and Wong (2018) which studies implications of trends in transaction costs for refinancing and state-dependent monetary policy.[16] While we use linked household-loan data that allows us to distinguish various types of prepayment, our primary empirical results on the relationship between total prepayment and interest rate incentives are similar. Our theoretical analysis is more distinct and we think highly complementary: we use a prepayment model which is simple but allows us to introduce endogenous mortgage pricing, aggregate consumption and which delivers transparent intuition for the precise determinants of path-dependence. We thus intentionally abstract from many features which they include in their model such as endogenous mortgage debt and housing, life-cycle amortization and rental markets.

The remainder of the paper proceeds as follows: Section 2 describes our data, Section 3 describes our empirical results, Section 4 describes our model setup and results and Section 5 concludes.

2 Data description

We briefly describe our primary mortgage-related data here. The appendix provides additional details as well as discussion of other data used in our analysis. Our primary mortgage data comes from Black Knight Financial Services McDash, and we supplement this using credit records from Equifax as well as information on the shares of mortgages by type from the CoreLogic LLMA data set.

Our main prepayment measures come from Black Knight Financial Services McDash loan origination and mortgage servicing records from approximately 180 million loans over the period 1992-2017. This data set includes detailed information on loan characteristics such as current interest rate and unpaid balances, appraisal values at origination, type of loan (rate-refi, cash-out, purchase), indicators for prepayment and borrower FICO scores. We measure prepayment shares as the fraction of all fixed rate first liens in the McDash Performance data set in a month with prepayment indicators.[17] While the data set provides information which distinguishes rate-refi, cash-out and new purchases at the time of loan origination, similar identifiers are not available at the time a loan is closed due to prepayment. This means that loan-level data can be used to measure prepayment but it cannot be used to directly distinguish between prepayment which is due to rate refinancing, cash-out and moves.

In order to distinguish between different types of prepayment as well as to measure additional individual level outcomes and covariates, we supplement the McDash data with additional information from the Equifax Credit Risk Insight Servicing McDash (CRISM) data set. This data set merges McDash

[16] They argue that trend declines in refinancing costs will reduce state-dependence and make monetary policy more effective. While this might seem at odds with our conclusion that monetary policy will face strong near-term headwinds, it is not. They focus on measuring average state-dependence in a stationary environment, while the headwinds we identify occur because the particular path of near-term interest rates is likely to lead to few households with refi incentives: even if refinancing costs decline to zero, rate gaps still matter since households with negative gaps have no incentive to refinance into higher rates.

[17] Regression results in the next section are very similar when including all mortgages instead of restricting to fixed rate mortgages. See Appendix Table A-4. In addition, our results equally weight mortgages, but redoing all results weighting by balances produces nearly identical results.

mortgage servicing records with credit bureau data (from Equifax) and is available beginning in 2005.[18] The structure of the data set makes it possible to link multiple loans by the same borrower together, which is not possible with mortgage servicing data alone. This lets us link the loan being paid off with any potential new loan so that we can precisely measure the reason for prepayment and distinguish refinancing from moves. It also allows us to measure equity extraction through cash-out refinancing.

While the exact prepayment type of each individual loan can only be done after 2005 when CRISM starts, there is some scope to infer the overall shares of prepayment due to refinancing vs. home moves using origination shares. In a stationary environment, every loan which is originated for the purpose of refinancing will be associated with one loan paying off for the purpose of refinancing. Prior to 2005, we thus infer the frequency of rate, cash-out and prepayment from moves by multiplying the origination shares of each type by the overall prepayment frequency. Appendix Figure A-1 validates this procedure in the post-2005 data. We measure these origination shares using CoreLogic LLMA data rather than McDash data because there is limited information on origination type in McDash data prior to 1998. The CoreLogic data is very similar to the McDash data set but the performance data does not include prepayment information prior to 1999 and we are not able to link CoreLogic loans to households as in the McDash data. Thus, while we use some information from each data set, our primary prepayment measures rely solely on McDash data. We then supplement this with information in CRISM and use information from CoreLogic data in a very limited way.

The CRISM data set links every loan in the McDash data set to an individual, and covers roughly 50% of outstanding US mortgage balances. Prior to 2005, the McDash data set has somewhat lower coverage, ranging from 10% market coverage in the early 90s to 20-25% in the late 90s. As a measure of representativeness and external validity, Appendix Figure A-2 shows that refinancing in our data closely tracks the refinancing applications index produced by the Mortgage Banker's Association from 1992-2017.[19] However, we also show robustness analysis restricting only to later sample years.

We supplement this mortgage related data with repeat sales house price indices from CoreLogic which we use to compute dynamic loan-to-value ratios. We do this by dividing the current unpaid balance for a loan by the property appraisal value at loan origination adjusted using location-specific CoreLogic house price indices. Finally, we use zip code level auto registration data from R.L. Polk available from 1998-2017. See Mian and Sufi (2012) for more information on this data set.

3 The Prepayment Incentive: Empirical Evidence

3.1 Overall Distribution of Loan-Level Rate Incentives and Prepayment

We begin our empirical analysis by looking at the overall distribution of "rate gaps" and their relationship to prepayment pooling all monthly observations from 1992-2017 together in the McDash data. For each outstanding loan i in month t we define the rate gap as $gap_{i,t} = m^*_{i,t} - m_t$, where $m^*_{i,t}$ is the current interest rate on the outstanding loan and m_t is the average 30-year FRM for new loans in month t. That is, we assume that a loan which refinances today will be replaced with a new loan at the current average

[18] See Beraja, Fuster, Hurst and Vavra (2018) and Di Maggio, Kermani and Palmer (2016) for other papers using this data.

[19] Note that we measure originations while this index measures applications. According to LendingTree, denials are roughly 8% after the financial crisis due to Dodd-Frank related changes in lending standards. This explains the level difference after the Financial Crisis but the series continue to highly comove.

30-year fixed rate.[20] We then sort loan-months into one-hundred 20 basis point wide gap bins and plot the fraction of loans in each bin (the gap distribution) as well as the fraction of loans in each bin which prepay (the non-parametric hazard). Figure 2 shows that there is a strong positive relationship between rate gaps and prepayment probabilities: loans with outstanding rates above the current market rate are much more likely to prepay than loans with outstanding rates below the current market rate. In addition, there is a substantial distribution of gaps and thus incentives to refinance across loan-months, ranging from loans whose annual rate would rise by 2 percentage points if they refinanced to those who whose rate would fall by 3 percentage points.

Figure 2: Prepayment Hazard and Density of Rate "Gaps"

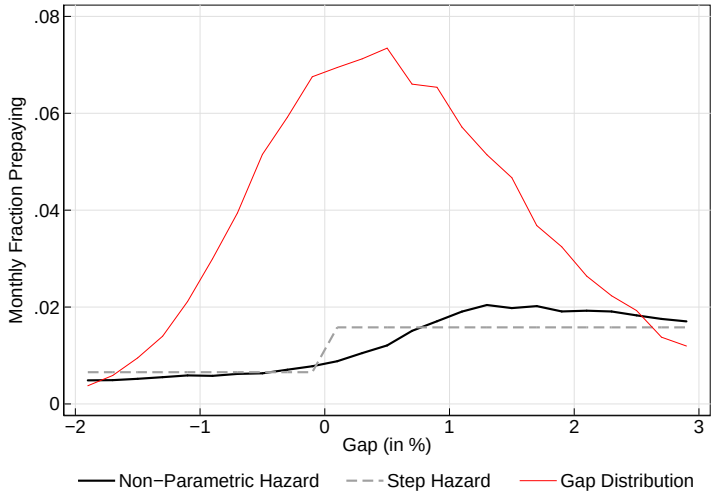


Figure shows the fraction of loans in 20-basis point gap bins ranging from -200 basis points to +300 basis points as well as the fraction of those loans prepaying. The step-hazard shows the fraction of loans with positive gaps and fraction of loans with negative gaps prepaying. The gap is the difference between the loan's current rate and the current 30 year fixed rate mortgage. We restrict to fixed-rate, first mortgages in McDash Performance data between 1992m1-2017m4.

Furthermore, the non-parametric prepayment hazard is relatively flat for gaps below zero and for gaps above 90 basis points and rises steeply between 0 and 90. To show a more explicit comparison, Figure 2 plots the average probability of prepayment for loans with positive gaps vs. loans with negative gaps (the step-hazard). While clearly not perfect, this figure shows that this step-hazard function is also not a terrible approximation of the full non-parametric hazard function, and we will use this step-hazard framework when we turn to modeling prepayments in order to derive a number of analytical simplifications. As we now show, this approximation holds even more strongly for "rate refinancing" and after introducing controls for other household and loan observables.

While Figure 2 shows a clear positive relationship between prepayment and rate gaps, it is possible that other characteristics which affect prepayment might also vary with rate gaps. For example, less attentive households might be likely to have both larger rate gaps and lower prepayment probabilities, which might confound any causal relationship between rate gaps and prepayment. In order to address

[20]For the current 30 year mortgage fixed rate, we use the current average 30 year agency conforming rate as provided by Freddie Mac. Using a common m_t abstracts from variation across borrowers with different FICO and LTV as well as variation in rates across lenders. We have redone our analysis calculating gaps which instead assume loan-specific reset targets and results are almost identical. See e.g. Appendix Table A-5.

the concern that the relationship between gaps and prepayment might be driven by some other factor, we run the following regression:

$$prepay_{i,j,t} = \mathbb{1}(gapbin)_{j,t} + X_{i,j,t} + \delta_i + \epsilon_{i,j,t} \tag{1}$$

where $\mathbb{1}(gapbin)_{j,t}$ is a dummy for the 20 basis point gap bin of household i with loan j in month t, $X_{i,j,t}$ is a vector of loan and household-level characteristics and δ_i is a household fixed effect.[21] The presence of household fixed effects removes any time invariant household characteristics which might affect both rate gaps and prepayment propensities (for example differences in financial sophistication) and time-varying controls for loan-age pick up well-known "burn-out" effects where loans which have not refinanced after a large number of years are unlikely to ever refinance. Controls for leverage and FICO pick up the fact that declines in house prices and income are likely to reduce refinancing propensities and are correlated with declines in interest rates. Dummies for previous loan-type pick up the fact that if a household has ever refinanced a loan, it is more likely to do so again in the future.

Figure 3 shows that after controlling for a variety of other observables, rate gaps have even more predictive power. It also shows that the total prepayment hazard estimated via equation 1 is even closer to a discrete step-function at zero than under our original time series regression specification. Quantitatively, our results imply that conditional on a negative rate gap, households have a 0.5% probability (per month) of prepayment, while this probability jumps to 2.5% per month when the rate gap turns positive.

Figure 3: Prepayment Hazard with Individual Controls

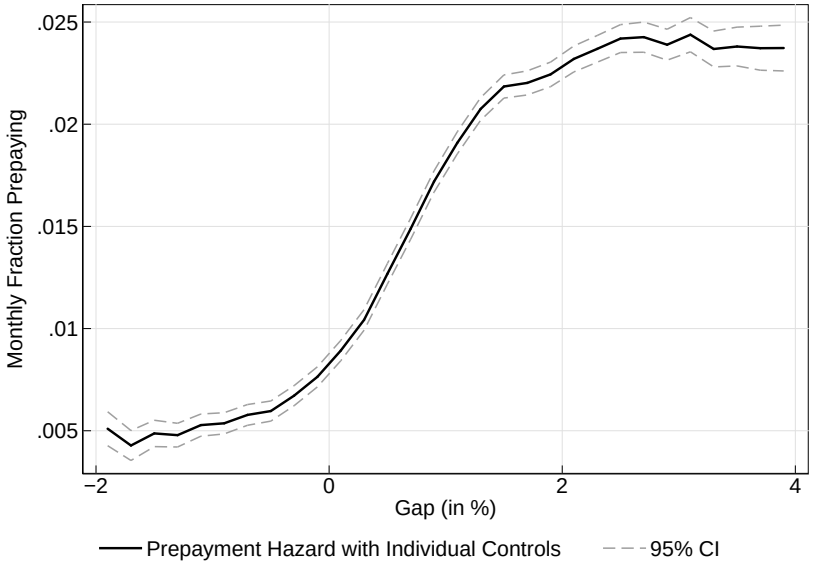


Figure shows the point estimates and 95-percent confidence intervals for the coefficients on the 20 basis point gap bin dummies in regression 1. Standard errors are clustered by household. In order to include household fixed effects and time-varying characteristics, figure uses CRISM data linked to credit records from 2005m6-2017m4.

[21] $X_{i,j,t}$ controls are: a quadratic in FICO, a quadratic in leverage (CLTV), a quadratic in loan age and dummies for whether the current loan was itself a new purchase loan, a cash-out refi or a rate refi, dummies for investor type (GSE, RFC, GNMA, on-balance sheet, private MBS), loan type (FHA, VA, conventional w/ and w/out PMI and HUD).

Most of our empirical results focus on total prepayment rather than on the decomposition of prepayment into rate-refinancing, cash-out refinancing and home-moves for several reasons: 1) The evolution of the rate gap distribution is driven by *all* prepayment including rate, cash-out and purchase prepayment. This means that concentrating only on rate refinancing would not provide a full view of the evolution of future rate incentives.[22] 2) The incentives to prepay are all linked together and cannot be decoupled from each other. For example, a rise in house prices increases the incentive for cashout refinancing. Since households cannot simultaneously do a cashout and non-cashout refi, this will lead to a decline in rate-refinancing. 3) Prepayment hazards as a function of individual loan gaps can be constructed for our entire 1992-2017 sample while we can only measure prepayment type for individual loans using CRISM data starting in 2005. This is because loan-level data alone can only identify which loans prepay but cannot link the prepaying loan to the new loan being originated and its purpose.

With these caveats in mind, beginning in 2005, the ability to link loans to households means that we can then link the loan which is prepaid to the new loan which is originated. This allows us to measure the type of prepayment based on the new loan which is originated at the same time and to verify that jumps in prepayment activity around gaps of zero are driven by refinancing. Figure 4 re-runs the regression specification in 1 by prepayment type to show that the probability of prepayment as a function of gaps varies substantially by prepayment type:

Figure 4: Prepayment Hazard and Density of Rate "Gaps" by Prepayment Type

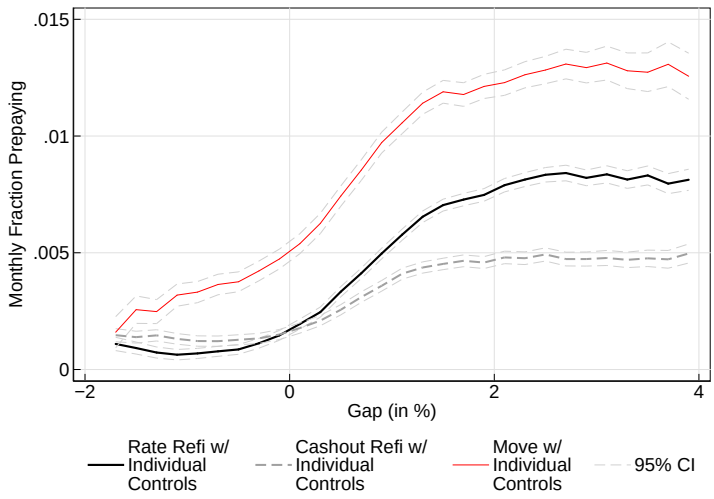


Figure shows the fraction of loans in 20 basis point bins which are paid off via rate-refinancing, cash-out refinancing or as a result of purchasing a different house using CRISM data from 2005m6-2017m4. See Appendix for descriptions of the identification of prepayment type. Standard errors are clustered by household.

Figure 4 shows that while the probability of all types of prepayment increases with the size of the gap, the jump at zero is strongest for rate-refinancing.[23] This is not particularly surprising since there is no incentive to engage in rate-refinancing into a higher rate while households may move or extract equity even if it requires taking out a mortgage at a higher rate than before.[24] However, this

[22]To take an extreme example, if all households moved houses today and took out new purchase mortgages, the gap distribution would be compressed at zero tomorrow, so there would be little incentive to do rate refinancing.

[23]The increasing hazard for moves demonstrates the "mortgage-lock" discussed in Quigley (1987).

[24]See the Appendix for discussion of measuring rate vs. cash-out refinancing, which requires an assumption on what

serves to validate the importance of the refinancing channel for prepayment which will be central to our theoretical analysis.[25]

3.2 Time-Series Variation in Rate Incentives and Aggregate Prepayment

Together Figures 2-3 show that there is a strong overall relationship between rate gaps and prepayment propensities when pooling the data across all months. We next move from these pooled relationships to time-series analysis in order to show that: 1) The distribution of rate gaps varies substantially across time. 2) This time-series variation strongly predicts time-series variation in prepayment. Given the stark difference in prepayment between loans with positive and negative gaps, our preferred summary statistic for the distribution of gaps at a point in time is the fraction of loans with positive rate gaps, which we label as $frac > 0$. Figure 5 shows that $frac > 0$ moves substantially across time, ranging from less than 20% in early 2000 to nearly 100% in 2003 and 2010. While we focus on $frac > 0$ as a summary statistic for the gap distribution, Appendix Figure A-4 shows that these movements in $frac > 0$ are closely associated with broad movements in the overall distribution of rate gaps. Figure 5 also shows that not only does $frac > 0$ move substantially across time, it is highly correlated with the fraction of loans prepaying in a given month, with a correlation of 0.53.[26]

Table 1 explores the time-series relationships between rate incentives and prepayment using more formal regression analysis. Column 1 shows that there is a very significant positive relationship between $frac > 0$ and prepayment. The R^2 of 0.282 means that this single variable explains just under thirty percent of the time-series variation in prepayment. Column 2 and 3 explore alternative thresholds as potential summary statistics for the distribution of rate gaps. While the fraction of loans with rate gaps of at least 50 or at least 100 basis points also have strong predictive power, they are mildly less successful (as measured by R^2) at predicting prepayment than the fraction of loans with rate gaps of at least 0. It is also straightforward to regress prepayment in each month on the fraction of loans in the full set of one-hundred bin dummies used in Figures 2-3. Importantly, this regression only increases the R^2 from 0.282 to 0.305, which means that $frac > 0$ summarizes nearly all the predictive content for prepayment which is delivered by knowing the entire non-parametric distribution of gaps. This also implies that using the constant step-hazard in Figure 2 interacted with the full time-varying gap distribution will produce 92.5% (.282/.305) of the variation in prepayment delivered by performing the same exercise with the non-parametric hazard. Thus, in a formal sense, the step-hazard produces predictions for prepayment which are nearly identical to the more complicated non-parametric hazard.

What other factors determine prepayment besides interest rate incentives?[27] Motivated by the analysis in Beraja, Fuster, Hurst and Vavra (2018), Column (4) shows that $frac > 0$ has less effect on

amount of closing costs may be rolled into the new loan. Furthermore, as discussed in footnote 20, there is some measurement error in rate gaps. These two issues likely account for the small probability of rate-refinancing even for negative gaps.

[25] It is also useful to decompose prepayment into moves and refinancing since the former may have effects on housing demand while the latter will primarily affect spending.

[26] In all specifications in this section we measure $frac > 0$ in month t and prepayment in month $t+1$ since McDash data measures originations rather than applications, and there is a lag of 1-2 months from application to origination.

[27] All of our results so far focus on the relationship between rate gaps and prepayment. However, it is also interesting to look at the relationship between the dollars saved when refinancing and prepayment rather than just the simple interest rate gap. This is because for a given interest rate gap, loans with larger balances obtain greater payment reductions by refinancing. Appendix Table A-3 recomputes the relationship between gaps and prepayment but measuring gaps using annual $ saved ($gap_{i,t} \times balance_{i,t}$) rather than gaps and finds similar conclusions.

