Equilibrium Mortgage Design with Heterogeneous Borrowers

August 18, 2026

Abstract

Conventional mortgages price discriminate: by refinancing, sophisticated borrowers pay less than do unsophisticated borrowers, even if their original mortgage terms were identical. I present a model of the mortgage market in which such discriminatory products are offered by lenders to maximize rent extraction from borrowers, who have heterogeneous search costs during mortgage origination. This can justify why mortgages that, for example, refinance automatically are not observed, despite the large and varied social benefits they could generate, as argued in previous literature. The menu of interest rates and "points" that borrowers are typically offered can also be rationalized by this mechanism.

1 Introduction

This paper considers why markets do not provide consumer debt products that better serve households who lack financial sophistication. Nowhere has this received more attention than in the market for residential mortgages, in which borrowers vary greatly in their financial sophistication and access. Sophisticated borrowers refinance at opportune times and so pay a lower effective price for credit than do unsophisticated borrowers, who fail to refinance. In countries like the United States and Denmark, in which fixed-rate mortgages (FRMs) are dominant, prudent borrowers refinance when market interest rates decline, but roughly half of households never do so (see Andersen *et al.* (2020), Zhang (2024), and Berger *et al.* (2024)). In countries like Australia and the United Kingdom, the typical contract is an adjustable-rate mortgage (ARM). Refinancing is important in these countries because such mortgages typically have an initial "teaser" (or "discount") rate, and savvy borrowers refinance upon the expiration of that low rate. And yet, many borrowers fail to do this promptly, as studied by Fisher *et al.* (2024). This pattern generates a regressive transfer, given that refinancing propensities tend to be higher among borrowers in stronger financial positions (see Deng and Quigley (2007), Agarwal *et al.* (2009), Agarwal *et al.* (2015), Keys *et al.* (2016), Johnson *et al.* (2018), Andersen *et al.* (2020), Abel and Fuster (2021), Gerardi *et al.* (2023), and Kiefer *et al.* (2025)). Recent work has also quantified substantial aggregate deadweight costs of the mortgage origination process, with refinancing playing an important role (see Fuster *et al.* (2024) and Zhang (2024)).

As a result, many authors have advocated mortgage designs that would better serve unsophisticated households and undo this inefficient and regressive transfer. A commonly touted approach for countries with FRMs is a provision to allow automatic refinancing when market interest rates fall, which would allow unsophisticated borrowers to reap the benefits of a refinance.[1] Indeed, Campbell and Ramadorai (2025) argue that mortgages with that feature

[1] Berger *et al.* (2024) use a structural model and find that the most-sophisticated borrowers have average effective coupon payments that are 181 basis points lower than the least-sophisticated borrowers over their

should be part of a financial "starter kit" that all U.S. borrowers have access to. But in equilibrium, such a feature is not offered. Instead, as pointed out by Zhang (2024), lenders exacerbate this regressive transfer by offering to partially cover borrowers' closing costs, which sophisticated borrowers take advantage of when refinancing.

In this paper, I present a model in which mortgage design is endogenous to offer a novel explanation of why observed mortgages have features that penalize unsophisticated borrowers and we seldom see alternatives that serve them better. The paper's substantive results are driven by the following fairly obvious observation, which has recently received substantial empirical backing: searching for a mortgage is a costly, difficult, and potentially confusing process for some people. This allows lenders to extract rents during the mortgage origination process. Models of heterogeneous refinancing typically assume perfect competition among lenders,[2] which largely relegates them to simply setting interest rates to break even. In this paper, lenders step to the fore and actively shape the mortgage market. The analysis shows that the possibility of rents has implications not just for how mortgage credit is *priced* (i.e. the interest rate) but also for how the mortgage contract is *designed*.

In the model, when a borrower initially buys their home, they are first matched to a single lender, who makes "captive" mortgage offer(s), which can include a standard FRM, a

lifetime. They perform a counterfactual in which the conventional mortgage is replaced by an automatically refinancing one and find that this alternative regime would feature far less cross-subsidization. Zhang (2024) also studies a structural model of the mortgage market in which borrowers have heterogeneous refinancing costs and finds a cross-subsidy of over $6,000 from unsophisticated to sophisticated types. He also finds that the automatically-refinancing loan can undo the distributional impacts of heterogeneous refinancing, and he additionally finds large efficiency gains from sparing the market from the resource costs associated with refinancing. Fisher *et al.* (2024) find a comparable result when eliminating teaser rates from ARMs in the U.K. They find that if that mortgage were replaced by one that does not reset, the equilibrium rate on that "single-rate" mortgage would be roughly one-third of the way between the pre- and post-reset rates in the current regime. As a result, that alternative contract would help unsophisticated borrowers, who make many payments at the high post-reset rate in the current regime.

[2]The importance of lenders' ability to extract rents is not new to the mortgage literature more generally, however. Scharfstein and Sunderam (2017), Wang *et al.* (2022), and Enkhbold (2026) emphasize the role of market power on monetary policy transmission. Buchak *et al.* (2024) study how the source of funding (balance sheet or securitization) interacts with market power. Jiang (2023) shows how banks use their role in funding shadow bank competitors to complement their market power. Benetton (2021) shows how changes in capital requirements differentially affect lenders' pricing behavior depending on the extent of their market power. Egan *et al.* (2017) structurally estimate a model of deposit competition and show quantitatively that banks' market power shapes the pricing and mix of products offered to households.

FRM with an automatic refinancing provision, or both, each with its own interest rate. The borrower can accept one of the offers or reject all of them and pay a search cost to receive competitive offers from multiple lenders. After the borrower has signed a mortgage, they can subsequently pay some resource cost to refinance into the current market interest rate. Borrowers fall into two categories: Sophisticated, who have low search costs during initial mortgage origination and who subsequently consider the option to refinance; and Unsophisticated, who have high search costs during origination and do not refinance. During the captive phase, lenders choose the menu of offerings to maximize the rents they earn, which are non-zero because of borrowers' search costs. Importantly, because Unsophisticated borrowers have higher search costs, the lender would like to charge them a higher price, but they do not know *ex ante* which type a given borrower is.[3] The core insight of the paper is that the standard FRM is a form of price discrimination: even given an identical contract upon buying their homes, the Unsophisticated borrower ends up paying more over time than does a Sophisticated one, due to the former's failure to refinance. Therefore, the model is able to produce an equilibrium in which lenders choose to withhold the automatic refinancing provision because such a feature would treat all borrowers equally, erase the FRM's ability to price discriminate, and therefore eliminate the windfall for lenders from some borrowers' failure to refinance. In other words, the model is able to generate the simple intuition that lenders benefit from the failure of some borrowers to refinance, something which models with perfect competition cannot contemplate because any such rents would be competed away.