Figure 5: Prepayment vs. Fraction with Positive Rate Gap Time-Series

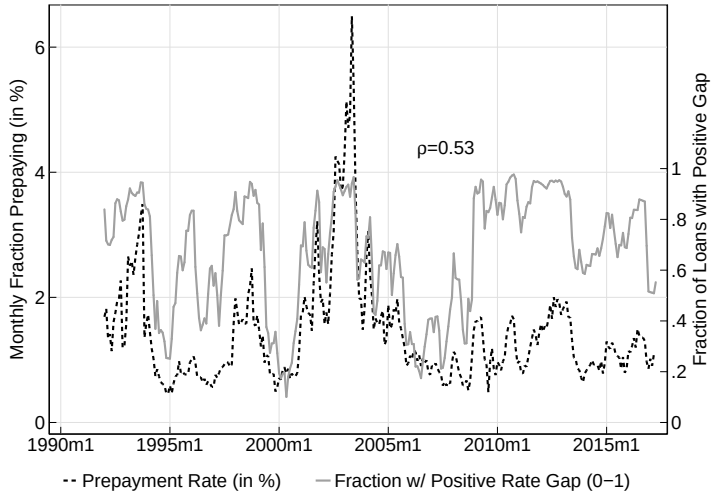


Figure shows the fraction of loans in McDash Performance data with positive rate gaps in each month as well as the fraction of loans prepaying in each month. The time sample is 1992m1-2017m4.

refinancing when leverage is high. Together, average leverage plus $frac > 0$ explain roughly half of the time-series variation in prepayment. Figure 5 shows that there is a large spike in prepayment in 2003. This unusual increase in refinancing above and beyond what can be explained by the behavior of mortgage rates coincides exactly with the "Mortgage-Rate Conundrum" documented by Justiniano, Primiceri and Tambalotti (2017) during which they argue mortgage credit was unusually lose. While we thus cannot explain this outlier in 2003, Column (5) shows that if we introduce a 2003 dummy, leverage and $frac > 0$ explain almost two-thirds of the variation in prepayment. Finally, Columns (6) and (8) show that $frac > 0$ alone has stronger explanatory power both in the period before the housing boom-bust and in the period after the housing boom-bust.[28] This suggests that rate-gaps are the most important determinant of refinancing behavior during "normal" times in the housing market. Column (7) shows that while $frac > 0$ continues to have important explanatory power during the housing boom-bust, its predictive power is significantly lower. Again, this is not surprising since a substantial fraction of prepayment activity during this particular period was related to house price movements and cash-out refinancing (Beraja, Fuster, Hurst and Vavra (2018)).

The results so far show a strong correlation between prepayment and $frac > 0$. However, this may not reflect a causal relationship since $frac > 0$ depends on past endogenous interest rates. It is thus possible that some confounding factor affects both $frac > 0$ and the propensity to prepay. As one way to address this concern, Table 3 re-estimate our baseline regressions using the cumulative value of the Gertler and Karadi (2015) high-frequency monetary policy shock series over the past six months as an instrument for $frac > 0$. Unsurprisingly, this reduces power and increases standard errors, but point estimates remain similar and $frac > 0$ remains a significant predictor of prepayment activity.[29] As an additional robustness check, in Appendix Table A-2, we show that there are similar relationships between $frac > 0$ and prepayment in specifications which also include calendar quarter fixed effects.

[28] The level of prepayment is lower after the housing bust but $frac > 0$ exhibits similar predictive power as measured by R^2.
[29] First-stage F-stats are fairly strong at 12 to 13 and exceed the 15% Stock-Yogo critical values for weak instrument bias.

In these specifications, identification only comes from monthly relationships between prepayment and rate incentives within quarters.[30] These specifications rule out most confounding factors, which would be unlikely to show up at these high frequencies.[31]

For reasons described above, we concentrate on total prepayment rather than just refinancing as our primary outcome of interest. However, our model will focus on effects of rate refinancing on spending. Since Figure 4 shows that there are different average relationships between the components of prepayment and rate incentives, it is important to also decompose total prepayment time-series relationships. In Table 2 we decompose the positive time-series relationship between total prepayment and $frac > 0$ into its constituent types (rate-refi, cashout-refi and purchase). As suggested by the overall loan-level relationship in Figure 3, $frac > 0$ is most important for explaining rate-refinancing.[32] $frac > 0$ alone explains roughly 40% of the time-series variance in rate-refinancing.

3.3 Effects of Rate Incentives on Outstanding Coupons

All of our empirical results thus far focus on prepayment (and its constituent components) as the outcome of interest. However, changes in the average outstanding mortgage rate m^* are arguably more important than prepayment rates since mortgage payments are what enter the household budget constraint and prepayment matters more for households who secure large payment reductions. Prepayment rates and changes in m^* are of course related: in each month $\Delta m^* = \int gap \times f(gap) \times h(gap) dgap$, where f is the density of gaps and h is the prepayment hazard in that month. The larger the prepayment hazard is for households with large gaps, the more m^* will decline. However, it is also clear that they need not move perfectly together since average rates will decline by more if the households prepaying have larger gaps. In Figure 6, we plot the actual time series for Δm^* and compare it to $\Delta m^*_{step} = \int gap \times f(gap)\ timesh_{step}(gap) dgap$, where h_{step} is the step hazard which arises from using only information on $frac > 0$. This figure shows that in addition to predicting prepayment, $frac > 0$ is a good predictor of changes in contractual mortgage coupon rates. Clearly the fit between the actual series and predicted series is not perfect, but the correlation is 0.76 and similar to the prepayment series, the only large miss is in 2003 where the step-hazard substantially understates actual declines in mortgage rates. Again this particular deviation of mortgage market activity in 2003 from historical relationships has received attention in Justiniano, Primiceri and Tambalotti (2017).

Table 4 Column (1) confirms this visual result with a regression specification to show that there is an extremely strong negative time-series relationship between $frac > 0$ and Δm^*.[33] Unsurprisingly, Column (2) shows that when the current market interest rate rises, so does the resulting average outstanding rate. More interestingly, column (3) shows that there is a strong interaction effect between $frac > 0$ and Δ FRM: interest rate pass-through into average coupons is much stronger when $frac > 0$ is large.[34] As we discuss below, this increase in rate pass-through with $frac > 0$ is a central impli-

[30] Allowing for a quadratic monthly trend instead of non-parametric quarterly trends also does not change results.

[31] For example, trends in lender concentration might affect both the level of rate gaps and passthrough of mortgage rates into prepayment and coupons, but these concentration trends primarily operate at lower frequencies.

[32] After 2005 we decompose prepayment using CRISM data, prior to 2005 we assume stationarity and decomposes using origination shares. See Section 2. This decomposition requires origination shares data from CoreLogic LLMA data, which has poor coverage prior to 1993. For this reason, regressions start from 1993 rather than 1992 as in Table 1.

[33] To account for a lag between applications and originations, in all specifications, $frac > 0$ is measured as of month t, Δm^* is measured between month t and month $t+1$ and Δ FRM is measured between month $t-1$ and month t.

[34] Note that the negative coefficient on Δ FRM does not imply that increases in interest rates lower coupons since the median

Figure 6: Prepayment vs. Fraction with Positive Rate Gap Time-Series

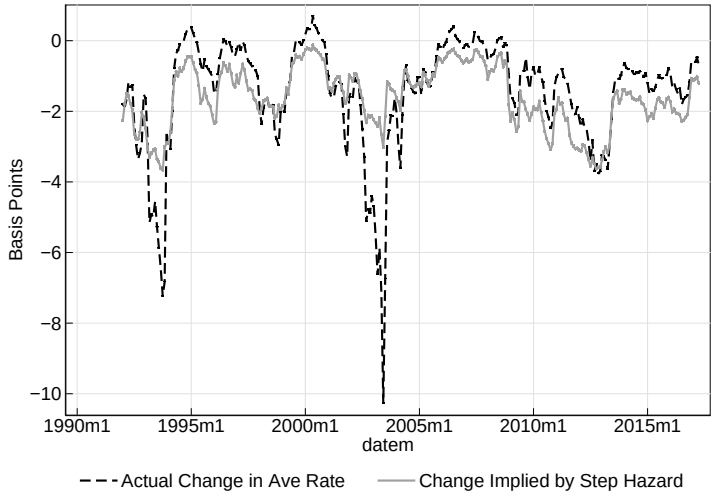


Figure shows the actual change in the average outstanding mortgage rate m^* as well as that predicted by interacting the step-hazard with changes in the gap distribution.

cation of our theoretical model and is a key indicator of path-dependence. While movements in rate gaps in our model will induce strong effects on interest rate pass-through, other forces can also have such effects: Beraja, Fuster, Hurst and Vavra (2018) emphasize that increases in leverage also reduce interest rate pass-through and the effectiveness of monetary policy. Motivated by this result, column (5) includes also interactions of interest rate changes in month t with average leverage in this same month. While we indeed find a negative interaction effect between leverage and pass-through, the interaction between $frac > 0$ and Δ FRM is if anything mildly amplified.

3.4 Cross-Region Evidence

Our empirical evidence focuses primarily on aggregate time-series relationships between $frac > 0$, prepayment and outstanding coupons. This is because mortgage rate movements which drive changes in $frac > 0$ are aggregate rather than regional. As first shown by Hurst, Keys, Seru and Vavra (2016), there is very little variation in market mortgage rates, m, across space. However, despite the lack of regional variation in m, there is nevertheless some regional variation in the cross-sectional distribution of $m^* - m$ and thus $frac > 0$ arising from differences in the timing of mortgage origination across regions.[35]

In Table 5 we exploit this variation across MSAs to show that there is a strong positive relationship between prepayment rates and $frac > 0$ at the MSA level, even after including both month and MSA $\times$ quarter fixed effects. Importantly, the presence of time fixed effects means that identification occurs en-

value of $frac > 0$ is approximately 0.7.

[35]These differences in timing could owe to many factors such as the differences in house price growth emphasized by Beraja, Fuster, Hurst and Vavra (2018) or the differences in lifecycle effects emphasized by Wong (2018). We have no instrument for cross-region rate gaps, so endogeneity is a concern. However, specifications below with MSA-quarter fixed effects absorb differences in gaps and prepayment which are constant within quarters, which eliminates many obvious concerns. When we also control for month fixed effects, any confounding channel must arise from a region-specific factor which varies at high frequencies independently of aggregate conditions and affects both current relative prepayment and gaps.

tirely from cross-region variation rather than aggregate time-series variation. This eliminates concerns that results might be driven by endogenous monetary policy since monetary policy does not vary across regions. This thus complements our previous identification strategy which dealt with endogeneity using high-frequency monetary policy shocks in Table 3. In addition, the presence of MSA $\times$ quarter fixed effects means that identification only comes from region-specific monthly variation within quarters and not from more permanent spatial differences. This eliminates concerns that results might be driven by differences in demographics, lender concentration or any other slower moving local characteristics.[36]

While moving to cross-region specifications can help to further alleviate identification concerns from aggregate time-series specifications, the primary advantage of these specifications is that they can be used to measure relationships between prepayment and some local spending responses. In particular, R.L. Polk collects high quality zip-code level data on local auto registrations. In Table 6, we look at the relationship between local prepayment, auto sales growth and changes in interest rates. In column (1), we regress auto sales growth from month t to $t+1$ on the change in prepayment between $t-1$ and t plus a month fixed effect so that we again only identify off of cross-region differences and not aggregate time-variation. This lag structure allows for a 1-month lag between mortgage origination and new auto purchases in order to allow for shopping effects. The coefficient shows that regions with larger increases in prepayment see larger increases in auto sales.[37] Results are, if anything, slightly stronger in Column (2), which includes additional MSA $\times$ quarter fixed effects so that identification does not come off of permanent or slower moving differences in spending growth and prepayment across space.

While this shows a strong relationship between changes in prepayment and changes in auto spending, there are many obvious threats to a causal interpretation. For example, an increase in expected future income with no change in income today might lead to an increase in auto spending and an increase in refinancing to finance that spending, in which case the coefficient on Δ freq would have an upward bias. Conversely, current income shocks might lead the coefficient to be biased down since greater income today might lead to greater desired spending and to a decrease in the need to refinance to fund spending. Furthermore, the McDash data does not contain a full census of local mortgage activity, so there is also likely measurement error in local prepayment rates, which will lead to attenuation bias. In Columns (3) and (4), motivated by our earlier empirical evidence that $frac > 0$ has strong effects on prepayment, we instrument for Δ freq using Δ frac > 0. The identifying assumption is that changes in the fraction of households with positive gaps (after controlling for aggregate month effects and region by quarter fixed effects) only affects auto sales growth through refinancing.[38] When instrumenting for changes in prepayment, point estimates are increased substantially, suggesting that the second type of OLS bias discussed above is more important.

In Columns (4)-(8), we show that changes in interest rates have greater effects on spending in locations where more households are prepaying their mortgages. In particular, we regress auto spending growth from month t to $t+1$ on the frequency of prepayment in month t (rather than the change in prepayment) but now interacted with the change in the 30 year fixed rate mortgage between month

[36] See e.g. Agarwal, Amromin, Chomsisengphet, Piskorski, Seru and Yao (2017) and Scharfstein and Sunderam (2016).

[37] The standard deviation of Δ freq is 0.388, so a 2 SD increase in frequency is associated with a 1.5 percentage point increase in the growth rate of auto sales (.388*2*.019=.0147). This is relative to an average growth rate of 2.2%.

[38] Unfortunately, there is no obvious instrument for regional $frac > 0$. For example, the monetary policy shock instrument we used in aggregate regressions will be absorbed by the aggregate month fixed effects. However, quarter $\times$ MSA controls address many obvious confounding stories, and controls for other observables such as high frequency changes in leverage have little effect on the conclusions.

$t-1$ and t. This shows that decreases in mortgage rates are correlated with larger increases in spending growth when more households are refinancing. This interaction between rate passthrough into spending and the level of prepayment will be a central prediction of the model we build in the following section. As discussed in our results looking at regional prepayment, endogeneity of the mortgage rate change is not likely to be a problem even in the presence of endogenous monetary policy, since all specifications have time fixed effects and are only identified off of cross-region variation. This means any endogenous relationship between FRM changes and aggregate conditions will be absorbed in the time fixed effects. However, there remains the previous concern that the frequency of prepayment may be related to other transitory local conditions which also affect auto spending growth. Here we are primarily interested in the interaction effect, and any effects which just move prepayment and spending together will not affect the interaction term. Nevertheless, we again instrument for the current prepayment frequency using the level (rather than the change) in the fraction of households with positive gaps. The identifying assumptions are similar to before. Again relationships become much stronger, presumably because we are reducing attenuation bias and eliminating the effects of current income and wealth shocks which one would expect to increase spending and reduce refinancing.

4 A Model of Mortgage Prepayment

We now turn to a theoretical model which we use to interpret our empirical results. This model has several goals: Most importantly, we use the model to explore counterfactuals which cannot be explored in the data. The data is directly informative about the current state of the economy but not its future evolution. In particular, our regressions show that $frac > 0$ has important predictive power for current prepayment, but not how future $frac > 0$ will respond to policy actions today. We use our model to argue that current policy actions have important implications for the evolution of future rate gaps and in turn the impact of future policy actions. In addition, we use our model to flesh out the consumption implications of the prepayment channel. Our cross-region results provide evidence of auto spending responses through the prepayment channel, but autos are only one piece of total consumption. Furthermore, since lending markets are national these regional regressions do not measure spending by lenders, who are negatively affected when borrowers refinance into lower rates.

In order to address these issues, we build a continuous-time small open economy framework which embeds a simple model of household mortgage refinancing into an incomplete markets environment with endogenous mortgage pricing. Our model includes (a) a continuum of home-owner households making consumption, savings and mortgage refinancing decisions, and (b) a competitive and deep-pocketed financial intermediary sector that extends fixed rate prepayable mortgage contracts and finances itself via deposits and short term debt. Households' labor income is risky, but households only have access to short term risk-free savings as a way to insure against this risk, as in Aiyagari (1994). The financial intermediary's net funding needs are filled through international funding at the short term interest rate, while the fixed rate mortgage interest rate is pinned down by the financial intermediaries' zero profit condition. This endogenous relationship between short rates and mortage rates leads to endogenous redistributional effects of rate changes: when mortgage rates decline, this lowers debt payments for borrowers who refinance. When short rates decline, this lowers returns for lenders.

As we explain in more detail below, we assume that households follow a "Calvo-style" refinancing

process: they can only refinance at Poisson arrival times and will do so if the rate on their mortgage is above the current market rate. In a sense which we formalize below, this is the simplest possible model of infrequent refinancing which still allows for state-dependent refinancing decisions that depend on rate gaps. Nevertheless, our empirical work shows it has good explanatory power for actual prepayment patterns and importantly, it allows us to break our analysis into two blocks: a mortgage refinancing component, and a consumption-savings component. In our model, refinancing decisions are orthogonal to consumption-savings decisions, which means that the mortgage block can be analyzed without specifying households' preferences, market access and labor income characteristics which are crucial for the consumption block. The consumption-savings component of our model is however entirely dependent on the refinancing behavior of households since consumption will depend on both current and expected future rates.

4.1 Uncertainty

There are two levels of uncertainty that households take into account when forming expectations and making consumption, savings and prepayment decisions. Households face aggregate uncertainty, in the sense that short term rates and mortgage rates follow specific stochastic processes. Households face idiosyncratic uncertainty, since their labor income fluctuates over time according to transition probabilities that may depend on the aggregate state of the economy.

4.1.1 Aggregate Uncertainty

Let $\{a_t\}_{t\geq 0}$ be a discrete state continuous time Markov process taking values in $\{1, ..., n_a\}$ and encoding the aggregate state of the economy. This stochastic process will drive the short term interest rate $r_t = r(a_t)$ as well as the mortgage rate $m_t = m(a_t)$. We take the mapping between short term interest rates and mortgage rates as given for now, and discuss in Section 4.4 how this mapping is pinned down via the zero profit condition of financial intermediaries. It is worth noting that our model of the transmission between short term rates and mortgage rates is a particular case of the more general framework studied by Berger, Milbradt and Tourre (2018). We assume that the discrete states are ordered by increasing short term interest rates: $r(1) < ... < r(n_a)$. Let $\Lambda = (\lambda_{ij})_{1\leq i,j\leq n_a}$ be the generator matrix for $\{a_t\}_{t\geq 0}$, which encodes transition probabilities from one state to another. Loosely speaking, on the time interval $[t, t+dt)$, the aggregate state of the economy transitions from state a_t to state j with probability $\lambda_{a_t,j}dt$.

4.1.2 Idiosyncratic Uncertainty

Let $\{s_t\}_{t\geq 0}$ be a second discrete state continuous time Markov process taking values in $\{1, ..., n_s\}$ and encoding the idiosyncratic state of a household of focus. The idiosyncratic state of such household is its annual income $Y_t = Y(s_t)$. This idiosyncratic process might be correlated with the aggregate process $\{a_t\}_{t\geq 0}$. One could imagine, for example that a period of low interest rates corresponds to a downturn for the aggregate economy, and that transitions from high income states to low income states are more frequent than in good economic times. To account for this correlation, let $\Theta^a = \left(\theta^a_{ij}\right)_{1\leq i,j\leq n_s}$

be the generator matrix for s, conditional on being in aggregate state a. For each aggregate state a, Θ^a is assumed to be conservative.

4.2 Endowment and Household Balance-Sheet

Our household of focus is born at $t = 0$ with a house, and liquid savings worth W_0.[39] The house is financed at birth with a fixed-rate prepayable mortgage with principal balance F and coupon rate $m_0 = m(a_0)$ at time $t = 0$. The mortgage has a constant face value and can be refinanced at the sole option of the household but only at random, exponentially distributed times (arrival intensity χ). When these opportunities arise, the household can choose to keep its existing mortgage or to refinance to the current mortgage market rate at no cost. We refer to these random "Calvo-style" opportunities to refinance as attention times: since there is no cost to refinance during a random attention time, a household will choose to refinance if and only if the current market rate is below his contractual rate. This modeling choice is a reduced form for a variety of costs (time and pecuniary) that a household faces and that help explain the failure of many households in the data to refinance even when it is optimal to do so (cf. Keys, Pope and Pope (2016)). This random arrival process of refinancing opportunities simplifies computations substantially by allowing us to abstract from a complex fixed-cost decision problem.[40] To focus on the simplest possible environment with refinancing, we assume that the mortgage balance remains fixed at F, so our model features rate but not cash-out refinancing.

The household also moves for exogenous reasons from its current house to a new house (with the same market value) at Poisson arrival rate ν. This exogeneous move is a device that forces the household to reset its mortgage rate to the current mortgage rate. While we model these moving shocks, we abstract from endogenous home purchase decisions and assume that mortgage balances remain constant when moving. Thus, the only effect of prepayment in our model is resetting the mortgage rate: the mortgage rate may reset to either a lower or a higher rate when the household moves from one house to another but will only reset to a lower rate when the household chooses to refinance. The combination of exogenous moving shocks together with random opportunities to refinance leads to a step-hazard for prepayment exactly of the type drawn in Figure 2.

The household receives labor income Y_t that is subject to uninsurable shocks. The household has access to a savings technology (at the short rate r_t) and solves a standard consumption-savings problem given liquid savings W_t, and we restrict $W_t \geq 0$: our household cannot take on unsecured debt, its only liability is its outstanding mortgage, and the net financial assets of our household is thus $W_t - F$.

Households in our model have no default option; we thus select $\{Y(s)\}_{s \leq n_s}$ and $\{m(a)\}_{a \leq n_a}$ so that labor income is always weakly greater than the required debt service via the parameter restrictions:[41]

[39] Our analysis abstracts from any effects of r on consumption which work through house price effects by assuming a constant house price. A number of papers have argued for important house price effects on spending (cf. Mian, Rao and Sufi (2013), Berger, Guerrieri, Lorenzoni and Vavra (2018), Stroebel and Vavra (2018), Kaplan, Mitman and Violante (2016) and Guren, McKay, Nakamura and Steinsson (2018)), and we believe modeling these effects would amplify our conclusions: interest rate histories which lead to greater refinancing behavior would also lead to greater housing demand, price increases and resulting consumption through this additional channel.

[40] Fixed costs lead to a more complicated real option decision. Closed form solutions such as those in Agarwal, Driscoll and Laibson (2015) require strong assumptions which do not hold in our context. Fixed costs break the separability of the mortgage block in our model and would substantially complicate the model solution. While we choose to focus on the simplest possible model with realistic state-dependent prepayment, the same basic economic forces would still apply with fixed refinancing costs. See e.g. Eichenbaum, Rebelo and Wong (2018) for related results in a model with fixed costs.

[41] We also implicitly assume that there is no DTI restriction on refinancing since there is no default. Default could be

$$\min_{s,a} \left[Y(s) - m(a)F \right] \geq 0$$

4.3 Household Problem

We assume that the life-time utility V_0 for a household satisfies:

$$V_t = \mathbb{E}_t \left[\int_t^{+\infty} \delta e^{-\delta(s-t)} \frac{C_s^{1-\gamma}}{1-\gamma} ds \right]. \tag{2}$$

Let $\boldsymbol{\tau_a} := \{\tau_{a,i}\}_{i\geq 0}$ be a sequence of attention times, and $N_t^{(\boldsymbol{\tau_a})}$ the related counting process. Let $\boldsymbol{\tau_m} := \{\tau_{m,i}\}_{i\geq 0}$ be a sequence of moving times, and $N_t^{(\boldsymbol{\tau_m})}$ the related counting process. The state vector for a given household is $S := (W, a, a^*, s)$. The index a indicates the current aggregate state, while the index a^* indicates the aggregate state at which the household refinanced last (i.e. the household's mortgage coupon is $m(a^*)$). The household solves the following problem:

$$\begin{aligned} V(S) :&= \sup_C \mathbb{E}^S \left[\int_0^{+\infty} \delta e^{-\delta t} \frac{C_t^{1-\gamma}}{1-\gamma} dt \right] \\ \text{s.t. } dW_t &= \left(Y(s_t) - C_t + r(a_t) W_t - m(a_t^*) F \right) dt, \qquad W_t \geq 0 \\ da_t^* &= (a_t - a_{t-}^*) \left[1_{\{m(a_t) < m(a_{t-}^*)\}} dN_t^{(\boldsymbol{\tau_a})} + dN_t^{(\boldsymbol{\tau_m})} \right] \end{aligned}$$

The superscript on the expectation operator above is used for conditioning on information at time 0. Changes in a_t^* occur for two reasons: either (i) at attention times, in which case a_t^* changes only if the current mortgage rate is lower than the rate that the household's old rate, or (ii) at moving times, in which case the household is "forced" to refinance its house at the current mortgage rate.

In Section A.1.2, we write down the Hamilton-Jacobi-Bellman (HJB) equation corresponding to the household value function V. Optimal consumption satisfies a standard first order condition: the marginal utility of consumption must equate the marginal value of liquid wealth. In order to calculate the numerical solution to the HJB equation and optimal consumption policies, we need to solve a system of $n_s n_a^2$ nested first order non-linear ODEs. In Section A.1.4, we describe our algorithm, which relies on a standard finite different scheme.

4.4 Link between Short Rates and Mortgage Rates

Households' refinancing decision is trivial: whenever they have to move (with hazard rate ν), they have to reset their mortgage; whenever they are attentive (with hazard rate χ), they refinance if and only if the current market mortgage rate is below the rate they are currently paying. This gives rise to a step-hazard for the probability of prepayment as a function of a households mortgage gap, with hazard

introduced as in Campbell and Cocco (2013) and interactions with DTI as in Greenwald (2017). This would complicate the model my making the distribution of mortgage debt a state-variable and would introduce another decision. However, default would likely amplify the path-dependence we identify: cuts in the current interest rates which lead to greater payment reductions will lead to greater default reductions (cf. Ganong and Noel (2017)). The path-dependent effect of rate cuts on payments is exactly our object of study.

ν for negative gaps and hazard $\nu + \chi$ for positive gaps.

The risk-neutral valuation of the mortgage cash-flows is studied in Berger, Milbradt and Tourre (2018), and the financial intermediaries' break-even condition leads to a mapping between the short term interest rate process $r(a_t)$ and mortgage rates $m(a_t)$. This mapping insures that the intermediary's expected profits are zero when extending a fixed rate prepayable mortgage financed with short term debt. Section A.1.1 provides details.

As discussed above, this model structure implies that refinancing decisions do not depend on the consumption and saving's behavior of a household. This allows us to perform experiments and compute impulse responses to shocks that are purely focused on mortgage outcomes without having to compute other model equilibrium outcomes. However, the entire model equilibrium must be solved when computing implications for consumption and savings.

4.4.1 Impulse Response Functions

Our impulse response function ("IRF") calculations are focused on the following outcome variables: aggregate (per annum) consumption, average mortgage coupons, and average prepayment rates. To compute the consumption IRF (for example), we define $\bar{C}(t; g_0, r_0)$, the aggregate consumption at time t as a function of the initial state of the economy (defined by an initial distribution $g_0(W, a^*, s)$ over liquid savings, coupons and income levels) and short term interest rate at time 0, as follows:

$$\bar{C}(t; g_0, r_0) := \sum_{s=1}^{n_s} \sum_{a^*=1}^{n_a} \int \mathbb{E}\left[C_{a_t, a_t^*, s_t}(W_t) \,|W_0 = W, s_0 = s, a_0^* = a^*, r(a_0) = r_0\right] g_0(W, a^*, s)\, dW$$

In the above, $C_{a_t, a_t^*, s_t}(W_t)$ is the consumption function for an optimizing household with liquid savings W_t, mortgage coupon $m(a_t^*)$, income level $Y(s_t)$, when the current short term rate is $r(a_t)$ (and the corresponding market mortgage rate is $m(a_t)$). The consumption IRF is then simply defined as:

$$IRF_C(t; g_0, r_0) := \bar{C}(t; g_0, r_0) - \bar{C}(t; g_0, r_0 - 1\%)$$

In other words, the IRF is defined as the change in aggregate consumption that would occur upon an unexpected 1% drop to short term interest rates. We can similarly define the IRF of the average mortgage coupon rate $\bar{m}^*(t; g_0, r_0)$ and of the average prepayment rate $\bar{P}(t; g_0, r_0)$.

4.4.2 Calibration

The model calibration leverages our empirical analysis. In particular, since ν represents the prepayment intensity conditional on a negative rate gap, and since $\nu + \chi$ represents the prepayment intensity conditional on a positive rate gap, we calibrate those parameters using our point estimate of the regression of prepayment rate onto (a) a constant term and (b) the fraction of households with positive gaps. We set ν to be equal to the intercept of this regression (equal to 2% per annum), and we set χ to be equal to the slope of this regression (equal to 20% per annum).[42]

[42]Note that to simplify exposition, we describe the probability of moving as a constant independent of rates. This rate ν is lower than the frequency of moving in the data, since the actual hazard of moving is increasing in the rate gap. However, since we calibrate χ to match total prepayment and because moving and refinancing in our model affect households identically, our setup is isomorphic to a model in which moving and refinancing hazards both have a positive step at zero. If we then split

We set $\gamma = 2$ following standard values in the macro literature. We then calibrate the rate of time preference δ so that the ergodic aggregate liquid savings in the economy is equal to the aggregate household mortgage debt in the economy. This yields a discount rate of $\delta = 10\%$ per annum. This calibration strategy implies that the current account balance is zero on average despite assuming a small open economy in which the financial intermediary is owned by deep-pocketed foreign investors. In this sense, the combination of interest rates and preferences we choose is consistent with a closed-economy equilibrium "on-average" but with stochastic foreign capital flows.

We intentionally keep our income process simple, with only 2 possible states: households are either employed, with annual income of \$60k, or unemployed, with annual income of \$10k. Transition intensities from one state to the other are calibrated to match historical U-E and E-U transitions hazard rates. We fix the mortgage face value F at the average value in our data of \$130k.

Finally, our discrete-state short term interest rate process is calibrated to approximate the model of Cox, Ingersoll Jr and Ross (2005). As discussed in section A.1.5, the approximation matches a set of conditional and conditional moments of the original Cox, Ingersoll Jr and Ross (2005) model to the corresponding moments of the discrete-state continuous time process. The interest rate process requires three parameters: a long-run average short rate $\bar{r}$, which we set to 4%; a speed of mean reversion κ, which we set to 13%; and a volatility parameter σ, which we set to 6%.

5 Mortgage Market Results

We begin by discussing implications of our model for mortgage market outcomes, and then turn to the implications of these mortgage responses for consumption. As explained above, mortgage market outcomes can be characterized without specifying anything about the income process, household preferences or the determination or distribution of savings in the economy and so rely on a much smaller set of parameters. Since mortgage market outcomes are determined independently of the more complicated elements of the model which characterize consumption, we fully explore these outcomes before returning to consumption.

5.1 Ergodic distribution

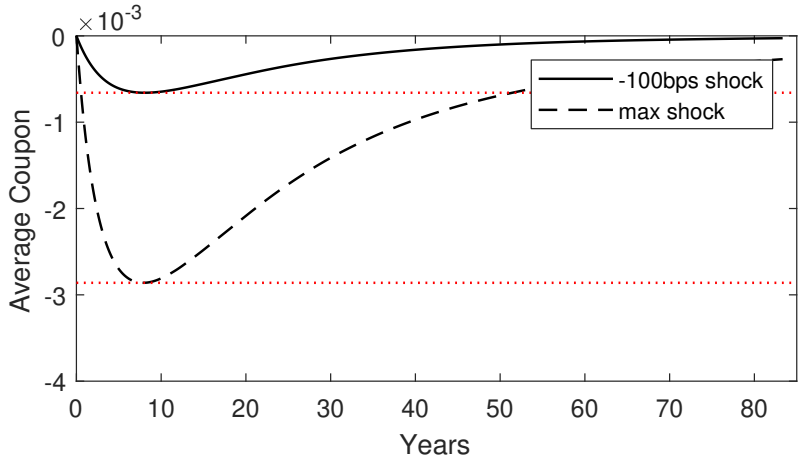
Let time 0 be the current time in the economy, from which we will begin all of our monetary policy experiments. In order to isolate the effects of previous rate histories, most of our experiments will compare two economies which have identical interest rates at time 0 but have different rate histories prior to that date. More specifically, most of our scenarios will compare impulse responses in a stationary ergodic coupon distribution to various alternatives. In order to define the baseline scenario in experiments, we compute the ergodic coupon distribution generated by the CIR process for the short-rate r and the prepayment hazard as calibrated in Section 4.4.2. In order to ensure that current interest rates are identical across experiments and that only the prior path of rates vary, we set the current rate $r_0 = \mu$ in our baseline economy and in most of our other scenarios. Beginning from the initial coupon distribution and r_0, r evolves stochastically and we then compute the impulse response of mortgage market

total prepayment from χ into moves and refinancing as in the data, this implies an average housing tenure of roughly nine years but otherwise delivers identical results for all model outcomes.

outcomes to a decline in r at some point $t \geq 0$.[43] We consider impulse responses to two alternative stimulative policies: 1) A decline in r of 100 basis points (which results in a decline in m of roughly 50 basis points in the baseline economy, as discussed in the Appendix). 2) Lowering r to zero, which we refer to as the "max" monetary stimulus.

Figure 7 shows the impulse response of average coupons to these two stimulus policies in the ergodic baseline computed at $t = 0$.[44] We depict the peak coupon responses to each shock as horizontal lines in red, which we will later use as a reference point when comparing to alternative scenarios. The 100 basis point decline in short rates leads to a peak response of average coupons of around 7 basis points while the max cut leads to a decline of roughly 29 basis points.[45]. We return to a more extensive discussion of magnitudes once we compare to alternative scenarios which more closely match actual rate behavior over the last 30 years and after bringing consumption back into the model. However, we note that these are fairly small but still non-trivial effects given average mortgage balances of $150,000: the max rate cut leads to an average decline in mortgage payments of $435 annually. We will soon see that effects are much larger in some alternative scenarios.

Figure 7: Baseline: IRF of Average Coupon m^* to 100 Basis Point & Max decline in r



5.2 Results: Effects of Secular Mortgage Rate Declines

In our first experiment, we investigate how a mortgage rate path such as we observed from 1992m1 to 2017m4 impacts IRFs. In particular, starting from the observed coupon distribution in 1992m1, we expose this distribution to the sequence of interest rates we observed from 1990 to 2017 to generate the initial distribution for our experiments.[46] The impact of this observed path is summarized in the cross-

[43] In most of our experiments we calculate the impulse response at exactly $t = 0$ but in some cases when they are of interest, we compute impulse responses at later dates.

[44] Since we are beginning from the ergodic distribution, impulse responses are almost identical when computed at any $t > 0$. They are not precisely so because we fix $r_0 = \mu$ rather than using a stochastic value for the initial r so that we can ensure r is identical across experiments. Due to Jensen's inequality and our mortgage pricing equation, this induces a slight upward trend in m and in turn a slight downward trend in IRFs.

[45] Note that the peak responses occur at roughly an 8 year horizon, and Figure 7 clearly indicates a long persistence in the IRF. This long persistence is a combination of the both the persistence of the interest rate shock, with an assumed half-life of 5.28 years in our calibration, and the relatively slow refinancing and prepayment behavior of the consumers.

[46] We pick the sequence of r so that $m(r)$ exactly matches the data, but instead using the observed sequence for r produces similar results.

sectional mortgage distribution. With this distribution in hand, we then follow the same procedure as in the baseline ergodic distribution – that is, we specify a dynamic process governing interest rate dynamics and mortgage pricing going forward. For the initial interest rate at 2017m4, we use r_0 that solves $m(r_0) = m_{2017m4}$ where m_{2017m4} is the empirically observed mortgage rate. Since μ in our CIR process is also calibrated to generate this same rate, this means that $r_0 = \mu$ so that the economy with the downward trend has the same interest rate in period 0 as the baseline ergodic economy but a different rate history. We also assume that both economies are governed by the same stochastic CIR process for r for $t > 0$.[47]

Figure 8 shows the behavior of short-term interest rates in the two economies. In the ergodic distribution, average rates are constant, while in the actual rate series they decline on average from 1992-2017. As previously described, we calibrate the model so that time period 0 corresponds to 2017m4 and interest rates are identical in both economies in period 0. We then compute the impulse response of the economy to a 100 basis point decline in the exogenous short rate r.[48]

Figure 8: Average Rate Behavior in Ergodic and Actual Simulation

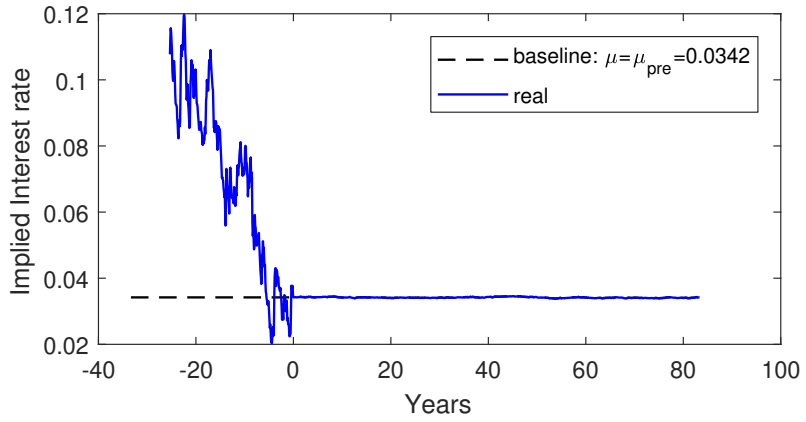


Figure 9 shows the response of prepayment to this rate decline while Figure 10 shows the response of the average coupon m^*. It is clear that both prepayment and average coupons respond much more in the economy calibrated to match actual interest rate declines over the last thirty years than they do in the stationary ergodic economy.

As Figure 11 shows, this is because there are many more households with positive rate gaps after the 30-year downward trend than in the ergodic distribution. This increases $frac > 0$, and as we show now this is the key force which amplifies mortgage responses. In particular, $frac > 0$ is a sufficient-statistic in our model for the response on impact of average coupons to a change in mortgage rates. Let x be the size of a household's mortgage gap $(m^* - m)$ and assume that x has density across households $f(x)$ and that the probability of prepayment for a given x is given by the hazard $h(x)$. Furthermore, assume that h is differentiable except at a discrete number of k points. Caballero and Engel (2007) proves that

[47] This assumption pins down $m(r)$ for $t > 0$. While we feed in an exogenous sequence for $t < 0$ to match the observed rate history, we must still take a stand on the stochastic process for $t < 0$ since this determines $m(r)$. We assume that $mu_{t<0} = mu_{t>0}$ in order to for now isolate effects that come from shifting coupon distribution from effects which come from changes in $m(r)$.

[48] Since we hold the level of current interest rates constant across these two scenarios they both have the same room to cut r to 0. This means that IRFs for the max shock are roughly scaled up versions of the IRFs to the 100 basis point shock and yield the same path-dependence conclusions, so we focus on 100 basis point shocks to simplify exposition. When we turn to scenarios which also vary the level of rates and thus the max room to cut, we show both 100 basis point and max cut IRFs.