Importantly, this price discrimination comes at a cost to lenders: the deadweight resource cost that active refinancing requires, which lenders (at least partially) bear. This creates

[3]It is important to note that even though some observable characteristics – such as education and income – are able to predict refinancing activity, there is still substantial heterogeneity in refinancing activity within a set of observables. For instance, Andersen *et al.* (2020) show that while higher-income borrowers in Denmark are better at refinancing than lower-income borrowers, borrowers in the bottom quintile of the income distribution succeed in getting more than 60% as much savings from refinancing as those in the top quintile. In a setting in which 50% of borrowers never refinance, this demonstrates that at both ends of the income distribution, there is a meaningful mix of borrower sophistication. This evidence supports the notion that lenders cannot tell based on observables what a borrower's sophistication level is.

a tradeoff for lenders when considering mortgage designs. While an automatic refinancing provision would eliminate their ability to price discriminate, it would increase social surplus, some of which they would be able to capture. The paper shows how other mortgage design features, such as prepayment penalties and points, can be used by lenders to navigate this tradeoff more deftly than the automatic refinancing provision. Prepayment penalties lessen deadweight refinancing effort without eliminating it and also lead to less extraction from Unsophisticated borrowers than the conventional mortgage, due to a lowered equilibrium interest rate. Paying some of borrowers' closing costs (i.e. "negative points") at origination, on the other hand, can increase extraction from Unsophisticated borrowers by increasing equilibrium interest rates, though by encouraging Sophisticated borrowers to refinance, this also increases deadweight loss. An off-the-shelf calibration of a model with heterogeneous refinancing behavior shows that lenders can increase their rents by offering points, whereas prepayment penalties would lower rents, and an automatic refinancing provision would lower rents even further. In other words, it can explain key features of observed mortgages: active refinancing is required for mortgages across the world, whether they be FRMs or ARMs; it is common for borrowers to be able to lower their closing costs by accepting a higher interest rate; and prepayment penalties have become quite rare in recent decades.[4] The results on prepayment penalties and points are not just interesting applications of the model but also evidence that the mechanism described above is important.

While positive issues of equilibrium mortgage design are less studied than normative issues of optimal mortgage design,[5] this is not the first paper to consider the former topic. Mayer *et al.* (2013) explain the prevalence of prepayment penalties on subprime loans in the early 2000s as a form of risk-sharing, as borrowers who receive a positive credit shock that allows them to qualify for a prime loan pay the penalty, essentially subsidizing the borrowers who

[4]France and Germany are exceptions, as prepayment penalties are common there. See Lea (2010).

[5]For additional work on the normative side, see Piskorski and Tchistyi (2011), Campbell (2013), Eberly and Krishnamurthy (2014), Campbell *et al.* (2021), Guren *et al.* (2021), Piskorksi and Seru (2018), and Bhagat (2021).

do not receive such a shock. Piskorski and Tchistyi (2020) consider what type of state-contingent mortgage would result in equilibrium, given heterogeneity in household tastes for homeownership, as well as idiosyncratic and aggregate income shocks and an option for mortgage default. Relatedly, Hartman-Glaser and Hebert (2020) consider mortgages with terms that depend on home price indexes and argue that they may not be observed if borrowers worry that the lender is sufficiently more knowledgeable about the indexes. The issue of why we do not see mortgages with features that better serve unsophisticated borrowers is far from answered, though.

On the institutional side, one explanation (see, e.g., Campbell (2006) and Campbell (2013)) is "inertia:" the observed FRM took hold for historical reasons and so is hard to displace because households are comfortable with it and important financial institutions (such as the government-sponsored enterprises (GSEs), Fannie Mae and Freddie Mac) have designed their strategies around FRMs as the standard contract. While inertia likely does play some role, this explanation ignores that there is substantive innovation in the mortgage market. For instance, even the FRM has a long history of innovation: it once had a 20-year term and was assumable by a home's new owner upon sale, and prepayment penalties have come and gone. ARMs as we know them were introduced in the U.S. in the 1960s and GSEs began engaging with them in the 1980s (see Lea and Sanders (2011)). And the housing boom of the 1990s and early 2000s witnessed a flurry of innovation, as mortgages with a variety of payment streams sprang up. So while it is hard to rule out a role for inertia, other financial innovations have overcome it.[6,7] Another proposed explanation comes from Berger *et al.* (2024), who point out that an automatic refinancing provision would lead to a higher initial interest rate and so could cause borrowers to run afoul of payment-to-income ratios.

[6]Indeed, mortgages with automatic refinancing provisions seem appealing because they would be easier to price, as there would be no need to model the substantial heterogeneity in borrower refinancing behavior that exists with current FRMs.

[7]Additionally, while it can be tempting to simply attribute the dominance of the FRM to its adoption by the GSEs, this likely has the direction of the causality flipped, at least in part: the GSEs focus on FRMs because they are dominant.

But the automatic refinancing provision could require a certain threshold of savings in order to trigger in such a way that average refinancing propensities are maintained and the initial interest rate would not be higher.

From a "behavioral" angle, Gabaix and Laibson (2006) argue that the process of marketing such a provision to unsophisticated borrowers would in fact make them sophisticated and undermine their interest in the product, as they would then prefer the standard FRM with a refinancing option subsidized by the still-unsophisticated. However, this mechanism cannot explain why other alternative mortgage types, such as one that cannot be refinanced (or one with high prepayment penalties), do not exist. After all, a mortgage that cannot be refinanced is in fact easier to understand than the FRM, which has a complex option to refinance. Because such a mortgage would come with a lower initial interest rate, it becomes hard to argue that this would be difficult to market. Another possibility is that borrowers are overconfident about their future refinancing behavior and so do not realize that they are overpaying with the standard FRM. However, this alone is not sufficient because the efficiency gains from the automatically refinancing mortgage would allow it to be offered at a lower cost of credit than the standard FRM, even assuming the latter is refinanced optimally. While not seeking to rule out a role for any of these explanations, the present paper shows that observed mortgage designs can be rationalized in a model in which borrowers understand mortgage contracts without need to appeal to inertia or institutional norms, which are hard to quantitatively evaluate. The key driver of this result is the relaxation of the assumption of perfect competition among lenders, stemming from borrowers' search costs.