Figure 9: IRF of Prepayment to 100 Basis Point decline in r

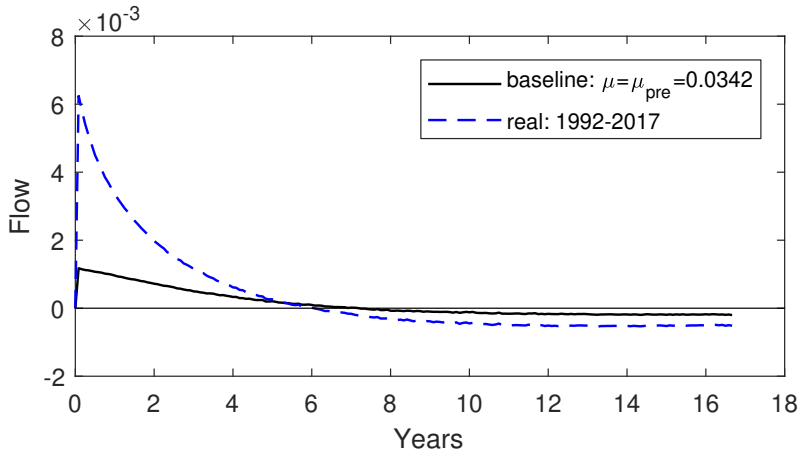


Figure 10: IRF of Average Coupon m^* to 100 Basis Point decline in r

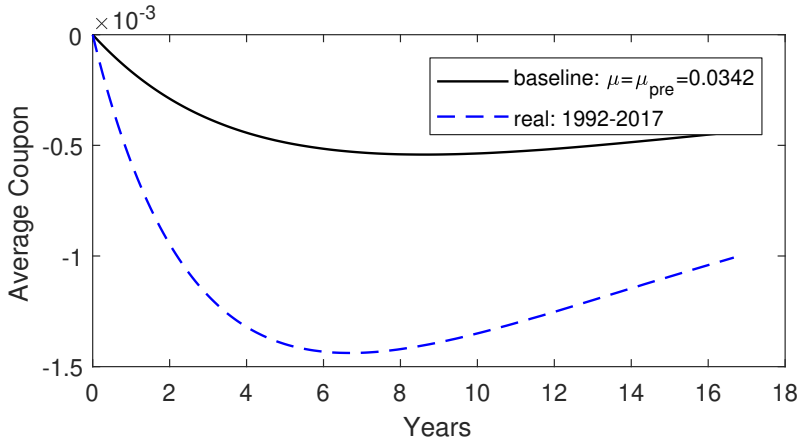
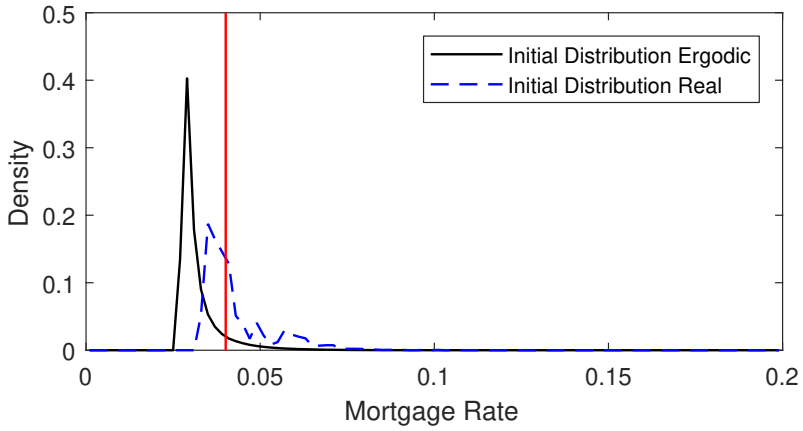


Figure 11: Distribution of m^* in Ergodic and Actual Simulation



in this general state-dependent setup, responses can be characterized as

$$\frac{d\bar{m}^*}{dm} = \int h(x)f(x)dx + \int xh'(x)f(x)dx + \sum_k D_k \tag{3}$$

where

$$D_k \equiv x_k[h(x_k^+)f(x_k^+) - h(x_k^-)f(_k^-)]$$

characterizes the density and jump in the hazard to the left and right of the k potential jumps. In our model, equation 3 can be simplified substantially by using properties of the hazard. Since the hazard is flat aside from the discrete jump up at 0, we know that $h'(x) = 0$ everywhere except 0, where it is undefined. Furthermore, we know that there is a single discrete, finite jump in the hazard at zero but that $D_k \equiv 0 \times [h(0^+)f(0^+) - h(0^-)f(0^-)] \equiv 0$. This then implies that

$$\frac{d\bar{m}^*}{dm} = \int h(x)f(x)dx = freq = \nu + (1 - F(0))\chi. \tag{4}$$

That is, the response of average coupons on impact to a change in the market mortgage rate is entirely given by the frequency of prepayment, which is in turn entirely determined by $frac > 0$, i.e., $(1 - F(0))$.[49] This implies that $frac > 0$ is a sufficient-statistic for determining mortgage coupon responses on impact in our model. Increases in $frac > 0$ increase the frequency of prepayment and lead to greater responses to mortgage rates. It is this effect that explains the differences in Figure 10.

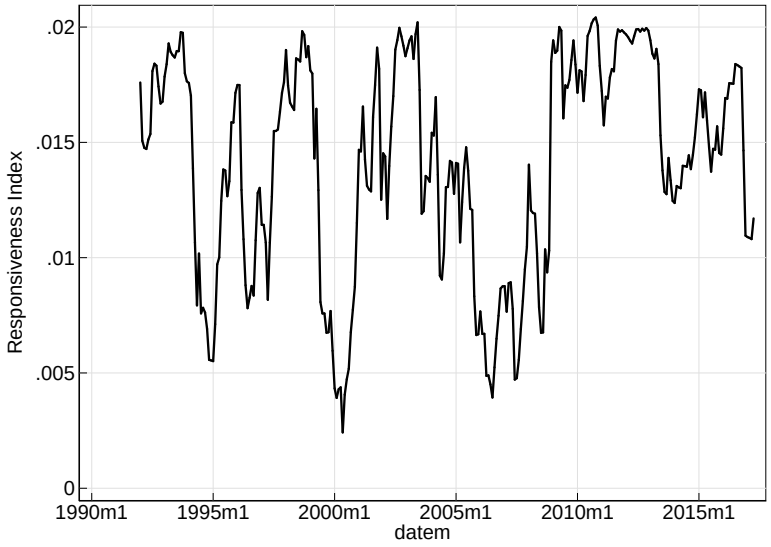
Caballero and Engel (2007) define the first term in 3 as the "intensive margin" contribution and the second two terms as the "extensive margin" contributions. Since the first term is just the frequency of prepayment absent the shock, they label this term the intensive margin because all households who would prepay absent the change in m will lower their coupons by more when they prepay if m falls. The extensive margin contribution gives the additional contribution that arises from changes in the mix of which households prepay in response to the shock. Typically in state-dependent models, these extensive margin effects are large since it is often households with large gaps who are on the margin of adjustment. The strength of these extensive margin effects will also typically depend on the entire distribution of gaps and can be difficult to characterize with closed-form solutions.

In contrast, despite the fact that our model exhibits very strong state-dependence with very different adjustment probabilities for positive and negative gaps, it implies an extensive margin effect which is exactly zero. The reason is because of the unique feature in our model that the only state-dependence occurs exactly at a gap of zero. A small decline in mortgage rates will switch some households from a small negative gap to a small positive gap and thus increase the frequency of prepayment, but since these households have gaps of zero, they contribute a negligible amount on the extensive margin when they adjust. This is a special feature of our model, and in fact it is easy to see from Formula 3 that our model is the *unique* state-dependent model in which responses to shocks can be fully characterized by the intensive margin (frequency) with no need to know other characteristics of the gap distribution. This is the formal sense in which ours is the simplest model of infrequent prepayment which still exhibits state-dependent responses to rate gaps, yet it nevertheless provides a good fit to the data.

Since $frac > 0$ is directly observable in the data, another interesting implication of our model is that we can directly measure the strength of mortgage rate pass-through (on impact) in historical data. In particular, it will be fully determined by $frac > 0$ plus the hazard parameters ν and χ. Figure 12 plots this implied "responsiveness index" from 1992-2017 and shows that there is substantial variation, with

[49] Note that this is a sufficient-statistic for impact effects but that the full dynamics of the impulse response are more complicated and require additional information. Note also that "impact" effects depend on the period over which time is measured. While our model is continuous time, it is natural to interpret these formulas at monthly intervals since that is the frequency at which we measure hazards and gap distributions in the data.

Figure 12: Responsiveness Index in Historical Data



max responsiveness to interest rate shocks nearly four times the minimum response. This is consistent with the strong empirical interaction between Δ FRM and $frac > 0$ documented in Table 4.[50]

Together the results in this subsection show that the prepayment channel of monetary policy has been substantially amplified by the downward trend in mortgage rates over the last thirty years, since this trend has increased the mass of households with positive gaps. This trend is unlikely to continue forever, and our results show that if mortgage rates return to a stable or increasing environment, there will be substantial headwinds on future interest rate stimulus.

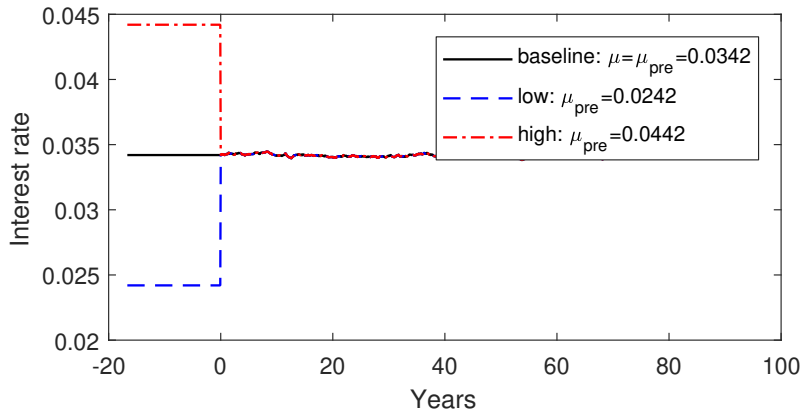
5.3 Results: Effects of Past Stimulus and Tightening

We next turn to the effects of past stimulus and tightening cycles. In particular, we compare impulse responses to 100 basis point declines in interest rates in the baseline ergodic economy to a scenario in which previous rates were high for an extended period of time and to a scenario in which previous rates were low for an extended period of time. In all three economies, current interest rates are again set identical in period 0 so that the economies only differ in their past paths.[51] Figure 13 illustrates the behavior of average rates in the three scenarios. The economy with previously high interest rates illustrates the effectiveness of monetary policy following a period of tight interest rates while the economy with previously low interest rates illustrates the effectiveness of monetary policy after a previous easing cycle. One can think of the latter economy as illustrating the effects of ending the long period of low interest rates after the Great Recession.

[50] With our endogenous pricing results, we can also define in a stable environment the monetary policy pass-through from short-rate interest rates to mortgage rates: $\frac{d\bar{m}^*}{dr} = \frac{d\bar{m}^*}{dm} m'(r)$, where $m'(r)$ is defined by the structural model.

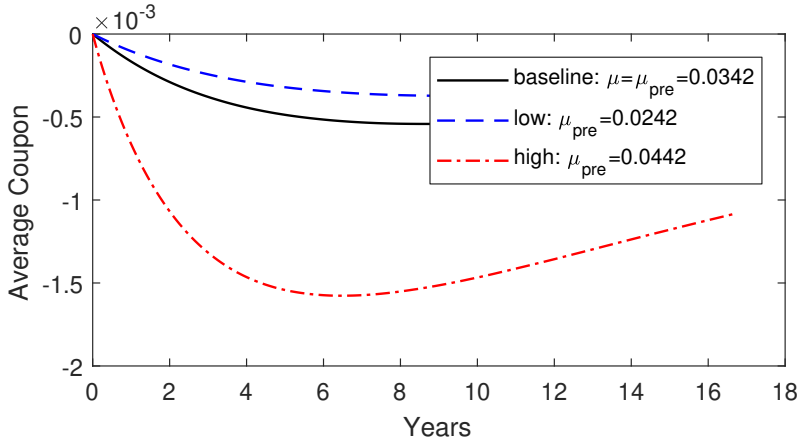
[51] The interest rate process and thus mortgage pricing are again the same in all three economies for $t > 0$. In the pre-period, we simulate r from a process with a μ which is either temporarily 100 basis point higher or lower. We interpret the period $t < 0$ as reflecting a long but not permanent level difference in mortgage rates, and as such, we assume that the mortgage pricing function is the same both before and after period 0 even though the stochastic process is identical. This isolates effects of the rate path from changes in mortgage pricing. We turn to mortgage pricing effects below.

Figure 13: Average Rate Behavior in Ergodic Vs. Prior Stimulus or Tightening



We again find substantial path-dependence effects: Figure 14 shows that average coupons respond much more strongly to the same 100 basis point decline in r in the economy with previously high rates than in the economy with previously low rates. The intuition is the same as before: in the economy with previously high rates, $frac > 0$ is much larger today for the same m than in the economy with previously low rates.[52] This shows that as the Fed exits the extended period of low interest rates, it is likely to have reduced stimulus ability. Furthermore, this figure also illustrates an asymmetry between rate increases and decreases: in the economy where rates decline, there is a larger absolute change in responsiveness than in the economy where rates increase. This is because in the stationary ergodic distribution, there are already few households with positive gaps (see Figure 11). When rates increase, $frac > 0$ still declines, but not by that much.[53] In contrast, in the economy where rates fall, there is a large increase in $frac > 0$ and thus responsiveness of m^* to interest rate shocks today. This illustrates that there are important asymmetries between increasing and decreasing rate environments. Our next set of experiments explore this asymmetry in more detail.

Figure 14: IRF of Average Coupon m^* to 100 Basis Point decline in r



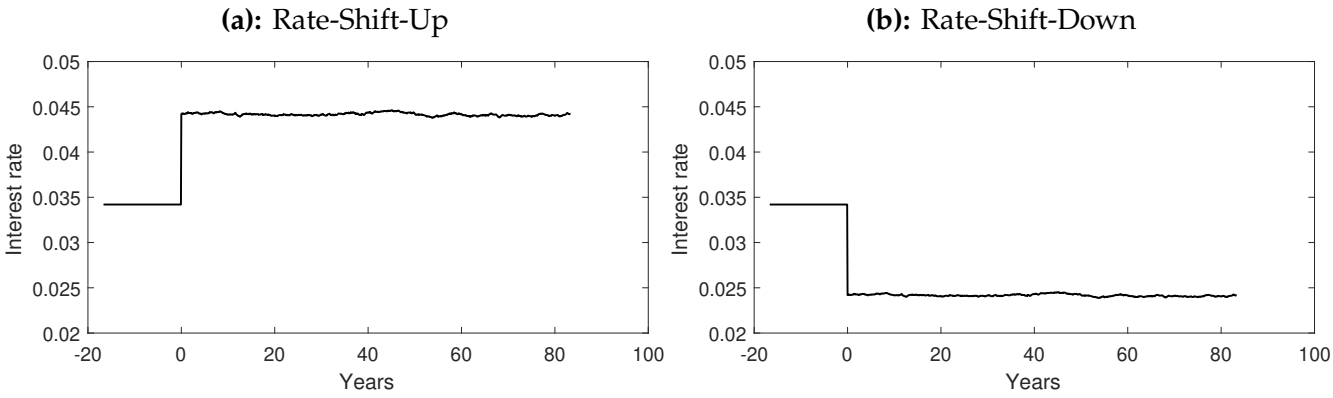
[52]Note that due to the same dynamics going forward across the three specification and the common starting point r_0 there is no difference in $m'(r_0)$.

[53]For a large enough increase in the rate regime, $frac > 0$ will hit zero and further increases have no effect on responsiveness.

5.4 Results: Asymmetries and Reloading

Results thus far focus on 100 basis point interest rate shocks immediately after shifts in interest rate regimes. In all of the prior experiments, rates in period 0 and the future interest rate dynamics were identical, but the coupon gap distribution differed due to differences in past rate history. In this section, we illustrate dynamic effects of shifting interest rate regimes and consider more dramatic stimulus policies. In particular, we consider two different economies which are both initially in an ergodic steady-state with an average r of 3.4%, as in our previous exercises and then we permanently increase interest rates in one economy and permanently decrease them in the other. In particular, in period 0 in Economy "Rate Shift Up", both the current r and the mean of the stochastic process governing future rates increase by 100 basis points.[54] Conversely, in Economy "Rate-Shift-Down", in period 0 the current r as well as the mean of the stochastic process r decrease by 100 basis points. Lastly, in Economy "Baseline", there is no change in period 0 to the dynamics. Figure 15 illustrates the average behavior of rates in the two economies in which the dynamics shift.

Figure 15: Average Rate Behavior in Rate-Shift-Up vs. Rate-Shift-Down Economies



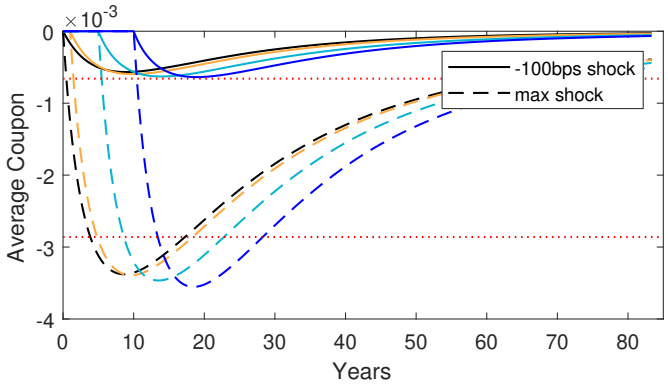
We then consider two different monetary policy experiments for all economies and compute their effects when enacted at different dates. As before, we consider both a 100 basis point decline in r as well as lowering r all the way to zero (which we label the "max" shock). We previously illustrated the impulse response to both the 100 basis point and the max shock for the baseline economy in Figure 7. Analogously, Figure 16 shows the impulse response to both the 100 basis point and the max shock in the regime shift economies, noting that the regime shift occurs in period 0. The two horizontal dashed red lines in both Figure 7 and Figure 16 show the peak coupon response to the 100 basis point and the max shock occurring at time 0 in the baseline economy with no regime shift. Thus, they are helpful in comparing the coupon responses in the regime shift economies to the baseline with no regime shift. The date at which the impulse starts indicates how far after the regime shift the monetary policy shock occurs. For example, the dashed yellow line in the left panel shows the effect of cutting rates from 4.4% to 0% one year after the regime shift from 3.4% to 4.4%. The dashed yellow line in the right panel shows the effect of cutting rates from 2.4% to 0% one year after the regime shift from 3.4% to 2.4%. The solid yellow lines again show impulse responses one year after the regime shifts, but show the effects of 100

[54] While shifting the mean of the interest rate process also changes mortgage pricing, we leave mortgage pricing constant using μ_{pre}. Thus, we interpret this as a change in future Fed policy which is not announced and thus not priced into the market mortgage rate. We show the effects of changes in mortgage pricing next.

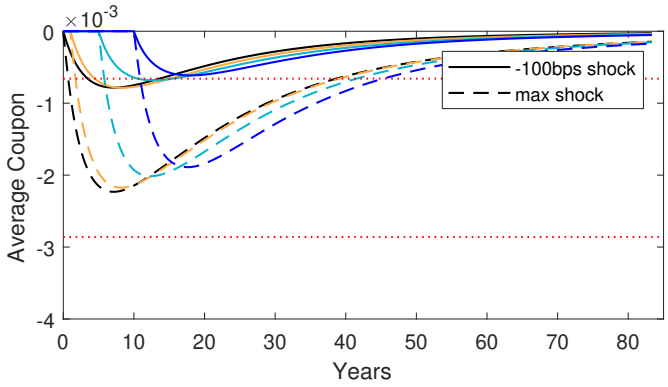
basis point cuts rather than cuts to zero.

Figure 16: Regime Shift: Average Coupon m^* to 100 Basis Point & Max decline in r

(a): Rate-Shift-Up



(b): Rate-Shift-Down



There are a number of takeaways from this figure. First, looking at 100 basis point responses on impact in black as compared to the baseline peak responses in red, we can see that the first 100 basis point declines in r initially have smaller effects on coupons in the Rate-Shift-Up economy than in the Rate-Shift-Down economy. This arises exactly from the effects emphasized in the previous experiments: when rates rise, $frac > 0$ goes down and this reduces the effects of a given change in r. However, looking at dynamics, we can see that over time, 100 basis point shocks become more powerful in the Rate-Shift-Up economy and less powerful in the Rate-Shift-Down economy.

Second, the maximum stimulus power (given by cutting r to 0), increases in the Rate-Shift-Up economy, but decreases in the Rate-Shift-Down economy. A higher average level of interest rates naturally results in a larger max cut and resulting response.

Third, the convergence rate with respect to different timing of the monetary shocks is asymmetric – the maximum stimulus available to the Fed converges to a new (lower) level in the Rate-Shift-Down economy and to a new (higher) level in the Rate-Shift-Up economy, but convergence is slower in the Rate-Shift-Down than Rate-Shift-Up economy. That is, the impulse response functions for a shock occurring in year 0, 1, 5 and 10 differ more from each other in the Rate-Shift-Down economy than in the Rate-Shift-Up Economy. The peak of the impulse response to cutting rates to zero in the baseline is 29 basis points. In the Rate-Shift-Up economy, this peak eventually converges to 44 basis points, and in the Rate-Shift-Down economy, this peak eventually converges to 18 basis points. Figure 16 shows that in the Rate-Shift-Down economy, impulse responses have converged to the new steady-state in less than 10 years while in the Rate-Shift-Down economy, impulse responses have only reached half of their new steady-state value after 10 years. This is because in response to rate declines, households rapidly prepay, causing the coupon distribution to converge to the new ergodic stationary distribution, whereas rate increases affect the coupon distribution only slowly through exogenous moves which force prepayment.

This experiment thus illustrates a strong asymmetry in Fed stimulus policy through the prepayment channel: when the Fed raises rates it reloads "ammunition", but it does so slowly. When the Fed lowers rates, it uses up its stimulus "ammunition" more rapidly than this reloading occurs.

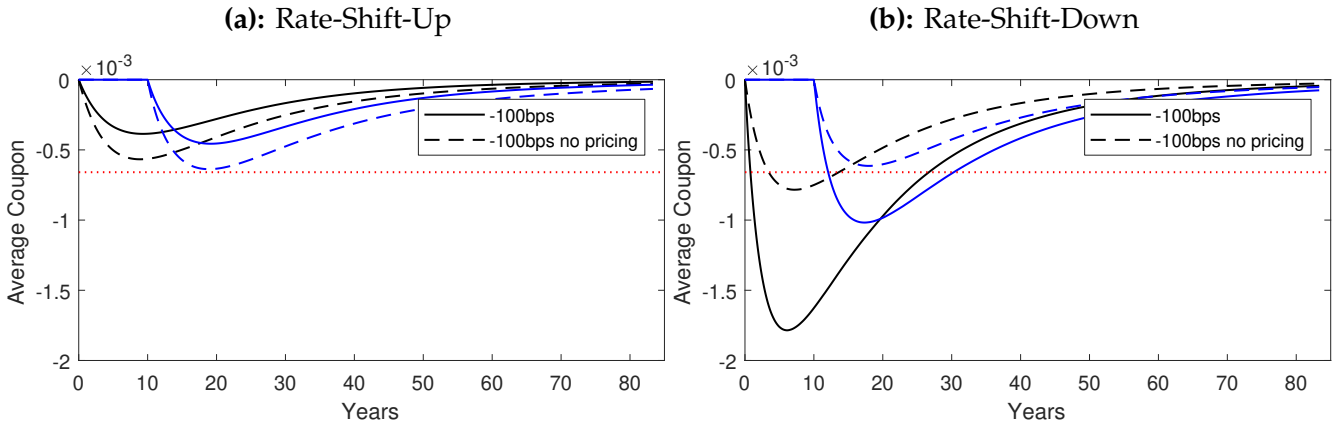
Overall, these results suggest that the Fed has much less room to stimuluate the economy today

than might otherwise be suggested by current rates.

5.5 Mortgage Pricing and Fed Information

In the previous experiment, we assumed that the stochastic process for r shifted up or down, but we did not shift mortgage pricing. We interpret this as a shift in the future behavior of Fed policy which is unannounced: in period 0, the market observes that r_0 jumps up, but it does not know that the stochastic process determining the future evolution of r has changed, and so it does not adjust the mortgage pricing function. That is, r_0 and thus $m_0 = m(r_0)$ changes, but there is no change in the function $m(.)$ which translates short rates to long rates.

Figure 17: Regime Shift: Average Coupon m^* to 100 bp decline in r with and without announcement



In Figure 17, we again do the same regime shift illustrated in Figure 15, but we now assume that the Fed announces that it is moving to a new higher or lower interest rate regime at the time of the regime shift so that the endogenous mortgage pricing function also changes. This figure shows that endogenous pricing amplifies the effects of regime shifts on the effectiveness of monetary policy. As we showed before, with no change in the pricing function, 100 basis point cuts are initially more powerful after r shifts down. This is because $frac > 0$ rises. When we also change mortgage pricing, this effect becomes even stronger. This is because the same regime shift down in r leads to a larger decline in $m(r)$ and thus a larger increase in $frac > 0$ when the market realizes that this shift is permanent and prices it into the mortgage rate. Conversely, 100 basis point cuts become less effective after regime shifts up, and this is amplified when these changes in the stochastic process are priced, since this leads to a larger increase in m and decline in $frac > 0$.

Of course, this is a simple example, and our model leaves out many important features that likely matter for determining the importance of Fed information on mortgage market activity. For example, with an exogenous prepayment hazard we cannot account for potentially important effects of Fed signals on household expectations of future rates potentially changing prepayment decisions today. Nevertheless, the results in this subsection suggest that mortgage pricing can potentially interact importantly with Fed information to affect the transmission of monetary policy.

6 Consumption results

The previous section shows strong path-dependence in the response of mortgage related outcomes to monetary policy. While we focus primarily on these mortgage related outcomes, this section explores the implications of this mortgage market path-dependence for path-dependence in spending. Since the conclusions for spending mirror those for mortgage outcomes, for brevity we focus primarily on consumption responses to 100 basis point short-term rate declines under the two different interest rate scenarios in 5.2. As before, in one scenario the economy is initially in the ergodic steady-state while in the other the prior path of rates follows the actual sequence of interest rates observed from 1992-2017.

Aggregate consumption in the ergodic distribution increases by 76 basis points on impact in response to the 100 basis point decline in short-term interest rates while it increases by 98 basis points in response to the same rate decline after the actual sequence of rates observed from 1992-2017. What drives these consumption responses? In our model, both short-rate changes and the resulting changes in mortgage rates induce effects on consumption: declines in short-rates cause income and substitution effects on savings together with a wealth effect from the change in the present value of labor income while declines in mortgage rates lead to declines in monthly mortgage payments. The response of rich households (net-lenders) to interest rate shocks is largely driven by the direct effects of short-term interest rates on consumption while the response of very poor hand-to-mouth households (net-borrowers) is driven by the behavior of m^*.

We can isolate role of the refinancing channel for spending by solving a version of the model in which r declines but households do not have the option to refinance.[55] When we re-compute impulse responses to a 100 basis point decline in r in the model with no refinancing option, aggregate consumption in the ergodic distribution increases by 48 basis points and aggregate consumption increases by 37 basis points in the model with 1992-2017 sequence of interest rates.

This shows that the refinancing channel accounts for 37% of the total consumption response in the baseline ergodic economy but that the refinancing channel increases to 62% of consumption responses in the 1992-2017 economy.[56] What does this imply for the drivers of path-dependence? In our model, there are two forces for path-dependence. We have emphasized path-dependence effects which arise through the mortgage market, but there is a second source of path-dependence which arises entirely from the behavior of the short-rate r. The short-rate follows a Markov-process and so has no path-dependence, but r influences aggregate savings and this induces an element of path-dependence into consumption dynamics even when the refinancing channel is shut down. Figure 18 shows that households in the economy with the rate series matched to 1992-2017 rates accumulate more savings than in the ergodic economy. This lowers MPCs and is a force for *lower* consumption responses in the actual series than in the ergodic series. This explains why the ergodic economy with no refinancing exhibits slightly higher consumption responses to rate changes (48 basis points) than the 1992-2017 economy with no refinancing (37 basis points).

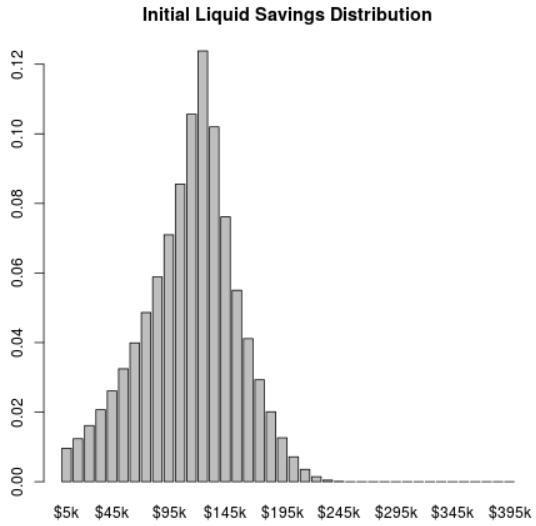
Overall, the refinancing channel leads to an increase in the consumption response in the 1992-2017

[55] More specifically, for $t < 0$ we solve our baseline model with refinancing in order to generate the distribution of household states in period 0, but we then solve for a new policy function for $t > 0$ in which the hazard of refinancing is set to zero. This eliminates the refinancing channel but leaves the distribution of household states at $t = 0$ all other channels which affect spending unchanged.

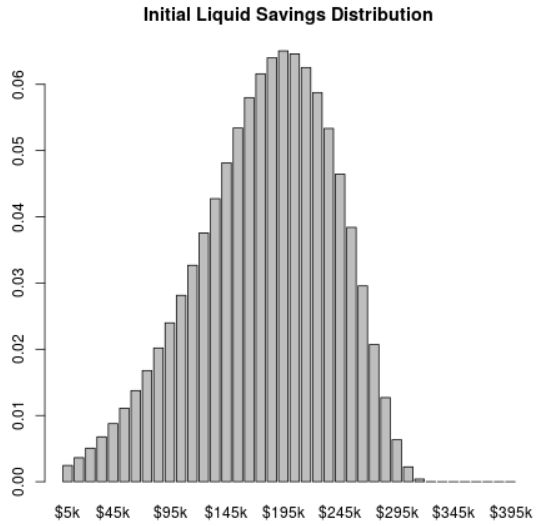
[56] 0.37=1-0.48/0.76 and 0.53=1-0.37/0.98.

Figure 18: Heterogeneity related to initial liquid savings

(a): Ergodic



(b): Real



economy of 61 basis points while it leads to an increase in the consumption response in the ergodic economy of only 33 basis points.[57] This means that the refinancing channel accounts for 150% of the IRF increase when moving from the ergodic to the 1992-2017 rate distribution. The savings effects account for -50% of the variation in IRFs. Savings effects are comparatively small because even though movements in past r generate path-dependence for responses to current r by affecting liquid wealth, the MPC varies little with liquid wealth for lenders who are far from the borrowing constraint. In contrast, constrained borrowers with an MPC of 1 have consumption which responds one-for-one to reductions m^*, and m^* responses to declines in r are much larger when $frac > 0$ is large. Figure 11 shows that households have much larger outstanding coupons m^* under the 1992-2017 rate distribution which in turn makes m^* and resulting consumption more sensitive to changes in rates.

Importantly, our model captures the fact that persistent interest rate declines lead to a lower return on rich households' savings, as emphasized by Greenwald (2017). However, since we assume the return on liquid savings is exogenous and equal to the Fed Funds rate, pass-through in our model is constant and does not depend on the fraction of borrowers refinancing. This means that the financial intermediary is subject to interest rate risk and suffers path-dependent profits and losses which depend on the coupon distribution in the economy. Our assumption that the financial intermediary is owned by foreign investors means that rich households do not adjust their consumption in response to these valuation changes. We make this assumption for tractability since it avoids having to forecast the equilibrium endogenous distribution of households using Krusell and Smith (1998) type methods. In reality, foreign mortgage credit is important, but many domestic households are also affected by shocks to profits in the mortgage market.[58] Since mortgage rate contracts are long-duration, these shocks are very persistent and can thus result in consumption offsets by lenders even if lenders are unconstrained

[57] 61=98-37 and 33=81-48.

[58] Beraja, Fuster, Hurst and Vavra (2018) argue that foreigners hold roughly 50% of the ultimate claims to mortgage debt.

permanent income consumers. Could these additional offsets be enough to undo the aggregate effects of refinancing on consumption?

To assess this in a simple way, we can look at the effects of these profit shocks on foreigners' consumption under the assumption that the deep pocketed foreigners are simple permanent income consumers who consume the annuity value of expected mortgage payments. Under this assumption, foreigners' consumption will decline more in response to a 100 basis point interest rate cut after the 1992-2017 rate history than in the ergodic distribution. If we include 100% of foreign lenders' consumption in our calculations as an upper bound on the role of these lender offsets, then the difference in the IRF between the 1992-2017 and baseline economy falls from 22 basis points to 14 basis points. This means that even in this unrealistic case in which domestic consumption responds 1-1 to profit shocks in the mortgage sector, the consumption response is still different after these alternative rate histories.[59]

7 Conclusion

The Great Recession led to a long period of low interest rates. While this helped stimulate the economy, the Fed has begun raising rates, in part so that it has ammunition to stimulate the economy in response to future recessions.[60] In this paper we argue that looking only at the level of *current* rates provides an incomplete view of the Fed's stimulative power, and that it may take an extended period of time with elevated rates before the Fed regains substantial ability to stimulate the economy.

In particular, we argue that the presence of vast amounts of US household debt in fixed-rate prepayable mortgage contracts leads to path-dependent consequences of monetary policy and thus stimulus power which depends on both current and past rates, since outstanding mortgage rates will depend importantly on the minimum rate obtained in the recent past. Using a simple model of mortgage prepayment embedded in a standard incomplete markets setup which we match to empirical loan-level prepayment relationships, we argue that even if the Fed raises rates substantially before the next recession arrives, it will likely have less ammunition available for stimulus than in recent recessions. In particular, we argue that: 1) The strength of monetary policy through mortgage prepayment has been amplified by the 30-year secular decline in mortgage rates. 2) Holding current rates constant, monetary policy is less effective if previous rates were low. 3) Monetary policy "reloads" stimulative power very slowly after raising rates. All three conclusions imply that the Fed is likely to face substantial headwinds for stimulus through mortgage markets over the next several years.

Our analysis focuses on path-dependence arising from fixed rate mortgages with an option to refinance. Mortgage contracts with fixed rates and no prepayment penalty are prevalent in the US and Denmark, but are otherwise uncommon. However, path-dependence can also arise even in fixed rate mortgage environments with large prepayment penalties, due to endogenous timing of home purchases and terms of mortgage contracts.[61] Interest rate cuts will encourage new home purchases and mort-

[59] Straub (2018) provides evidence that the consumption of rich households does not respond 1-1 to permanent shocks, which would also weaken the importance of this channel even with no foreign lenders.

[60] https://www.bloomberg.com/news/articles/2018-06-18/summers-warns-biggest-economies-not-prepared-for-next-downturn

[61] For example, France, Canada, Germany, Japan, the Nertherlands and Switzerland all have substantial mortgage debt with rates fixed for one-year or longer with varying degrees of penalties for prepayment. The exact size of the prepayment penalty will also have effects on incentives to refinance and is not necessarily prohibitive. In Germany prepayment penalties are waived when moving.

gages, but this will have less effect if rates were already low in the recent past. Furthermore, fixed rate mortgage terms in many countries are 5-10 years, and loans then refinance after this fixed period. This means that as interest rates affect the fraction of households starting new mortgage contracts in a given year, this will have echo effects which affect future monetary policy 5-10 years in the future.[62] Finally, current monetary policy can affect the incentives to take out fixed vs. variable rate mortgages in countries where both contracts are common, which will then affect future monetary policy. Thus, the dynamics of path-dependence are likely more complicated in other countries, but our insight that the past history of rates matters for current transmission of monetary policy through the mortgage market is not specific to the US. An exploration of these dynamics and the quantitative size of path-dependence in countries with different mortgage contracts is an interesting avenue for future research.

References

Abel, Joshua, and Andreas Fuster. 2018. "How Do Mortgage Refinances Affect Debt, Default, and Spending? Evidence from HARP."

Agarwal, Sumit, Gene Amromin, Souphala Chomsisengphet, Tim Landvoigt, Tomasz Piskorski, Amit Seru, and Vincent Yao. 2017. "Mortgage Refinancing, Consumer Spending, and Competition: Evidence from the Home Affordable Refinancing Program." National Bureau of Economic Research Working Paper 21512.

Agarwal, Sumit, Gene Amromin, Souphala Chomsisengphet, Tomasz Piskorski, Amit Seru, and Vincent Yao. 2017. "Mortgage Refinancing, Consumer Spending, and Competition: Evidence from the Home Affordable Refinancing Program." *NBER Working Paper No. 21512*.

Agarwal, Sumit, John Driscoll, and David Laibson. 2015. "Optimal Mortgage Refinancing: A Closed-Form Solution." *Journal of Money, Credit and Banking*, 45(4): 591–622.

Aiyagari, S Rao. 1994. "Uninsured idiosyncratic risk and aggregate saving." *The Quarterly Journal of Economics*, 109(3): 659–684.

Auclert, Adrien. 2017. "Monetary policy and the redistribution channel." National Bureau of Economic Research.

Barles, Guy, and Panagiotis E Souganidis. 1991. "Convergence of approximation schemes for fully nonlinear second order equations." *Asymptotic analysis*, 4(3): 271–283.

Beraja, Martin, Andreas Fuster, Erik Hurst, and Joseph Vavra. 2018. "Regional Heterogeneity and the Refinancing Channel of Monetary Policy." *Quarterly Journal of Economics*.

Berger, David, and Joseph Vavra. 2015. "Consumption Dynamics During Recessions." *Econometrica*, 83(1): 101–154.

Berger, David, and Joseph Vavra. 2018*a*. "Dynamics of the U.S. Price Distribution." *European Economic Review*, 103.

Berger, David, and Joseph Vavra. 2018*b*. "Shocks vs. Responsiveness: What Drives Time-Varying Dispersion?" *Journal of Political Economy*.

[62] For example, the Bank of Canada Financial System Review in November 2017 expresses concern that 47% of all Canadian mortgages will refinance in 2018 since many households took out new mortgages when rates were low in 2013.

Berger, David, Konstantin Milbradt, and Fabrice Tourre. 2018. "A Theory of Mortgage Rate Pass-Through."

Berger, David, Veronica Guerrieri, Guido Lorenzoni, and Joseph Vavra. 2018. "House Prices and Consumer Spending." *Review of Economic Studies*, 85(3): 1502–1542.

Caballero, Ricardo, and Eduardo Engel. 2007. "Price stickiness in Ss models: New interpretations of old results." *Journal of Monetary Economics*, 54.

Campbell, John, and Joao Cocco. 2013. "A Model of Mortgage Default." *The Journal of Finance*, 70(4): 1495–1554.

Cox, John C, Jonathan E Ingersoll Jr, and Stephen A Ross. 2005. "A theory of the term structure of interest rates." In *Theory of Valuation*. 129–164. World Scientific.

Di Maggio, Marco, Amir Kermani, and Christopher Palmer. 2016. "How Quantitative Easing Works: Evidence on the Refinancing Channel." NBER Working Paper 22638.

Di Maggio, Marco, Amir Kermani, Benjamin J. Keys, Tomasz Piskorski, Rodney Ramcharan, Amit Seru, and Vincent Yao. 2017. "Interest Rate Pass-Through: Mortgage Rates, Household Consumption, and Voluntary Deleveraging." *American Economic Review*, 107(11): 3550–88.

Eichenbaum, Martin, Sergio Rebelo, and Arlene Wong. 2018. "State Dependency and the Efficacy of Monetary Policy: The Refinancing Channel."

Ganong, Peter, and Pascal Noel. 2017. "The Effect of Debt on Default and Consumption: Evidence from Housing Policy in the Great Recession." Harvard University Working Paper.

Gertler, Mark, and Peter Karadi. 2015. "Monetary Policy Surprises, Credit Costs, and Economic Activity." *American Economic Journal: Macroeconomics*, 7(1): 44–76.

Greenwald, Daniel L. 2017. "The Mortgage Credit Channel of Macroeconomic Transmission." MIT Sloan Working Paper.

Greenwald, Daniel, Tim Landvoigt, and Stijn Van Nieuwerburgh. 2018. "Financial Fragility with SAM?"

Guren, Adam M, Alisdair McKay, Emi Nakamura, and Jon Steinsson. 2018. "Housing Wealth Effects: The Long View." National Bureau of Economic Research Working Paper 24729.

Guren, Adam M, Arvind Krishnamurthy, and Timothy J McQuade. 2018. "Mortgage Design in an Equilibrium Model of the Housing Market." National Bureau of Economic Research Working Paper 24446.

Hedlund, Aaron, Fatih Karahan, Kurt Mitman, and Serdar Ozkan. 2017. "Monetary Policy, Heterogeneity and the Housing Channel."

Hurst, Erik, Benjamin J. Keys, Amit Seru, and Joseph S. Vavra. 2016. "Regional Redistribution Through the US Mortgage Market." *American Economic Review*, 106(10).