The assumption of substantive search costs is supported by a growing empirical literature showing that mortgage lenders do extract rents differentially across borrowers and that borrower sophistication plays a key role in driving this.[8] Woodward and Hall (2010) and Wood-

[8]Nelson (2025) also finds a qualitatively important role for lender rents in the market for credit cards. Bjonnes *et al.* (2024) document extensive variation in execution costs for OTC trades, with client sophistication being the primary driver of that variation. Egan (2019), Egan *et al.* (2019), and Egan *et al.* (forthcoming) demonstrate extraction from relatively inexperienced retail households by financial advisers.

ward and Hall (2012) provide evidence that borrowers struggle to assess the true cost of their mortgages, given that the price has multiple dimensions (interest rate and points/closing costs). This leads to substantial variation in the compensation that borrowers pay to mortgage brokers, with (local) education levels being a notable driver of the variation, suggesting an important role for financial sophistication. Agarwal *et al.* (2017) similarly find that borrowers who overpay brokers tend to have lower levels of education and furthermore that they are less likely to refinance. In the same vein, Gurun *et al.* (2016), Agarwal *et al.* (2024), and Bhutta *et al.* (2025) all show evidence that borrowers vary in their abilities to get favorable mortgage terms, conditioning on observables. Furthermore, this literature shows suggestive evidence that search behavior is an important driver of the variation, with financial sophistication being an important driver of that. This is empirical evidence of precisely the relationship that drives the conceptual results in the present paper: the borrowers who are penalized by the FRM due to their lack of refinancing are largely the same ones vulnerable to extraction by lenders during mortgage origination.

While the discussion above repeatedly emphasizes that FRM refinancing leads to different costs of credit for different borrowers, the reader should keep in mind that, as previously discussed, refinancing is important for observed ARMs (with teaser rates), and so this mechanism is also relevant for understanding markets where ARMs are dominant, like the U.K. Miles (2004) and Campbell (2006) make the point that this teaser rate structure is not exogenous and may exist to reward sophisticated borrowers at the expense of the unsophisticated. This paper strengthens that intuition by pointing out that, under the reasonable assumption that lenders can extract additional rents from unsophisticated borrowers, this is precisely the type of feature we would expect to see for ARMs. So while the model below primarily discusses FRMs, the core ideas extend more generally to explain why we do not see more equitable and efficient mortgages that better serve those who lack financial sophistication.

2 Model of Mortgage Design

This section presents a model of mortgage design. Section 2.1 follows Agarwal *et al.* (2013) and Berger *et al.* (2025) to present an analytically tractable, quantitatively relevant model of refinancing, *given the design of the mortgage and the initial interest rate.* The tractability of that partial equilibrium model allows it to be embedded within the broader model of equilibrium mortgage design, as presented in Section 2.2: conditional on borrower refinancing behavior as described in Section 2.1, lenders choose optimal mortgage design. The model has a number of features useful for the quantitative analysis in subsequent sections. To highlight the key economic forces governing the qualitative results, Internet Appendix A presents a version of the model that is simplified on a number of dimensions, such as having only two periods rather than unfolding in continuous time. As highlighted in that appendix, the results are driven by three features of the model: 1) lender market power, subject to borrowers' participation constraints; 2) unobservable heterogeneous refinancing propensity among borrowers; 3) negative correlation between a borrower's refinancing propensity and the cost of its outside option. Together, these features can lead to the use of the inefficient, inequitable FRM by lenders. Market power incentivizes lenders to extract differentially from borrowers, and the FRM is an effective way to do this because it charges a higher effective price to people who do not refinance, the same borrowers with worse outside options – precisely who the lenders would like to extract most from.

2.1 "Partial Equilibrium" Mortgage Costs, Taking Mortgage Contract As Given

2.1.1 Model Setting

Consider a borrower with a mortgage, which has an exogenous principal balance that without loss of generality is normalized to \$1. Assume all agents are risk-neutral[9] and have discount

[9]Internet Appendix H presents a model in which households are risk averse. Results from that model are discussed below.

rate ρ. This paper will primarily focus on two types of mortgages: a FRM and a Self-Refinancing Mortgage (SRM). The former is akin to the standard contract in the U.S., which can be refinanced only if the borrower bears some refinancing cost. The latter is similar, except that it refinances costlessly any time the market rate on such a mortgage is below the borrower's current rate.[10] In setting up the model, this section focuses on the FRM; the SRM is discussed subsequently.

Assume that the real short-term interest rate, r, follows a Brownian motion:

$$dr = \sigma \cdot dz, \tag{1}$$

where dz represents standard Brownian increments. Assume inflation is a constant π. Given this setup, the nominal short-term interest rate, $i_t = r_t + \pi$, is a random walk.[11]

For now, assume that the interest rate on a FRM, m_t^{FRM}, exceeds the nominal short-term interest rate by a constant wedge:

$$m_t^{FRM} = i_t + \Delta^{FRM}. \tag{2}$$

Section 2.2 shows that under the assumptions of the model of equilibrium mortgage design, an equilibrium exists in which this assumption is indeed true.

I make the simplifying assumption that mortgages are "interest-only" contracts, meaning the borrower pays interest but does not pay down the balance of the loan during the life of the contract. The principal of the loan is paid back in entirety upon the realization of an exogenous "moving shock," which arrives with a hazard rate of μ. The assumption that the mortgage is interest-only simplifies the partial equilibrium problem of the borrower by removing the mortgage balance as a state variable (as it does not vary), but more crucially

[10] Given the other assumptions of the model, a mortgage that automatically refinances when interest rates fall by a certain non-zero threshold would generate all the same results as the SRM.

[11] Internet Appendix G presents a model with mean reversion in interest rates.

it ensures that m_t^{FRM} follows a Brownian motion, as discussed in Section 2.2.[12]

2.1.2 Borrower Refinancing Behavior

Let m_0^{FRM} be the interest rate the borrower is currently paying on her mortgage – the one she agreed to at origination. At any time during the life of the loan, she can pay some cost C^{Refi} to replace m_0^{FRM} with m_t^{FRM}, the current mortgage rate in the market. Assume that the borrower does not always pay attention to interest rates, a point with empirical backing from Andersen *et al.* (2020) and Maturana and Nickerson (2019) and also used to model the mortgage market by Zhang (2024), Berger *et al.* (2024), and Berger *et al.* (2025). In particular, following Berger *et al.* (2025), she is only aware of her refinancing incentive when she receives an "attention shock," which arrives at a hazard rate of χ.[13]

The change in the interest rate that a borrower could achieve through refinancing, $y_t \equiv m_t^{FRM} - m_0^{FRM}$, is the critical state variable of the problem, as she must determine how far the interest rate has to fall in order justify the cost of refinancing. Note that y follows a Brownian motion, with $dy = \sigma \cdot dz$.