Justiniano, Alejandro, Giorgio E Primiceri, and Andrea Tambalotti. 2017. "The Mortgage Rate Conundrum." National Bureau of Economic Research Working Paper 23784.

Kaplan, Greg, Kurt Mitman, and Gianluca Violante. 2016. "Non-durable Consumption and Housing Net Worth in the Great Recession: Evidence from Easily Accessible Data." *NBER Working Papers 22232*.

Keys, Benjamin J., Devin G. Pope, and Jaren C. Pope. 2016. "Failure to refinance." *Journal of Financial Economics*, 122(3): 482 – 499.

Krusell, Per, and Anthony Smith. 1998. "Income and Wealth Heterogeneity in the Macroeconomy." *Journal of Political Economy*, 106(5): 867–896.

Mian, Atif, and Amir Sufi. 2012. "The Effects of Fiscal Stimulus: Evidence from the 2009 Cash for Clunkers Program." *Quarterly Journal of Economics*, 1107–1142.

Mian, Atif, Kamelesh Rao, and Amir Sufi. 2013. "Household Balance Sheets, Consumption, and the Economic Slump." *Quarterly Journal of Economics*, 1687–1726.

Nakamura, Emi, and Jon Steinsson. 2018. "High-Frequency Identification of Monetary Non-Neutrality: The Information Effect*." *The Quarterly Journal of Economics*, 133(3): 1283–1330.

Quigley, John M. 1987. "Interest Rate Variations, Mortgage Prepayments and Household Mobility." *The Review of Economics and Statistics*, 69(4): 636–643.

Scharfstein, David, and Adi Sunderam. 2016. "Market Power in Mortgage Lending and the Transmission of Monetary Policy."

Straub, Ludwig. 2018. "Consumption, Savings, and the Distribution of Permanent Income."

Stroebel, Johannes, and Joseph Vavra. 2018. "House Prices, Local Demand, and Retail Prices." *Journal of Political Economy*, forthcoming.

Vavra, Joseph. 2014. "Inflation Dynamics and Time-Varying Volatility: New Evidence and an Ss Interpretation." *Quarterly Journal of Economics*, 129(1).

Winberry, Thomas. 2016. "Lumpy Investment, Business Cycles and Stimulus Policy." University of Chicago Working Paper.

Wong, Arlene. 2018. "Population Aging and the Transmission of Monetary Policy to Consumption." Northwestern University Working Paper.

Table 1: Effects of Rate Gaps on Prepayment Propensities

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
frac > 0	2.051***			7.861**	5.117*	2.058***	2.696***	2.066***
	(0.386)			(3.650)	(2.714)	(0.357)	(0.866)	(0.394)
frac > 50bp		1.989***						
		(0.389)						
frac > 100bp			2.249***					
			(0.450)					
mean CLTV				-0.975	-2.130			
				(2.456)	(1.991)			
(frac > 0) $\times$ mean CLTV				-8.158	-4.259			
				(5.281)	(3.978)			
$\mathbb{1}_{year=2003}$					1.737***			
					(0.621)			
Constant	0.0416	0.365**	0.673***	0.228	1.062	-0.0455	-0.0264	-0.398
	(0.196)	(0.145)	(0.107)	(1.646)	(1.304)	(0.194)	(0.401)	(0.298)
Adj. R^2	0.282	0.277	0.246	0.488	0.623	0.558	0.298	0.608
N	304	304	304	304	304	108	120	76
Date Range	92-17m4	92-17m4	92-17m4	92-17m4	92-17m4	92-00	01-10	11-17m4

Newey-West standard errors in parantheses. *=10%, **=5%, ***=1% significance. Mean LTV is the ratio of a loan's outstanding balance to value estimated using appraisal values at origination updated using local house price indices from CoreLogic. Loan level data from McDash Performance data+appraisal values from McDash origination data is used to calculate LTV. Prepayment fractions are measured in month $t+1$ while rate incentives and LTV are measured in month t, since McDash data measures origination not application and there is a 1-2 month lag from application to origination.

Table 2: Effects of Rate Gaps on Prepayment Propensities by Type

	(1) Total	(2) Rate	(3) Cashout	(4) Purchase
frac > 0	2.053***	1.261***	0.296**	0.400***
	(0.399)	(0.187)	(0.128)	(0.132)
Constant	0.0424	-0.411***	0.0670	0.380***
	(0.199)	(0.0948)	(0.0672)	(0.0690)
Adj. R^2	0.280	0.418	0.0713	0.101
N	292	292	292	292
Date Range	93-17m4	93-17m4	93-17m4	93-17m4

Newey-West standard errors in parantheses. *=10%, **=5%, ***=1% significance. Prepayment is measured using loan level data from McDash Performance data. After 2005, we decompose prepayment by type using CRISM data which links new and old loans. Prior to 2005, we decompose prepayment by type using origination shares by type from CoreLogic LLMA data. Regressions begin in 1993 rather than 1992 since reliable CoreLogic origination data on prepayment shares does not begin until 1993. See Appendix for additional discussion. Prepayment fractions are measured in month $t+1$ while rate incentives and CLTV are measured in month t, since McDash data measures origination not application and there is a 1-2 month lag from application to origination.

Table 3: Robustness: Instrumenting for Rate Gaps with High Frequency Monetary Policy Shocks

	(1) Total	(2) Rate	(3) Cashout	(4) Purchase
frac > 0	2.762**	1.259***	0.874**	0.657
	(1.200)	(0.455)	(0.418)	(0.484)
Constant	-0.322	-0.393	-0.269	0.235
	(0.769)	(0.282)	(0.279)	(0.314)
N	235	235	235	235
Date Range	93-12m7	93-12m7	93-12m7	93-12m7

Newey-West standard errors in parantheses. *=10%, **=5%, ***=1% significance. This table instruments for $frac > 0$ using the sum over the prior 6 months of the high frequency monetary policy shock series from Gertler and Karadi (2015), available through 2012m7. After 2005, we decompose prepayment by type using CRISM data which links new and old loans. Prior to 2005, we decompose prepayment by type using origination shares by type from CoreLogic LLMA data. Regressions begin in 1993 rather than 1992 since reliable CoreLogic origination data on prepayment shares does not begin until 1993. Prepayment fractions are measured in month $t+1$ while rate incentives and CLTV are measured in month t, since McDash data measures origination not application and there is a 1-2 month lag from application to origination.

Table 4: Effects of FRM Changes and Gaps on Average Coupon Changes

	(1)	(2)	(3)	(4)
frac > 0	-0.0486***		-0.0470***	-0.0531***
	(0.00579)		(0.00531)	(0.00654)
Δ FRM		0.0188***	-0.0286**	0.0369
		(0.00516)	(0.0124)	(0.0356)
Δ FRM × (frac > 0)			0.0542**	0.0679***
			(0.0231)	(0.0244)
Δ FRM × mean LTV				-0.118*
				(0.0673)
mean LTV				0.0521**
				(0.0217)
Constant	0.0180***	-0.0142***	0.0175***	-0.0109
	(0.00284)	(0.00178)	(0.00273)	(0.0111)
Adj. R^2	0.497	0.0497	0.516	0.560
N	303	303	303	303
Date Range	92-17m3	92-17m3	92-17m3	92-17m3

Newey-West standard errors in parantheses. *=10%, **=5%, ***=1% significance. Average LTV is average across loans of the ratio of a loan's outstanding balance to value estimated using appraisal values at origination updated using local house price indices from CoreLogic. Loan level data from McDash Performance data+appraisal values from McDash origination data is used to calculate LTV. Δ FRM is the change in the current 30 year FRM is the monthly average of the Freddie Mac weekly PMMS survey 30 year fixed rate mortgage average: https://fred.stlouisfed.org/series/MORTGAGE30US. To account for a lag between application and origination, in all specifications, $frac > 0$ and LTV is measured as of month t, Δm^* is measured between month t and month $t+1$ and Δ FRM is measured between month $t-1$ and month t.

Table 5: Effects of Rate Gaps on Prepayment Propensities by MSA

	(1) Tot Prepay	(2) Rate-Refi	(3) Cashout	(4) Purchase
frac > 0	3.36***	2.06***	0.63***	0.65***
	(0.36)	(0.25)	(0.13)	(0.25)
Quarter X MSA FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Adj. R^2	0.98	0.93	0.94	0.79
N	87219	87219	87219	87219
Date Range	92-17m4	92-17m4	92-17m4	92-17m4

Standard errors two-way clustered by MSA and month. *=10%, **=5%, ***=1% significance. Prepayment is measured using loan level data from McDash Performance data. After 2005, we decompose prepayment by type using CRISM data which links new and old loans. Prior to 2005, we decompose prepayment by type using origination shares by type from CoreLogic LLMA data. See Appendix for additional discussion. Prepayment fractions are measured in month $t+1$ while rate incentives are measured in month t, since McDash data measures origination not application and there is a 1-2 month lag from application to origination.

Table 6: Auto Sales Growth Responses to Refinancing and Mortgage Rate Changes

	(1) OLS	(2) OLS	(3) IV	(4) IV	(5) OLS	(6) OLS	(7) IV	(8) IV
Δ freq	0.0190***	0.0299***	0.0809***	0.135***				
	(0.00725)	(0.00942)	(0.0277)	(0.0397)				
freq					-0.00130	0.0329***	0.00365	-0.124
					(0.00339)	(0.0111)	(0.0146)	(0.0968)
freq $\times\Delta$ FRM					-0.0343*	-0.0446*	-0.203***	-0.386**
					(0.0191)	(0.0229)	(0.0661)	(0.179)
Quarter $\times$ MSA FE	No	Yes	No	Yes	No	Yes	No	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	85328	84585	85328	84585	85328	84585	85328	84585
Date Range	98-17m4	98-17m4	98-17m4	98-17m4	98-17m4	98-17m4	98-17m4	98-17m4

Standard errors two-way clustered by MSA and month. *=10%, **=5%, ***=1% significance. Prepayment is measured using loan level data from McDash Performance data. Δ freq is the change in prepayment and Δ FRM is the change in the 30-year fixed rate mortgage between month $t-1$ and t. To account for lags between origination and spending, the outcome in all regressions is auto sales growth from R.L. Polk measured between t and $t+1$. IV specifications instrument for freq and Δ freq using frac > 0 and Δ frac >0.

A.1 Model Appendix

A.1.1 Mapping between Short Rates and Mortgage Rates

This appendix is written for the case where the mortgage taken by the household amortizes exponentially at rate α. Note that in the main text, we assume that $\alpha := 0$, so that the mortgage face value F is not a state variable of the household problem.

A.1.1.1 Mortgage Rates

In this section, we explain how to map short term rates into mortgage rates when risk-neutral financial intermediaries lend to households. Those calculations are conceptually the same as those developed in Berger, Milbradt and Tourre (2018). The value of a mortgage, from the financial intermediary standpoint, can be encoded via a matrix P_{a,a^*}, where the first subscript represents the current aggregate state and the second subscript represents the aggregate state at which the household refinanced last (in other words the mortgage we are considering pays a coupon rate $m(a^*)$). In other words, the price P_{a,a^*} is defined via:

$$P_{a,a^*} = \mathbb{E}^a \left[\int_0^\tau e^{-\int_0^t (r(a_s)+\alpha)ds} \left(\alpha + m(a^*)\right) dt + e^{-\int_0^\tau (r(a_s)+\alpha)ds} \right]$$

In the above, τ is the prepayment time, a stopping time that is the minimum of (a) an exponentially-distributed time τ_ν representing a move, and (b) the first exponentially distributed attention time τ_χ for which the mortgage rate $m(a_{\tau_\chi})$ is below $m(a^*)$. The matrix P encoding those mortgage prices then verifies the following risk-neutral pricing equation:

$$\left(r(a) + \alpha + \nu + \chi 1_{\{a<a^*\}}\right) P_{a,a^*} = m(a^*) + \alpha + \nu + \chi 1_{\{a<a^*\}} + \sum_{a'=1}^{n} \lambda_{a,a'} P_{a',a^*}$$

Assuming the $n_a \times 1$ vector of mortgage rates $(m(1), ..., m(n_a))$ is known, the above is a set of n_a^2 linear equations in n_a^2 unknown P_{a,a^*}, which can be solved easily. If we note $P_{\cdot,a^*}$ the $n_a \times 1$ column vector with a^{th} entry P_{a,a^*}, if $\text{diag}(r)$ is a diagonal matrix with a^{th} diagonal entry $r(a)$, if I is the $n_a \times n_a$ identity matrix, if 1_{n_a} is an $n_a \times 1$ vector of ones, and if X_{a^*} is an $n_a \times n_a$ diagonal matrix with a^{th} diagonal entry $\chi 1_{\{a<a^*\}}$, $P_{\cdot,a^*}$ can be computed via:

$$P_{\cdot,a^*} = \left(\text{diag}(r) + (\alpha+\nu)I + X_{a^*} - \Lambda\right)^{-1} \left((m(a^*) + \alpha + \nu)I + X_{a^*}\right) 1_{n_a}$$

In a risk-neutral environment, it must be the case that the price of the mortgage, at time of origination, is equal to the notional of such mortgage. In other words, we must have the mortgage market equilibrium condition $P_{a,a} = 1$. Note $\Sigma_{a^*} := \text{diag}(r) + (\alpha+\nu)I + X_{a^*} - \Lambda$, then the mortgage rate can be computed in semi-closed form:

$$m(a^*) = \frac{1 - \left[\Sigma_{a^*}^{-1} \left((\alpha+\nu)I + X_{a^*}\right) 1_{n_a}\right]_{a^*}}{\left[\Sigma_{a^*}^{-1} 1_{n_a}\right]_{a^*}}$$

A.1.2 Solution to Household Problem

We note $V_{a,a^*,s}(W)$ the value function in aggregate state a for a household in idiosyncratic state s, with liquid wealth W and mortgage debt F paying a fixed mortgage rate $m(a^*)$. The household Hamilton-Jacobi-Bellman (HJB) equation can be written:

$$\delta V_{a,a^*,s} = \sup_C u(C) + \sum_{a'=1}^{n_a} \lambda_{a,a'} V_{a',a^*,s} + \sum_{s'=1}^{n_s} \theta^a_{s,s'} V_{a,a^*,s'} + \chi \left[V_{a,\min(a,a^*),s} - V_{a,a^*,s} \right] + \nu \left[V_{a,a,s} - V_{a,a^*,s} \right] + \left(r(a) W + Y(s) - C - (m(a^*) + \alpha) F \right) V'_{a,a^*,s} \quad \text{(A1)}$$

The optimal consumption function $C_{a,a^*,s}(W)$ solves the first order condition $u'\left(C_{a,a^*,s}(W)\right) = V'_{a,a^*,s}(W)$, which can be written:

$$\delta C^{-\gamma}_{a,a^*,s} = V'_{a,a^*,s}$$

We can then reinject the optimal consumption policy into the HJB equation satisfied by $V_{a,a^*,s}$ to obtain:

$$0 = u\left(C_{a,a^*,s}\right) - \delta V_{a,a^*,s} + \sum_{a'=1}^{n_a} \lambda_{a,a'} V_{a',a^*,s} + \sum_{s'=1}^{n_s} \theta^a_{s,s'} V_{a,a^*,s'} + \chi \left[V_{a,\min(a,a^*),s} - V_{a,a^*,s} \right] + \nu \left[V_{a,a,s} - V_{a,a^*,s} \right] + \left(r(a) W + Y(s) - C_{a,a^*,s} - (m(a^*) + \alpha) F \right) V'_{a,a^*,s} \quad \text{(A2)}$$

The above equation is a first-order non-linear ordinary differential equation (for the continuous state variable W). The non-linearity stems from the fact that consumption is controlled – its value depend on $V_{a,a^*,s}$ and its first partial derivative w.r.t. W. The boundary conditions are worth discussing. Since we restrict $W_t \geq 0$, we need to impose a weakly positive savings rate at $W = 0$. This means that we must have:

$$\begin{aligned} & C_{a,a^*,s}(0) \geq Y(s) - (m(a^*) + \alpha) F \\ \Leftrightarrow\ & V'_{a,a^*,s}(0) \leq \delta \left(Y(s) - (m(a^*) + \alpha) F \right)^{-\gamma} \end{aligned}$$

We can also look at the asymptotic behavior of the consumption, savings and value functions as $W \to +\infty$. Postulate that in such case, the value function takes the form:

$$V_{a,a^*,s}(W) \underset{W \to +\infty}{=} v_a \frac{W^{1-\gamma}}{1-\gamma} \quad \text{(A3)}$$

Using the first order condition for consumption leads to:

$$C_{a,a^*,s}(W) \underset{W \to +\infty}{=} \delta^{1/\gamma} v_a^{-\frac{1}{\gamma}} W$$

Reinjecting our guess in (A3) into the HJB equation satisfied by the value function leads to the following set of n_a non-linear equations for the unknown constants v_a:

$$\frac{\gamma \delta^{1/\gamma}}{1-\gamma} v_a^{-\frac{1}{\gamma}} - \frac{\delta}{1-\gamma} + \sum_{a'=1}^{n_a} \frac{\lambda_{a,a'}}{1-\gamma} \frac{v_{a'}}{v_a} + r(a) = 0$$

In the case of a unique aggregate interest rate state, we obtain immediately that the marginal propensity to consume of very wealthy households is equal to:

$$\delta^{1/\gamma} v_a^{-\frac{1}{\gamma}} = \left(\left(\frac{1}{\gamma} \right) \delta + \left(1 - \frac{1}{\gamma} \right) r \right)$$

In other words, the marginal propensity to consume out of financial wealth for very wealthy households is equal to a weighted average of the household's rate of time preference δ and the short term rate r (weighted by the IES and 1 minus IES respectively). In the case where the household has "log" time-

separable preferences, the asymptotic behavior of the value function is the following:

$$V_{a,a^*,s}(W) \underset{W\to+\infty}{=} v_a + \ln W$$

In this particular case, the set of constants $\{v_a\}$ satisfy a set of linear equations, while the marginal propensity to consume is equal to δ (this is the so-called "myopic" behavior).

A.1.3 Fokker Planck Equation

The aggregate state a_t evolves entirely exogenously in our model. For a given known realization of the aggregate state $\{a_t\}_{t\geq 0}$, the joint density g_t over (a) savings W, (b) coupons $m(a^*)$ and (c) income $Y(s)$ satisfies the following Fokker Planck equation:

$$\begin{aligned}\partial_t\left[g_t\left(W,a^*,s\right)\right] = &-\partial_W\left[\mu_W\left(W,a_t,a^*,s\right)g_t\left(W,a^*,s\right)\right]\\ &+\sum_{s'=1}^{n_s}\theta_{s',s}^{a_t}g_t\left(W,a^*,s'\right)-\left(\nu+\chi 1_{\{m(a_t)<m(a^*)\}}\right)1_{\{a_t\neq a^*\}}g_t\left(W,a^*,s\right)\\ &+1_{\{a_t=a^*\}}\left[\nu\sum_{a'\neq a_t}g_t\left(W,a',s\right)+\chi\sum_{a':m(a')>m(a_t)}g_t\left(W,a',s\right)\right]\end{aligned}$$

This equation will be leveraged in our numerical scheme when computing impulse response functions.

A.1.4 Numerical Implementation

We describe our numerical implementation for the case $\alpha = 0$ – i.e. the case where the mortgage is a console instrument. The more general case $\alpha > 0$, which adds an additional state variable, can be computed using a similar logic. We will compute the equilibrium of the model numerically by determining the value function $\{V_{a,a^*,s}\}_{1\leq a,a^*\leq n_a,1\leq s\leq n_s}$ at n_w discrete points $\{W_k\}_{1\leq k\leq n_w}$ of the state space. We create an equally spaced grid with steps Δ_w. We note $V_{a,a^*,s,k} = V_{a,a^*,s}(W_k)$. We define the forward and backward difference approximations of $V'_{a,a^*,s}$ as follows:

$$V'_{a,a^*,s,k,B} := \frac{V_{a,a^*,s,k}-V_{a,a^*,s,k-1}}{\Delta_w} \qquad V'_{a,a^*,s,k,F} := \frac{V_{a,a^*,s,k+1}-V_{a,a^*,s,k}}{\Delta_w}$$

The HJB satisfied by $V_{a,a^*,s}$ can be written:

$$\begin{aligned}0 = u\left(\left(u'\right)^{-1}\left(V_{a,a^*,s,k}\right)\right)-\delta V_{a,a^*,s,k}+\sum_{a'=1}^{n_a}\lambda_{a,a'}V_{a',a^*,s,k}+\sum_{s'=1}^{n_s}\theta_{s,s'}^{a}V_{a,a^*,s',k}\\ +\chi\left[V_{a,\min(a,a^*),s,k}-V_{a,a^*,s,k}\right]+\nu\left[V_{a,a,s,k}-V_{a,a^*,s,k}\right]+A_{a,a^*,s,k}V'_{a,a^*,s,k}\end{aligned}$$

In the above, we have used $A_{a,a^*,s,k} := \left(r(a)W + Y(s) - C_{a,a^*,s}(W_k) - m\left(a^*\right)F\right)$. The finite difference approximation of $V'_{a,a^*,s,k}$ will be:

$$V'_{a,a^*,s,k} := V'_{a,a^*,s,k,B}1_{\{A_{a,a^*,s,k}<0\}} + V'_{a,a^*,s,k,F}1_{\{A_{a,a^*,s,k}\geq 0\}}$$

This upwinding strategy insures that our finite difference scheme is monotone, in the sense of Barles and Souganidis (1991). We compute our value functions iteratively using a false transient (aka an artifical time-derivative) as follows:

$$\frac{V^{t+\Delta_t}_{a,a^*,s,k}-V^{t}_{a,a^*,s,k}}{\Delta_t} = u\left(\left(u'\right)^{-1}\left(V^t_{a,a^*,s,k}\right)\right)-\delta V^{t+\Delta_t}_{a,a^*,s,k}+\sum_{a'=1}^{n_a}\lambda_{a,a'}V^{t+\Delta_t}_{a',a^*,s,k}+\sum_{s'=1}^{n_s}\theta_{s,s'}^{a}V^{t+\Delta_t}_{a,a^*,s',k}$$

$$+\chi\left[V^{t+\Delta_t}_{a,\min(a,a^*),s,k}-V^{t+\Delta_t}_{a,a^*,s,k}\right]+\nu\left[V^{t+\Delta_t}_{a,a,s,k}-V^{t+\Delta_t}_{a,a^*,s,k}\right]$$
$$+\left[A_{a,a^*,s,k}\right]^+\left(V'_{a,a^*,s,k,F}\right)^{t+\Delta_t}+\left[A_{a,a^*,s,k}\right]^-\left(V'_{a,a^*,s,k,B}\right)^{t+\Delta_t}$$

Here, Δ_t is the time-step we impose. Usually, we pick $\Delta_t = 1$. The above is a system of $n_s \times n_a^2 \times n_w$ linear equations that can be written:

$$\left[(1+\delta\Delta_t)I-\Delta_t M_t\right]V^{t+1}=V^t+\Delta_t\Phi_t$$

where M_t is a square matrix, Φ_t is a vector with elements $\{u\left(C^t_{a,a^*,s,k}\right)\}$, and V^{t+1} is our unknown value vector. We organize the linear system such that element $V_{a,a^*,s,k}$ can be found in row $k+(a-1)n_w+(a^*-1)n_an_w+(s-1)n_a^2n_w$. Note then that the diagonal elements of the matrix M_t are equal to:

$$-\left[\frac{|A_{a,a^*,s,k}|}{\Delta_w}-\lambda_{a,a}-\theta^a_{s,s}+\chi 1_{\{m(a)<m(a^*)\}}+\nu\right]$$

Lower and upper diagonals (i.e. one diagonal off the main diagonal) can then be filled with $-\left[A_{a,a^*,s,k}\right]^-/\Delta_w$ (for the lower diagonal) and $\left[A_{a,a^*,s,k}\right]^+/\Delta_w$ (for the upper diagonal). The rest of the matrix can be filled with block diagonal elements, corresponding to aggregate state transitions ($\lambda_{a,a'}$), idiosyncratic state transitions ($\theta^a_{s,s'}$), and refinancings ($\nu+\chi 1_{\{m(a)<m(a^*)\}}$).

Our algorithm iterates over time, until the point where the artificial time derivative is close to zero. Note that the matrix M_t then converges to a matrix M. Both M_t and M are Markov intensity matrices – their row sums are always equal to zero. Finally, note that if we were instead implementing an explicit – as opposed to implicit – finite difference scheme, the equation that would need to be solved at each point of the state space and for each time period is:

$$V^{t+1}=\left[(1-\delta\Delta_t)I+\Delta_t M_t\right]V^t+\Delta_t\Phi_t$$

The ergodic distribution of our economic model is then computed by focusing on the implied transition intensity matrix M, and by finding the column vector π that solves $\pi' M = 0$ — in other words, the left-eigen-vector of M, associated with the eigen-value 0, that verifies $\sum_{i=1}^N \pi_i = 1$. Numerically, finding an eigen-vector related to a particular eigen-value of a large matrix is complex. Instead, we proceed by solving a linear system as follows:

$$\begin{pmatrix}\pi_1 & \pi'_2\end{pmatrix}\begin{pmatrix}m_{11} & m_{12}\\ m_{21} & M_{22}\end{pmatrix}=0$$

In the above, m_{11} is a scalar, m_{21} has dimension $(N-1)\times 1$, m_{12} has dimension $1\times(N-1)$, and M_{22} is an $(N-1)\times(N-1)$ matrix. Up to a renormalization, we can assume that $\pi_1 = 1$, and we then simply solve the linear system $m_{12}+\pi'_2 M_{22}=0$, which has a solution $\pi_2 = -(M'_{22})^{-1}m'_{12}$.

Finally, in order to compute consumption impulse response functions, we use Monte-Carlo simulations to generate 1,000 random realizations of the aggregate state $\{a_t\}_{t\geq 0}$. For each realization, we compute the time-series evolution of the density $g_t(W,a^*,s)$ by (a) constructing the $n_a\times n_s\times n_w$ matrix B_{a_t}, which corresponds to the discretized version (using a monotone scheme) of the Feynman-Kac operator (A4) (defined below for any function f), and (b) applying the recursive equation $G_{t+dt}=\left(I+B^T_{a_t}dt\right)G_t$, where $B^T_{a_t}$ is the transpose of B_{a_t}.

$$B_{a_t}f_{a^*,s}(W):=\sum_{s'=1}^{n_s}\theta^{a_t}_{s,s'}f_{a^*,s'}+\chi\left[f_{\min(a_t,a^*),s}-f_{a^*,s}\right]+\nu\left[f_{a_t,s}-f_{a^*,s}\right]$$
$$+\left(r(a_t)W+Y(s)-C_{a_t,a^*,s}-m(a^*)F\right)f'_{a^*,s}\quad\text{(A4)}$$

The underlying rational for this strategy relies on the fact that the Fokker-Planck operator is the adjoint of the Feynman-Kac operator, and in a discrete state setting, the adjoint operation merely corresponds to taking the transpose of the Feynman-Kac matrix.

A.1.5 Continuous and Discrete State Markov Processes

Several methods have been implemented over the years to approximate continuous state Markov processes by discrete state Markov chains. Since we use a small number of discrete states in our numerical application, we use a procedure which matches conditional and unconditional moments of the original process and the approximating process. We use the model of Cox, Ingersoll Jr and Ross (2005) for the short term interest rate in order to illustrate our procedure:

$$dr_t = -\kappa\left(r_t - \mu\right)dt + \sigma\sqrt{r_t}dZ_t$$

In order to approximate this process by a discrete state continuous time Markov process, we first choose a discrete number of risk-free rates $\{r_a\}_{1\leq a\leq n_a}$, and then minimize the distance between (i) a set of conditional and unconditional moments of the continuous state process and (ii) the same set of conditional and unconditional moments of our discrete state continuous time Markov process. Table A-1 highlights the set of conditional and unconditional moments we use. In such table, we note π the vector encoding the ergodic distribution of the discrete state Markov process; π satisfies $\pi'\Lambda = 0$, and $\pi' 1_{n_a} = 1$.

Table A-1: Target Moments

Moment	Cox, Ingersoll Jr and Ross (2005)	Discrete State Model
$\mathbb{E}\left[r_t\right]$	μ	$\sum_{i=1}^{n_a}\pi_i r_i$
$\text{var}\left[r_t^2\right]$	$\frac{\mu\sigma^2}{2\kappa}$	$\sum_{i=1}^{n_a}\pi_i r_i^2 - \left(\sum_{i=1}^{n_a}\pi_i r_i\right)^2$
$\mathbb{E}\left[r_{t+s}\middle\vert\mathcal{F}_t\right]$	$r_t e^{-\kappa s} + \mu\left(1-e^{-\kappa s}\right)$	$\sum_{j=1}^{n_a}\left(e^{\Lambda s}\right)_{s_t,j} r_j$
$\text{var}\left[r_{t+s}\middle\vert\mathcal{F}_t\right]$	$\frac{\mu\sigma^2}{2\kappa}\left(1-e^{-\kappa s}\right)^2 + \frac{r_t\sigma^2}{\kappa}\left(e^{-\kappa s}-e^{-2\kappa s}\right)$	$\sum_{j=1}^{n_a}\left(e^{\Lambda s}\right)_{s_t,j} r_j^2 - \left(\sum_{j=1}^{n_a}\left(e^{\Lambda s}\right)_{s_t,j} r_j\right)^2$

Our choice of n_a is driven by computational considerations – we pick $n_a = 7$ in order for our numerical algorithm to be able to solve for an equilibrium relatively quickly. Since we impose that the matrix Λ is tri-diagonal, this means that we need to compute $2(n_a - 1) = 14$ transition intensities $\left(\Lambda_{ij}\right)_{i\neq j}$. We pick an equally spaced grid such that $r_1 = 0.01$ and $r_{n_a} = 0.07$. In addition to the mean and variance of the ergodic distribution of r_t, we also match the conditional mean and variance at 2 horizons: 0.25 years, and 2 years. This gives us 16 different moments (14 conditional, and 2 unconditional), for 14 free parameters. The resulting ergodic distribution of our discrete state process has a mean of 4% and a standard deviation of 2.3%, which correspond exactly to the moments of the ergodic distribution of the original continuous state process.

$$\Lambda = \begin{bmatrix} -0.57 & 0.57 & 0 & 0 & 0 & 0 & 0 \\ 1.36 & -2.39 & 1.03 & 0 & 0 & 0 & 0 \\ 0 & 0.87 & -2.13 & 1.25 & 0 & 0 & 0 \\ 0 & 0 & 1.36 & -2.49 & 1.13 & 0 & 0 \\ 0 & 0 & 0 & 1.38 & -2.78 & 1.40 & 0 \\ 0 & 0 & 0 & 0 & 0.95 & -1.82 & 0.86 \\ 0 & 0 & 0 & 0 & 0 & 0.48 & -0.48 \end{bmatrix}$$

Note that the autocorrelation function of the discrete state continuous time Markov process approximation takes the following form:

$$R(s) = \frac{\sum_{i=1}^{n_a} \pi_i r_i \sum_{j=1}^{n_a} \left[\left(e^{s\Lambda} \right)_{ij} - \pi_j \right] r_j}{\sum_{i=1}^{n_a} \pi_i r_i^2 - \left(\sum_{i=1}^{n_a} \pi_i r_i \right)^2}$$

One can show that asymptotically, $R(s) \approx e^{-\kappa s}$, where $-\kappa$ is the largest non-zero (and of course negative) eigen-value of the matrix Λ. Our numerical procedure leads to $\kappa = 0.19$, which is very close to the persistence parameter of our continuous state process.

A.2 Data Appendix

A.2.1 Baseline McDash Prepayment Sample

This section describes in more detail our loan-level sample restrictions as well as our identification of prepayment type in CRISM data. Our primary loan-level data set for measuring prepayment and loan gaps is the McDash loan performance and origination data from 1992-2017produced by Black Knight Financial Services. The origination data provides a number of origination characteristics such as origination date, amount, loan purpose and appraisal value while the performance data provides dynamic info on these loans like current unpaid balance, current interest rate and flags for prepayment. Our prepayment analysis primarily requires information from the dynamic loan performance data: we define loan prepayment in month t as any loan with termination flag "voluntary payoff" in that month and a termination date of month t. We restrict our analysis to fixed rate first mortgages, but results are similar when including all mortgages in the McDash data set. In order to maintain a consistent sample when running cross-MSA results, we also drop any loan with missing information on MSA-division. We also drop any loan with missing information on the current interest rate in the McDash loan performance data set, since we cannot measure gaps for these loans. We define the interest rate gap as the current interest rate minus the monthly average 30 year FRM from the Freddie Mac PMMS, and we bin loans in 20 basis point bins by interest rate gaps, from $< -5\%$ to $> +5\%$. We also compute \$ gaps in addition to rate gaps, which we define as the current unpaid balance times the interest rate gap. For that analysis, we drop any loan with missing unpaid balance in the McDash performance data set.

Since most of our analysis is focused on prepayment rates, the majority of our analysis can be performed using only loan performance data. However, while performance data is required to measure prepayment, it cannot be used to decompose prepayment into that arising from refinancing and moves. This is because loan purpose is collected at origination but not at termination, so the performance data set tells us if a loan prepays but not why. Conversely, origination data can be used to measure the share of *new* originations which are due to refinancing and moves, but it cannot be used to measure the share of *old* mortgages which are prepaying. So origination data cannot tell us what share of mortgages prepay and why, since it contains the wrong denominator. This means that measuring the frequency of prepayment by type requires linking information on loan performance for terminating loans with loan origination information for newly originating loans. After 2005, we are able to use the linked Equifax/CRISM data which we describe in the next subsection to precisely link each individual prepaying loan to a newly originating loan so that we can measure exactly why each individual loan is prepaying. Prior to 2005, these links are unavailable, so we cannot measure the reason that any individual loan prepays However, in a stationary environment, performance data and origination data can be combined to proxy for the *share* of prepayment arising from different types. In particular, in an environment with no net flows in and out of the mortgage market, every loan which prepays due to refinancing or due to moving must be matched by a newly originated loan with the same purpose. While we cannot link the new and old loan together, the shares must remain unchanged. This allows

us to proxy for rate-refi, cash-out refi and movement frequencies in a month using only loan level data without links to individuals.

In particular, we compute $freq_t^{type} = freq_t^{prepay} \times share_t^{type}$ where $freq_t^{type}$ is the frequency of a given type of prepayment, $freq_t^{prepay}$ is the frequency of prepayment in performance data and $share_t^{type}$ is the share of a given type of loan purpose in origination data. The McDash origination data set only collects loan purpose after 1998, so we measure $share_t^{type}$ using originations data from CoreLogic LLMA. This data set has a structure nearly identical to the McDash data, but it contains reliable loan purpose at origination info as early as 1993. However, we continue to measure $freq_t^{prepay}$ using McDash data because CoreLogic performance data does not measure prepayment before 1999, and has roughly half the market coverage of the McDash Performance Data set. Combining prepayment frequencies from McDash with loan purposes shares at origination from CoreLogic thus leverages the comparative advantages of the two data sets. While stationarity is clearly a strong assumption, after 2005, we can use the CRISM data to compare our proxies for frequency by type under the stationarity assumption with actual frequencies. Figure A-1 shows they are very similar.

Overall, the McDash Performance data set contains information on approximately 180 million loans. After 2005, the McDash Performance data set covers roughly 50% of total U.S. mortgage debt as measured by the Federal Reserve. Prior to 2005, coverage is somewhat lower, ranging from around 10% market coverage in the early 90s to 20-25% in the late 90s. As a measure of representativeness and external validity, Appendix Figure A-2 shows that refinancing in our data closely tracks the refinancing applications index produced by the Mortgage Banker's Association from 1992-2017.[63] This suggests that despite the changing sample sizes, the McDash Performance data is broadly representative of the U.S. mortgage market over the entire 1992-2017 period.[64]

We use the McDash Performance data rather than CoreLogic Performance data because the CoreLogic performance data does not measure prepayment before 1999, and has roughly half the market coverage of the McDash performance data. As shown in Appendix Figure A-3, using CoreLogic instead of McDash Performance data leads to somewhat noiser refinancing series and eliminates the first 7 years of data. In addition, we cannot link loans to individuals in CoreLogic data like we can in the McDash data using the links we describe next. Nevertheless, we have repeated our analysis of prepayment using series derived from CoreLogic LLMA Performance data and arrive at similar conclusions.

A.2.2 Linked CRISM Sample: Measuring Refinancing

After 2005, we link loans in the McDash data set to Equifax credit records, which allows us to decompose prepayment into different types and to control for several individual level observables. Our analysis and description of this data closely follows Beraja, Fuster, Hurst and Vavra (2018). The linked Equifax/McDash CRISM data set provides the linked Equifax credit records for each McDash mortgage for the lifetime of the loan, including an additional 6 months before origination and after termination. This link is done directly by Equifax. Credit records provide a consumer's total outstanding debt amounts in different categories (first-lien mortgages, second-lien mortgages, home equity lines of credit [HELOCs], auto loans, etc.). Additionally, in any month, Equifax provides the origination date, amount, and remaining principal balance of the two largest (in balance terms) first mortgages, closed-end seconds, and HELOCs outstanding for a given consumer.

In order to reduce the computational burden, we begin the analysis of CRISM data by extracting all the loan and individual characteristics from a random 10% sample of all individuals in the Equifax/McDash data at some point between 1992 and 2017. This 10% CRISM sample includes all

[63] Note that we measure originations while this index measures applications. According to LendingTree, denials are roughly 8% after the financial crisis due to Dodd-Frank related changes in lending standards. This explains the level difference after the Financial Crisis but the series continue to highly comove.

[64] Note that even in the months with the fewest observations, we still have more than 5 million mortgages, so only lack of representativeness and not sampling error is a potential concern.

mortgage loans (approximately 11 million) for approximately 5.9 million individuals[65] We further restrict our CRISM sample to those consumers who start our sample with two or fewer loans in each category and never have more than three of any of these types of loans outstanding.[66] These sample restrictions leave roughly 96% of the 5.9 million individuals in our analysis sample. In creating this loan-level data set, we assume that the month in which the loan stops appearing in Equifax is the month that it was terminated.

While McDash loans are linked to individual credit records directly be Equifax using social security number, individuals can have multiple loans and the loan information in Equifax does not always exactly match that in McDash since they come from independent sources.[67] We thus have to construct a unique match between a loan in McDash and the possible set of linked loans in Equifax. As in Beraja, Fuster, Hurst and Vavra (2018), we consider an Equifax loan/McDash loan pairing a match if the origination date of the Equifax loan is within 1 month and the origination amount is within $10,000 of the McDash loan. If more than one loan is matched, we use the origination amount, date, termination date, zip code (where available, or 3-digit zip code and MSA-div where not)[68], and termination balance as tiebreakers. We are able to match roughly 93% of McDash loans to an Equifax loan using these restrictions.

As in our primary analysis, we begin with all remaining outstanding fixed rate first liens in the McDash which are voluntarily paid off. We then look for any loan in the Equifax data set that has an open date within 4 months of the McDash loan's termination date. We classify these new loans as a refinance if either:

- The loan also appears in McDash and is tagged as a refinance in the purpose-type variable.
- The loan also appears in McDash and is tagged as an "Unknown" or "Other" purpose type, and has the same property 5 digit zip code (where available, or 3-digit zip code and MSA-div where not) as the original loan.
- The loan appears only in Equifax but the borrower's Equifax address does not change in the 6 months following the termination of the original loan.

This allows us to compute one of our primary outcomes of interest, the count of first-lien FRM loans which refinance in month t divided by the total number of McDash first-lien FRM loans with Performance data in that month. (We have also considered results which compute balance weighted shares, and they are very similar).

A.2.3 Linked CRISM Sample: Decomposing refinancing into rate and cash-out

To compute the cash-out and rate-refinancing share of loans, we must further break these refinancing loans down by type. In particular, we need to compute how the balance of the new loan compares to the outstanding balance of the loan(s) being prepaid. We begin by labeling any loan in the Equifax data set that terminates between -1 and 4 months from a new McDash loan's close date a "linked" loan, including first mortgages as well as closed-end seconds and HELOCs, and we call the new loan a refinance if:

- The loan is a known refinance in McDash.