Following Berger *et al.* (2025), Internet Appendix B shows how to find $y^*(\chi)$, the threshold value such that a borrower with attention parameter χ who is paying attention at time t will refinance if and only if $y < y^*(\chi)$, which results from a system of three equations in three unknowns (the other two unknowns being coefficients in the borrower's value function) that is guaranteed to have a unique solution.

2.1.3 Expected Future Costs, Evaluated At Origination

Define $K^{FRM}(m_0^{FRM}, \chi)$ to be the net present value (NPV) of expected costs, at the time of origination, for a borrower with attention parameter χ who signs her FRM contract with

[12]The assumption of an interest-only mortgage differs subtly from the model of Agarwal *et al.* (2013), as discussed in Internet Appendix F. There, I demonstrate that the assumption does not substantively impact the paper's findings.

[13]This is similar to the common assumption of random attention for firms when they set prices, as modeled by Calvo (1983) and then many subsequent paper.

interest rate m_0^{FRM}. Then:

$$K^{FRM}(m_0^{FRM},\chi) = P^{FRM}(m_0^{FRM},\chi) + D^{FRM}(\chi), \tag{3}$$

where $P^{FRM}(m_0^{FRM},\chi)$ accounts for costs that are captured by the lender in the form of mortgage payments and $D^{FRM}(\chi)$ accounts for deadweight refinancing costs that are captured by no one.[14] This decomposition of overall borrower costs into a portion paid to the lender and a portion that is deadweight loss is critical for the analysis that follows. Appendix B derives the following solutions for $P^{FRM}(m_0^{FRM},\chi)$ and $D^{FRM}(\chi)$ conditional on $y^*(\chi)$, which – as discussed in the previous subsection – itself is implicitly solved for:

$$P^{FRM}(m_0^{FRM},\chi) = \frac{m_0^{FRM}+\mu}{\rho+\mu+\pi} + \frac{\chi\cdot\left(\eta\cdot\frac{y^*(\chi)}{\rho+\mu+\pi} - \frac{1}{\rho+\mu+\pi}\right)}{(\rho+\mu+\pi+\chi)\cdot(\eta+\psi)\cdot exp(-\psi\cdot y^*(\chi)) - \chi\cdot\eta}; \tag{4}$$

$$D^{FRM}(\chi) = \frac{\chi\cdot\eta\cdot C^{Refi}}{(\rho+\mu+\pi+\chi)\cdot(\eta+\psi)\cdot exp(-\psi\cdot y^*(\chi)) - \chi\cdot\eta}, \tag{5}$$

where $\psi \equiv \frac{\sqrt{2\cdot(\rho+\mu+\pi)}}{\sigma}$ and $\eta \equiv \frac{\sqrt{2\cdot(\rho+\mu+\pi+\chi)}}{\sigma}$.

2.1.4 Self-Refinancing Mortgages

The same setup and analysis can be applied to the SRM. To start, again assume that the interest rate on the SRM is a constant wedge above the nominal short-term interest rate, which will be confirmed in Section 2.2:

$$m_t^{SRM} = i_t + \Delta^{SRM}. \tag{6}$$

[14]The deadweight loss does not depend on the initial interest rate, m_0^{FRM}, because the borrower's refinancing behavior is identical regardless of where the mortgage's interest rate starts, as shown in Internet Appendix B.

There are two key distinctions between the SRM and the FRM. First, the SRM has no refinancing cost. Second, a borrower's attention parameter, χ, is irrelevant, as the mortgage refinances costlessly and automatically whenever $y < 0$. Effectively, $C^{Refi} = 0$ and $\chi \to \infty$. Letting $K^{SRM}(m_0^{SRM})$ be the NPV of expected costs, at the time of origination, for a borrower who signs her SRM with interest rate m_0^{SRM}, we have:

$$K^{SRM}(m_0^{SRM}) = P^{SRM}(m_0^{SRM}), \tag{7}$$

where $P^{SRM}(m_0^{SRM})$ is the expected payment to the lender.[15] SRM outcomes differ from the FRM in two critical ways. First, there is no deadweight loss, since borrowers are not required to exert effort in order to refinance. In other words, all costs to the borrowers are recovered by the lenders. Second, conditional on an initial interest rate, all borrowers bear the same costs. Because refinancing does not require input from the borrowers, borrower sophistication is irrelevant and there is no heterogeneity in payments or costs.

2.2 Equilibrium Mortgage Design

During the mortgage origination process, lenders choose what mortgage products to offer borrowers in order to maximize profits, given borrowers' refinancing behavior and the ensuing costs and payments, as described in Section 2.1.

2.2.1 The Mortgage Origination Setting

Assume that at every instant t, a cohort of risk-neutral households appears, each needing to borrow \$1 for a mortgage. Assume fraction S_{Soph} are "Sophisticated" borrowers, who would refinance a FRM optimally given some attention parameter $\chi > 0$, as described in Section 2.1; fraction $(1 - S_{Soph})$ is "Unsophisticated," and their attention parameter is 0; they do

[15] Specifically, $P^{SRM}(m_0^{SRM}) = \left(m_0^{SRM} + \mu - 1/\psi\right) \cdot \frac{1}{\rho+\mu+\pi}$, which follows from L'Hopital's Rule.

not refinance their mortgages.[16,17]

Assume there are $N \geq 2$ lenders. Mortgage origination proceeds as follows. Each borrower j matches to a single lender and receives offer(s) during this "captive phase" – an SRM and/or a FRM, each at some specified interest rate. The borrower can accept one of the offers, or she can reject all offers and pay a search cost to "shop around," at which point she enters a "Bertrand stage" in which she receives offers from all lenders and chooses the one with the lowest cost.[18] Unsophisticated borrowers have higher search costs: $C^{Search}_{Uns} \geq C^{Search}_{Soph} \geq 0$. Borrowers' search costs enable lenders to extract rents by offering supra-competitive interest rates during the captive phase: if chosen correctly, borrowers will accept paying an interest rate with a profit margin because it is costly to find a better one. Critically, while lenders know the *share* of borrowers that is Sophisticated, they do not know whether any individual borrower is Sophisticated or Unsophisticated. Therefore, lenders must maximize profits by offering the same mortgage(s) to all borrowers.[19] The assumption that Unsophisticated borrowers have higher search costs is critical to the results of the model, as that will incentivize lenders to set up a price discrimination mechanism to extract higher rents from them. Costly refinancing will be central to that mechanism.

[16] As discussed in Section 3, empirical analyses have shown that a large fraction of borrowers can be appropriately modeled as having an attention parameter of 0.

[17] A more streamlined version of the model would not include heterogeneous attention but would assume instead that refinancing requires paying a search cost, which as discussed below is assumed to be another dimension of heterogeneity between Sophisticated and Unsophisticated borrowers. This would be streamlined in the sense of allowing heterogeneity to be described by a single parameter rather than a pair. However, inattention is considered as a separate friction in the model because the empirical literature, such as Andersen *et al.* (2020), has argued that it is an important source of heterogeneous refinancing. A key pattern in the data is that borrowers commonly refinance after eschewing an earlier opportunity to achieve even higher payment savings, a result that is difficult to rationalize solely through search costs. Note, however, that the theoretical results would go through even if inattention were not included and heterogeneous refinancing came from differences in search costs. That would still generate a situation in which lenders want to extract from Unsophisticated borrowers and the FRM would be effective in doing so due to their lack of refinancing, the key mechanism behind the results below.

[18] The assumption of Bertrand competition is not central and is made for simplicity. The key attribute that is needed is that searching leads to a lower interest rate (in expectation).

[19] This should be thought of as being conditioned on observables. So if lenders observe borrowers' credit scores, they have to offer the same mortgage to all borrowers with the same credit score, but they can offer different mortgages to borrowers with different credit scores. As mentioned above, Andersen *et al.* (2020) show that even with observable categories, there is substantial variation in sophistication among borrowers.

One more assumption is useful for ruling out extreme lender behavior when Sophisticated borrowers have low search costs. The assumption is that Unsophisticated borrowers will be matched to whichever lender has a higher market share among Sophisticated borrowers; if lenders have the same market share among Sophisticated borrowers, the Unsophisticated borrowers will be split evenly among them. Essentially, Unsophisticated borrowers are attracted to large lenders that have "name recognition." This ensures that lenders will still want to attract Sophisticated borrowers even if they cannot extract much from them (due to low search costs), as doing so will entice Unsophisticated borrowers, who are more profitable. Without this assumption, lenders would ignore Sophisticated borrowers if their search costs are low, essentially eliminating heterogeneity in the model and obviating any incentive to price discriminate. An alternative explanation for why lenders would still market to Sophisticated borrowers is that they want to attract their business for other types of non-mortgage needs. Essentially, the name recognition formulation amounts to forcing lenders to attract Sophisticated borrowers. Without some mechanism for this, the heterogeneity at the heart of the paper could disappear with low search costs for the Sophisticated type.

Concretely, the probability that borrower i is matched to lender j for the captive phase is $P_{ij} = 1/N$ if the borrower is Sophisticated. Denote by Φ_j^{Soph} the share of Sophisticated borrowers who borrow from lender j in equilibrium. If $\tilde{N} \leq N$ is the number of lenders with $\Phi_j^{Soph} = \max\{\Phi_k^{Soph}\}$, then the probability that Unsophisticated borrower i is matched to lender j in the captive phase is:

$$P_{ij} = \begin{cases} 1/\tilde{N} & \text{if } \Phi_j^{Soph} = \max\{\Phi_k^{Soph}\} \\ 0 & \text{if } \Phi_j^{Soph} < \max\{\Phi_k^{Soph}\} \end{cases}. \tag{8}$$

2.2.2 *Reservation Interest Rates*

Given the above setup, lenders must choose which mortgages to offer and at what interest rates. To solve for these choices, we must first describe borrowers' reservation interest rates, the highest interest rates they would be willing to accept during the captive phase of the mortgage origination process. Denote these reservation interest rates by $\bar{m}_j^{FRM}$ and $\bar{m}_j^{SRM}$. A borrower of type j would prefer an FRM with any interest rate below $\bar{m}_j^{FRM}$ over performing a costly search for more offers; the reservation interest rate for a SRM is defined analogously.

A lender who makes a loan to the borrower is foregoing the opportunity to earn the short term interest rate, i_t, until a moving shock forces the repayment of the mortgage.[20] Denoting by i_0 the nominal short-term interest rate at the time of origination, the expected NPV to lenders of earning the short-term interest rate is equal to $\frac{i_0+\mu}{\rho+\mu+\pi}$, so this is what the borrower will pay in mortgage costs if she proceeds to the Bertrand phase.[21] This means that during the captive phase, the borrower's outside option to reject all offers has a cost of $\frac{i_0+\mu}{\rho+\mu+\pi} + C_j^{Search}$, with the latter term representing the search cost required to receive a mortgage offer in the competitive Bertrand phase.

Therefore, the reservation interest rates are characterized by the following equations:

$$\frac{i_0+\mu}{\rho+\mu+\pi} + C_j^{Search} = K^{SRM}(\bar{m}_j^{SRM}) = K^{FRM}(\bar{m}_j^{FRM}, \chi_j), \tag{9}$$

which state that the reservation interest rates on the SRM and FRM are those that generate the same overall costs to borrowers as paying a search cost and then signing a mortgage with

[20] This implicitly assumes that when refinancing, the borrower's new loan is with the original lender, or at least that the lender replaces the refinanced mortgage with another one. This is simplifying in that it keeps the lender's opportunity cost of capital identical across borrower types.

[21] This is a standard zero profit assumption. I will focus exclusively on symmetric Nash Equilibria in which all lenders offer mortgages that yield the opportunity cost of capital in the Bertrand phase. Note that lenders could have an incentive to offer mortgages that earn less than the opportunity cost as a way of gaining more market share among Sophisticated borrowers in order to capture the Unsophisticated market segment. However, in equilibrium, all Sophisticated borrowers will accept offers in the captive phase, meaning that there are no rejections, so offering zero-profit mortgages in the Bertrand phase is part of a Nash Equilibrium.

a cost equal to the lender's opportunity cost of capital.

The reservation interest rates on a FRM are therefore:

$$\bar{m}_j^{FRM} = i_0 + v_j + \omega_j^{FRM}, \tag{10}$$

where v_j accounts for expected profit[22] and ω_j^{FRM} is a premium for the option to refinance.[23]

Similarly, the reservation interest rates on a SRM are:

$$\bar{m}_j^{SRM} = i_0 + v_j + \omega^{SRM}. \tag{11}$$

This has the same type-specific profit margin as the FRM, but the SRM's premium for the option to refinance is the same for all borrowers, since it generates the same refinancing behavior for all borrowers.[24]

Sophisticated borrowers will have a lower reservation interest rate on the SRM, because it is cheaper for them to shop around and find a better offer. However, it is ambiguous which borrower type will have a higher reservation interest rate on an FRM. Intuitively, Unsophisticated borrowers have higher search costs, which imply a higher reservation interest rate, as their outside option of searching is costlier. On the other hand, Sophisticated borrowers' higher propensity to refinance could lead to a sufficiently high refinancing premium such that they have a higher reservation interest rate. Intuitively, despite their lower search costs, Sophisticated borrowers may be willing to accept a higher FRM interest rate because they know they are likely to refinance it downward subsequently. So, we could have $\bar{m}_{Uns}^{FRM} > \bar{m}_{Soph}^{FRM}$ or $\bar{m}_{Uns}^{FRM} < \bar{m}_{Soph}^{FRM}$, and the direction of the inequality is one key determinant of equilibrium design, which we turn to now.