[65] Results are extremely similar when using 5% and 20% samples since the CRISM sample is very large so sampling error is not important.

[66] This restriction allows us to infer the origination month, origination balance, and balance of the third largest loan of any loan type even though this information does not appear explicitly in Equifax, where if the third largest loan is also the newest loan, we assume its origination month to be the first month it appears in Equifax. We also drop loans that do not have complete consecutive Equifax records.

[67] For example, balances may differ slightly since they may be reported to credit bureaus and servicers at different dates.

[68] To ensure anonymity, McDash reports 5-digit zip code for loans in higher volume locations and 3-digit zip code for loans in lower volume locations.

- The loan has an "Unknown" or "Other" purpose type in McDash and a linked loan in McDash that has a matching property zip code (5 digit when available or 3-digit + MSA-div when not).
- The loan has an "Unknown" or "Other" purpose type in McDash and a linked loan that appears only in Equifax, but the consumer's Equifax address does not change in the 6 months after the new loan was opened.

If there is more than one linked loan that is a first mortgage in Equifax, we link only the loan that is closest in balance to the origination amount of the new mortgage. We only link those Equifax loans that exist in the Equifax data for at least three months to prevent the refinanced loan balance from being counted in the old balance of the loan.

For each of these cases, we can then calculate the cash-out amount as the difference between the origination amount on the refinance loan and the balance of the linked loan(s) at termination. In order to capture the correct origination amount on the refinance loan, we want to ensure that we are also including any "piggyback" second liens that are opened with the refinance loan that we find in McDash. Thus, we look for any loan in the Equifax record linked to our refinance loan that has an Equifax open date within three months of our refinance loan and an origination balance of less than 25% of our loan's origination balance if labeled a first mortgage and less than 125% of the refinance loan's origination balance if labeled a HELOC or CES, and add the balance of these piggyback seconds to the refi origination amount when calculating cash-out amounts.[69] To eliminate outliers, we also drop cash-out and "cash-in" amounts that are greater than $1,000,000. These amount to dropping less than 0.05% of the refinance loans.

After measuring the change in the balance, we then call a refinancing a cash-out if, after subtracting 2 percent from the new loan to cover closing costs, the new mortgage balance is at least $5,000 above the old mortgage. Using a more restrictive definition of cash-out reduces the overall share of cash-out and the sensitivity of cash-out to rate gaps while using a less conservative cutoff does the reverse since it reclassifies some rate-refis as cash-outs. These definition by construction has no effect on the decomposition of prepayment into refi vs. moves.

A.2.4 Leverage controls

In many of our aggregate regressions we control for average leverage in our data set and in our individual level regressions, we control for individual leverage. In order to measure leverage at the loan-level we start with all McDash FRM first mortgages. For each mortgage we estimate its current value as the appraisal value at origination updated using local house price indices from CoreLogic. We use zip code level house price indices to update values when the 5-digit zip is available in McDash and in the CoreLogic indices, and we otherwise use MSA level house price indices. We then compute leverage LTV for a given loan as the ratio of the current unpaid balance to this estimate of value. Our aggregate controls then take the average leverage across loans.

This procedure will tend to understate leverage for individuals with multiple loans, but it can be applied over the entire 1992-2017 sample. After 2005, we can construct a more accurate measure of leverage using CRISM data. Following Beraja, Fuster, Hurst and Vavra (2018), we begin with first-lien McDash FRM loans. For each month, we then take the corresponding Equifax record and assign all outstanding second liens to the outstanding first liens in Equifax using the rule that each second lien is assigned to the largest first lien (in balance terms) that was opened on or before the second lien's opening date. We then add the assigned second lien balance(s) to the McDash balance of our original loan as our measure of secured debt on a property, which is the numerator of $CLTV$. We then divide by the value constructed exactly as described above.

[69] We impose these upper bounds because we want to avoid picking up other first lien mortgages (to purchase another property) the borrower might originate at the same time.

A.3 Additional Figures and Tables

Figure A-1: Construction of Prepay Shares by Type

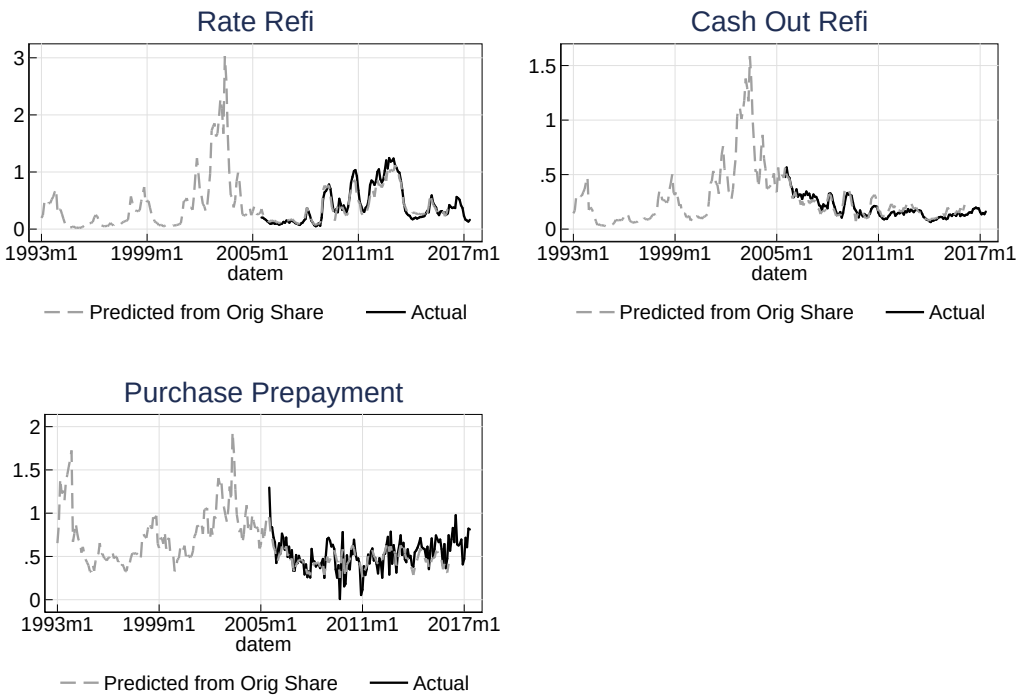


Figure A-2: Comparison of Refi Measured with McDash Data to Mortgage Bankers' Association data

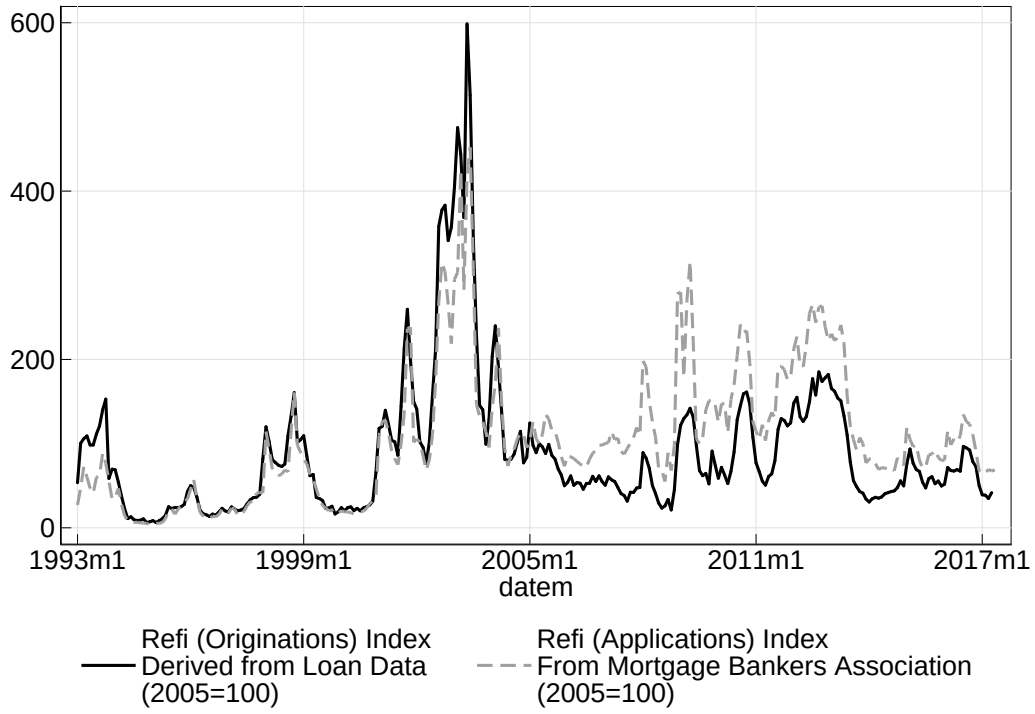


Figure shows an index of refinancing computed using McDash Performance data and CoreLogic Origination Purpose data compared to the Mortgage Banker's Association Refinancing Application Index. Note that the loan-level index measures originations while the MBA index measures applications. Indices are normalized to 100 in 2005m4.

Figure A-3: Comparison of Refi Measured with CoreLogic Performance Data to Mortgage Bankers' Association data

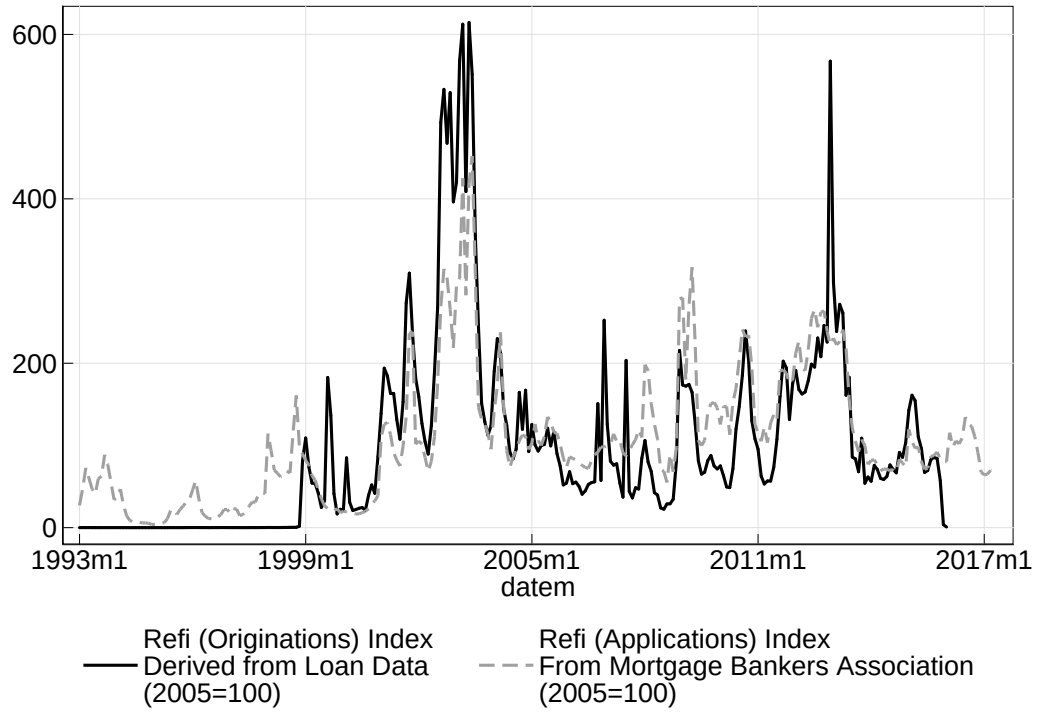


Figure shows an index of refinancing computed using alternative CoreLogic Performance data and CoreLogic Origination Purpose data compared to the Mortgage Banker's Association Refinancing Application Index. Note that the loan-level index measures originations while the MBA index measures applications. Indices are normalized to 100 in 2005m4. We do not use this series in any of our reported results since it is noiser and available for a shorter time-window than our baseline in A-2, but redoing results with this series leads to similar conclusions.

Figure A-4: Distribution of Gaps at Two Dates

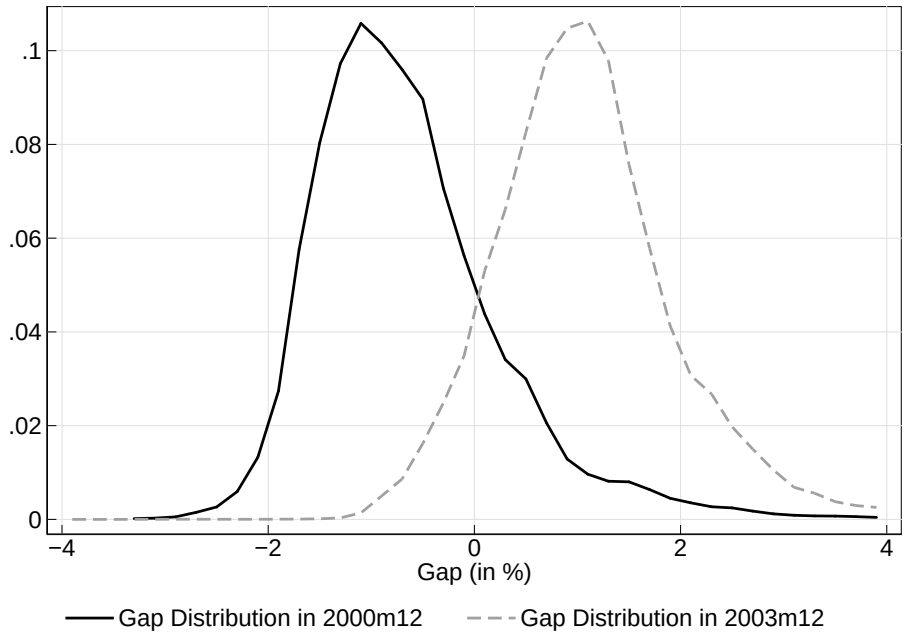


Table A-2: Robustness to Time Fixed Effects: Effects of Rate Gaps on Prepayment Propensities

	(1)	(2)	(3)	(4)	(5)	(6)
frac > 0	2.012***			1.682***	2.355***	1.478***
	(0.381)			(0.594)	(0.578)	(0.225)
frac > 50bp		2.244***				
		(0.435)				
frac > 100bp			2.929***			
			(0.513)			
Constant	0.102	0.170	0.325	0.351	0.118	-0.318*
	(0.292)	(0.286)	(0.241)	(0.452)	(0.420)	(0.180)
Time(Quarter) FE	YES	YES	YES	YES	YES	YES
Adj. R^2	0.919	0.927	0.928	0.826	0.940	0.927
N	304	304	304	108	120	76
Date Range	92-17m4	92-17m4	92-17m4	92-00	01-10	11-17m4

This table repeats results in Table 1 but including calendar quarter time fixed effects so that identification only occurs off of within quarter time-variation. We do not include Columns (4) and (5) from that table since the 2003 dummy is absorbed by quarter fixed effects and CLTV varies little within quarter. Prepayment fractions are measured in month $t+1$ while rate incentives are measured in month t, since McDash data measures origination not application and there is a 1-2 month lag from application to origination. Newey-West standard errors in parantheses. *=10%, **=5%, ***=1% significance.

Table A-3: Effects of $ Gaps on Prepayment Propensities

	(1)	(2)	(3)	(4)	(5)	(6)
frac > $0	2.029***					
	(0.379)					
frac > $200		1.914***				
		(0.383)				
frac > $400			1.831***			
			(0.398)			
frac > $600				1.765***		
				(0.413)		
frac > $1000					1.643***	
					(0.435)	
frac > $1500						1.537***
						(0.440)
Constant	-0.0162	0.245	0.457***	0.627***	0.882***	1.047***
	(0.194)	(0.157)	(0.131)	(0.111)	(0.0918)	(0.0937)
Adj. R^2	0.279	0.273	0.246	0.212	0.140	0.0829
N	304	304	304	304	304	304
Date Range	92-17m4	92-17m4	92-17m4	92-17m4	92-17m4	92-17m4

Newey-West standard errors in parantheses. *=10%, **=5%, ***=1% significance. Loan level data from McDash Performance data. Annual $ Gaps are defined as the current outstanding balance times the rate gap. Prepayment fractions are measured in month $t+1$ while rate incentives and CLTV are measured in month t, since McDash data measures origination not application and there is a 1-2 month lag from application to origination.

Table A-4: Robustness to Including All First Mortgages: Effects of Rate Gaps on Prepayment Propensities

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
frac > 0	1.914***			8.695**	5.914**	2.007***	2.403***	1.899***
	(0.399)			(3.891)	(2.953)	(0.300)	(0.911)	(0.386)
frac > 50bp		1.837***						
		(0.408)						
frac > 100bp			2.051***					
			(0.478)					
mean CLTV				-0.847	-2.016			
				(2.546)	(2.110)			
(frac > 0) × mean CLTV				-9.506*	-5.528			
				(5.644)	(4.328)			
$\mathbb{1}_{year=2003}$					1.638***			
					(0.582)			
Constant	0.218	0.516***	0.796***	0.268	1.098	0.0694	0.267	-0.197
	(0.197)	(0.150)	(0.112)	(1.711)	(1.392)	(0.158)	(0.413)	(0.273)
Adj. R^2	0.235	0.222	0.187	0.491	0.624	0.571	0.225	0.561
N	304	304	304	304	304	108	120	76
Date Range	92-17m4	92-17m4	92-17m4	92-17m4	92-17m4	92-00	01-10	11-17m4

This table replicates our baseline time-series analysis in Table A-4, but including all first-liens rather than restricting the analysis to fixed rate mortgages. Newey-West standard errors in parantheses. *=10%, **=5%, ***=1% significance. Mean LTV is the ratio of a loan's outstanding balance to value estimated using appraisal values at origination updated using local house price indices from CoreLogic. Loan level data from McDash Performance data+appraisal values from McDash origination data is used to calculate LTV. Prepayment fractions are measured in month $t+1$ while rate incentives and LTV are measured in month t, since McDash data measures origination not application and there is a 1-2 month lag from application to origination.

Table A-5: Robustness to Estimating Heterogeneous Rate Targets: Effects of Rate Gaps on Prepayment Propensities

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
frac > 0	1.995***			8.006**	5.290*	2.085***	2.566***	2.009***
	(0.374)			(3.658)	(2.729)	(0.359)	(0.853)	(0.379)
frac > 50bp		1.948***						
		(0.383)						
frac > 100bp			2.208***					
			(0.435)					
mean CLTV				-0.934	-2.031			
				(2.361)	(1.921)			
(frac > 0) × mean CLTV				-8.415	-4.554			
				(5.276)	(3.981)			
$\mathbb{1}_{year=2003}$					1.763***			
					(0.624)			
Constant	0.0848	0.394***	0.692***	0.223	1.016	-0.0545	0.0667	-0.356
	(0.185)	(0.139)	(0.102)	(1.592)	(1.268)	(0.194)	(0.383)	(0.284)
Adj. R^2	0.268	0.265	0.231	0.479	0.620	0.566	0.273	0.600
N	304	304	304	304	304	108	120	76
Date Range	92-17m4	92-17m4	92-17m4	92-17m4	92-17m4	92-00	01-10	11-17m4

This table replicates our baseline time-series analysis in Table A-4, but estimating rate gaps using a loan-specific reset rate $gap_{i,t} = mi, t^* - m_i, t$ rather than assuming all loans reset to the average 30 year FRM: $gap_{i,t} = m^*_{i,t} - m_t$. We predict m_i, t separately for each month t based on a regression of originated loans in month t on a constant, a quadratic in the current FICO score and in current LTV. When FICO or LTV is missing we estimate using only a quadratic in the non-missing variable and when both are missing we again use the average 30 year FRM to estimate gaps. Newey-West standard errors in parantheses. *=10%, **=5%, ***=1% significance. Mean LTV is the ratio of a loan's outstanding balance to value estimated using appraisal values at origination updated using local house price indices from CoreLogic. Loan level data from McDash Performance data+appraisal values from McDash origination data is used to calculate LTV. Prepayment fractions are measured in month $t + 1$ while rate incentives and LTV are measured in month t, since McDash data measures origination not application and there is a 1-2 month lag from application to origination.