[22]Specifically, $v_j = C_j^{Search} \cdot (\rho + \mu + \pi)$.

[23]Specifically, $\omega_j^{FRM} = \frac{\chi \cdot \left(1 - \eta \cdot (y^*(\chi) - (\rho+\mu+\pi) \cdot C^{Refi})\right)}{(\rho+\mu+\pi+\chi) \cdot (\eta+\psi) \cdot exp(-\psi \cdot y^*(\chi)) - \chi \cdot \eta}$.

[24]Specifically, $v_j = C_j^{Search} \cdot (\rho + \mu + \pi)$ (as with the FRM), and $\omega^{SRM} = 1/\psi$.

2.2.3 Lenders' Profit-maximizing Mortgage Offer(s)

The main conceptual result of the paper occurs in the case when $\bar{m}^{FRM}_{Uns} > \bar{m}^{FRM}_{Soph}$, as this condition allows for the possibility that no SRM will be offered in equilibrium. To see why, note that if the lender offers the SRM in the captive stage, the highest interest rate that will be acceptable to Sophisticated borrowers is $\bar{m}^{SRM}_{Soph}$, and profit will be:

$$\Pi_{SRM} = P^{SRM}(\bar{m}^{SRM}_{Soph}) - \frac{i_0 + \mu}{\rho + \mu + \pi}. \tag{12}$$

The first term is the expected payments received, which are equal between Sophisticated and Unsophisticated borrowers since the SRM refinances without their input; the second term is the opportunity cost of capital.

If the lender instead offers the FRM in the captive stage, the highest interest rate that will be acceptable to Sophisticated borrowers is $\bar{m}^{FRM}_{Soph}$, and profit will be:

$$\Pi_{FRM} = S_{Soph} \cdot P^{FRM}(\bar{m}^{FRM}_{Soph}, \chi) + (1 - S_{Soph}) \cdot P^{FRM}(\bar{m}^{FRM}_{Soph}, 0) - \frac{i_0 + \mu}{\rho + \mu + \pi}. \tag{13}$$

The first term represents payments from Sophisticated borrowers, the second term represents payments from Unsophisticated borrowers, and the last term is the opportunity cost of capital. Unlike the SRM, the FRM generates different revenue streams from Sophisticated and Unsophisticated borrowers, due to their different refinancing propensities.[25]

With that setup, the paper's main conceptual result is:

Proposition 1 *If $\bar{m}^{FRM}_{Uns} > \bar{m}^{FRM}_{Soph}$, then there exists an equilibrium with no SRM and only an FRM with interest rate $m^{FRM}_t = i_t + \Delta^{FRM}$, where $\Delta^{FRM} = \upsilon_{Soph} + \omega^{FRM}_{Soph}$, if and only*

[25] Lenders are identical, and as the focus is on symmetric equilibria in which they take the same actions, all lenders face the same share of Sophisticated borrowers, which is equal to S_{Soph}.

if the following condition holds:

$$\underbrace{(1-S_{Soph})\cdot\left(P^{FRM}(\bar{m}^{FRM}_{Soph},0)-P^{SRM}(\bar{m}^{SRM}_{Soph})\right)}_{\substack{\text{Additional revenue extracted from}\\ \text{Unsophisticated borrowers by FRM}}} > \underbrace{S_{soph}\cdot D^{FRM}(\chi)}_{\substack{\text{Reduced revenue extracted from}\\ \text{Sophisticated borrowers by FRM}}} \quad . \tag{14}$$

If Expression 14 does not hold, equilibrium will include only an SRM. In this case, the interest rate will be $m^{SRM}_t = i_t + \Delta^{SRM}$*, where* $\Delta^{SRM} = \upsilon_{Soph} + \omega^{SRM}$.

Proposition 1 is proven in Internet Appendix C, but the intuition is as follows. Expression 14 shows the core tradeoff between SRMs and FRMs from the lenders' perspective. On the one hand, the FRM generates less revenue from Sophisticated borrowers because it forces them to exert effort to refinance. Offering a SRM would spare the borrowers those costs and therefore generate surplus, which the lender could capture by increasing the interest rate on the SRM to $\bar{m}^{SRM}_{Soph}$. So, the righthand side of the inequality is the deadweight loss for Sophisticated borrowers, scaled by their share.

On the other hand, FRMs extract more revenue from Unsophisticated borrowers than do SRMs. Intuitively, the FRM gets the Unsophisticated borrowers to pay a premium for a refinancing option that they will not use. Of course, if they could search freely, they would prefer a SRM that allows them to use the refinancing option, or alternatively a FRM with a lower interest rate to better reflect their inability to refinance it. But, because of their high search costs, they are forced to accept the FRM and waste its option to refinance, to the benefit of lenders. This is the lefthand side of the inequality.

The FRM can therefore be understood as a mechanism to price discriminate. Lenders would like to charge a higher price for mortgage credit to Unsophisticated borrowers, who find it more difficult to shop around. The SRM is incapable of doing that because it treats all borrowers the same. On the other hand, the FRM succeeds in charging more to the Unsophisticated borrowers; even though all borrowers agree to the same interest rate, in expectation Sophisticated borrowers pay less over time because they refinance. The FRM therefore allows lenders to charge a high "sticker price" for mortgage credit, which the Unsophisticated borrowers pay. Sophisticated borrowers are willing to accept the same mortgage because they can ultimately pay a lower price by refinancing. The mechanism is not perfect, however, because it requires effort by Sophisticated borrowers to pass the lenders' screen and get the lower price. That cost shrinks social surplus and therefore shrinks the lenders' ability to extract revenue. If Expression 14 did not hold, this effect would dominate and lenders would do better by offering the SRM.[26]

Equilibrium mortgage design can differ when $\bar{m}^{FRM}_{Uns} < \bar{m}^{FRM}_{Soph}$:

Proposition 2 *If $\bar{m}^{FRM}_{Uns} < \bar{m}^{FRM}_{Soph}$, then Sophisticated borrowers will agree to FRMs with an interest rate of $m^{FRM}_t = i_t + \Delta^{FRM}$, where $\Delta^{FRM} = v_{Soph} + \omega^{FRM}_{Soph}$, and Unsophisticated borrowers will agree to SRMs with an interest rate of $m^{SRM}_t = i_t + \Delta^{SRM}$, where $\Delta^{SRM} = v_{Uns} + \omega^{SRM}$, in equilibrium if and only if the following condition holds:*

$$\underbrace{(1 - S_{Soph}) \cdot \left(P^{SRM}(\bar{m}^{SRM}_{Uns}) - P^{SRM}(\bar{m}^{SRM}_{Soph})\right)}_{\substack{\textit{Additional revenue extracted from} \\ \textit{Unsophisticated borrowers when FRM is offered}}} > \underbrace{S_{soph} \cdot D^{FRM}(\chi)}_{\substack{\textit{Reduced revenue extracted from} \\ \textit{Sophisticated borrowers by FRM}}} . \qquad (15)$$

If Expression 15 does not hold, equilibrium will include only an SRM. In this case, the interest rate will be $m^{SRM}_t = i_t + \Delta^{SRM}$, where $\Delta^{SRM} = v_{Soph} + \omega^{SRM}$.

This, too, is proven in Internet Appendix C, but the intuition is quite similar to the price discrimination explanation of Proposition 1, even if the mechanism differs somewhat. In the previous result, the lender price discriminated by having the borrowers all agree to the same mortgage, but then allowing the Sophisticated borrowers to get a lower price *ex-post* by refinancing. In Proposition 2, the lender price discriminates by having them agree to different mortgages *ex-ante*. This is possible because $\bar{m}^{FRM}_{Uns} < \bar{m}^{FRM}_{Soph}$ and $\bar{m}^{SRM}_{Uns} > \bar{m}^{SRM}_{Soph}$. As a result, the lender can get the Unsophisticated borrower to agree to a SRM at her

[26]Internet Appendix C shows why – if $\bar{m}^{FRM}_{Soph} < \bar{m}^{FRM}_{Uns}$ – lenders cannot do better by simultaneously offering both the FRM and SRM.

reservation interest rate ($\bar{m}^{SRM}_{Uns}$) by ratcheting up the FRM interest rate to a level she is not willing to accept. On the other hand, the Sophisticated borrowers will be willing to accept this high FRM interest rate because of their likelihood of refinancing it downward subsequently. Intuitively, the Unsophisticated borrower does not want the FRM because it has an expensive refinancing option that she will not use, so they are willing to accept a SRM with a high profit margin; Sophisticated borrowers are willing to pay for the FRM's refinancing option because they will use it, and they do not want to pay the SRM's high profit margin.

Just as in Proposition 1, Proposition 2 shows the tradeoff associated with the choice of whether to offer the FRM: while it allows for increased revenue from Unsophisticated borrowers via price discrimination (lefthand side of Expression 15), it has the drawback of requiring costly refinancing from Sophisticated borrowers, and so lenders are able to extract less from them (righthand side). When the expression holds, offering the FRM is worthwhile because it allows lenders to separate borrowers *ex-ante* and price discriminate; when the expression does not hold, lenders should offer only an SRM.

2.3 Comparison to Standard Assumption of Perfect Competition

Proposition 1 rationalizes how lenders can have the ability to offer an SRM, which is efficiency-enhancing and seemingly well-tailored to Unsophisticated borrowers who fail to refinance, and yet choose not to offer it in equilibrium. This is not a result that is easily rationalized in a standard model that assumes Perfect Competition ("PC").

To see why, consider the model above without search costs ($C^{Search}_j = 0$), so lenders must offer mortgages that earn zero profit in expectation. Suppose temporarily that the only possible mortgage is an FRM. Then, defining $\hat{m}^{FRM}_{PC}$ to be the interest rate on the FRM, we would have:

$$S_{Soph} \cdot P^{FRM}(\hat{m}_{PC}^{FRM}, \chi) + (1 - S_{Soph}) \cdot P^{FRM}(\hat{m}_{PC}^{FRM}, 0) = \frac{i_0 + \mu}{\rho + \mu + \pi}. \tag{16}$$

The lefthand side is a weighted average of the payments from Sophisticated and Unsophisticated borrowers, and the righthand side is the opportunity cost of the capital. Because Unsophisticated borrowers pay more in expectation than Sophisticated borrowers at a given initial interest rate, this implies that Unsophisticated borrowers pay more than the lenders' opportunity cost of capital.

The introduction of a SRM will unravel the FRM market such that no borrowers will choose the latter type of mortgage. To see why, note that lenders can offer the SRM with an interest rate of m_{PC}^{SRM} such that:

$$P^{SRM}(m_{PC}^{SRM}) = \frac{i_0 + \mu}{\rho + \mu + \pi}. \tag{17}$$

Unsophisticated borrowers will choose the SRM, as it allows them to pay only the opportunity cost of capital, which is less than what they were paying in the equilibrium characterized by Equation 16. This causes adverse selection in the FRM market, as lenders will be left with only Sophisticated borrowers who refinance often, so the $\hat{m}_{PC}^{FRM}$ characterized above would yield negative profit. Lenders would instead have to offer FRMs with a higher interest rate, m_{PC}^{FRM}, such that:

$$P^{FRM}(m_{PC}^{FRM}, \chi) = \frac{i_0 + \mu}{\rho + \mu + \pi}. \tag{18}$$

Such a mortgage would allow lenders to break even on Sophisticated borrowers. However, Sophisticated borrowers would not choose such a mortgage, because the payments to lenders only capture part of their costs; when the refinancing costs ($D^{FRM}(\chi)$) are factored in, they would end up paying more than the opportunity cost of capital. Therefore, they, too, would choose the SRM.[27]

This deepens the puzzle of why SRMs are not observed: under a typical Perfect Competition assumption, they should not only exist, but they should dominate the market, quite the opposite of what is observed. Assuming Perfect Competition eliminates the tradeoff lenders face when weighing the merits of FRMs and SRMs; the total absence of lender rents eliminates their ability to price discriminate, which is the benefit of offering the FRM. With that benefit removed, the SRM dominates due to the inefficiency caused by the FRM's costly refinancing. The model in this paper makes a minor deviation from Perfect Competition: lenders are prepared to compete fiercely, but (at least some) borrowers have to pay a cost to initiate that process. This gives lenders a measure of power to extract rents, and if the conditions are right, they can do so by using a FRM to price discriminate and extract a supra-competitive return. As such, the SRM can be absent from the market.

2.4 Robustness

This subsection explores the robustness of the results to various modeling assumptions. First, the model above greatly simplifies the real world's mortgage finance pipeline by simply modeling "lenders." In contrast to this monolith, there is in fact a complex "originate-to-distribute" model in which numerous entities play a role, from mortgage originators to final investors in mortgage-backed securities (MBS), each with a unique role and compensation structure. Of particular concern for the present paper is that the mortgage originators – the people who actually interact with borrowers and can exploit the variation in search costs as discussed above – sell the mortgages promptly. Given that they therefore do not have a direct stake in the flow of interest payments and instead make money from origination fees, is it valid for the model to assume they seek to maximize the NPV of mortgage payments?

To the extent that participants in the mortgage credit system are able to competently eval-

[27] Berger *et al.* (2024) make a similar point about unraveling, but because they do not consider the efficiency consequences of refinancing, they find that the most sophisticated borrowers may still end up in a FRM under Perfect Competition.

uate the products they buy and sell, the answer to the above question is "yes." For instance, suppose the SRM were more profitable for MBS investors (i.e. Equation 14 does not hold), but some originator sought to originate FRMs instead to maintain the possibility of future fees from refinancing. This strategy would not work well because the securitizers would not want the low-yield mortgages, since their customers (the MBS investors) do not want low-yield securities. In that way, pressure from the MBS investors at the end of the pipeline requires intermediaries to act in their interest and produce mortgages with the best yield. This is competition at work. If we view the originators as agents of their MBS investor principals, there is not a clear conflict here that would cause a principal-agent problem, since MBS investors are able to see the interest rates being charged to borrowers and so can evaluate the originators' offerings appropriately.[28] Put differently, the originators have some flexibility to dictate terms to borrowers due to some borrowers' lack of sophistication; however, the originators' sophistication advantage disappears when they turn around and sell these mortgages to other savvy members of the financial sector. So while they can get borrowers to accept mortgages they do not "want," it is much harder to argue that they can get securitizers or MBS investors to accept mortgages they do not want. Given this conception of competition, it follows that the mortgage lending industry can indeed be viewed as seeking to maximize the value of the mortgages for the purposes at hand.

There are also implicit assumptions embedded in the model's refinancing process. First, the assumption maintained throughout the paper is that these costs, such as fees and time, are resource costs captured by nobody. This is a conservative assumption from the perspective of this paper. If, as is likely, some of borrowers' cost of refinancing is instead captured by lenders, that would mean that FRM refinancing is in fact associated with less deadweight loss than the full C^{Refi}, which makes the FRM more appealing to lenders. The assumption

[28]A problem does arise in the context of unobservable features of mortgages. For instance, Agarwal *et al.* (2012) showed that loans with high prepayment risk were more likely to be securitized prior to the financial crisis that began in 2007, while during the crisis the adverse selection shifted towards loans with high credit risk being more likely to be securitized. However, at issue in this paper is observable features of mortgage contracts: their design and their interest rates. As a result, such adverse selection is not relevant.

that the full cost is a resource cost thus pushes against the adoption of the FRM, the result the paper endeavors to explain.

Additionally, the assumption that borrowers can refinance by paying the resource cost C^{Refi} effectively assumes perfect competition among lenders during the refinancing process. This is internally consistent with the rest of the model if $C^{Search}_{Soph} = 0$, since only Sophisticated borrowers refinance. However, there is a more substantive restriction that borrowers cannot refinance FRMs into SRMs. If they could, they would, since eliminating subsequent refinancing effort would create surplus that lenders could capture. Without this restriction, the SRMs would emerge in a perverse equilibrium in which Unsophisticated borrowers are stuck with FRMs and Sophisticated borrowers eventually end up in SRMs when they first refinance. In fact, Sophisticated borrowers may want to refinance the instant after buying their home, effectively signaling to lenders their sophistication and, once passing that screen, receiving the SRM. In reality, this may be a dangerous strategy for lenders to pursue, because once the SRM has been introduced, it may be challenging to keep them away from Unsophisticated borrowers. Nonetheless, this does demonstrate the difficulty of avoiding the SRM, given the efficiency gains it generates.

It has also been assumed that borrowers who reject offers and reach the competitive phase will end up paying the cost of capital, $\frac{i_0+\mu}{\rho+\mu+\pi}$. While this may seem like the natural result of Bertrand competition, it does warrant further exploration. In particular, if the lenders offer an FRM that covers their cost of capital, it will in fact cost borrowers *more* than the cost of capital when refinancing costs are accounted for. Therefore, borrowers in this phase get the cost of capital only if the SRM is used in that phase. Note the irony: as this phase is not reached, the model succeeds in producing a no-SRM equilibrium, but it is underpinned by a "shadow SRM" that is never used but pins down this cost. If one is uncomfortable with an equilibrium in which the off-equilibrium path includes mortgages that no agent in the model has actually ever seen, we could instead require the mortgage in this phase to be FRMs, too. This would result in a worsening of the outside option for Sophisticated borrowers, allow

Parameter	Meaning	Value	Source
ρ	Discount rate	0.05	Agarwal *et al.* (2013)
π	Inflation rate	0.03	Agarwal *et al.* (2013)
μ	Hazard rate of move	0.1	Agarwal *et al.* (2013)
σ	Interest rate volatility	0.0109	Agarwal *et al.* (2013)
C^{Refi}	Refinancing cost	0.018	Agarwal *et al.* (2013) (for $250k mortgage)
S_{Soph}	Sophisticated share	0.55	Berger *et al.* (2024)
χ	Attention for Sophisticated	0.56	Berger *et al.* (2024)
C_{Soph}^{Search}	Sophisticated search cost	0	Generate $m_0^{FRM} = 0.06$,
i_0	Short-term rate at origination	0.0533	as in Agarwal *et al.* (2013)
S_{Mixed}	Share of non-Sophisticated with low search costs	0.34	Generates profit margin of 0.62%, as in MBA (2022)

Table 1: Parameters used in baseline calibration

lenders to extract more from Unsophisticated borrowers in the captive phase over and above their search cost, and therefore strengthen the appeal of the FRM to lenders. So assuming that borrowers can get the cost of capital in the competitive phase is another conservative assumption, pushing against the dominance of the FRM.

3 Quantitative Evaluation

Based on Proposition 1, the model's prediction that the SRM will be absent from the market requires Expression 14 to hold. This section shows that it does indeed hold when the model is calibrated using parameter values from the literature and that result is robust to relaxing a number of the model's assumptions, such as risk neutrality and a unit root for interest rates. The section then shows how mortgage design depends on the model's primitives around the baseline parameters.

3.1 Baseline Calibration

To assess whether Expression 14 holds (i.e. a FRM yields more revenue for lenders than a SRM), the model is calibrated primarily using off-the-shelf parameters from Agarwal *et al.* (2013) and Berger *et al.* (2024), as shown in the top panel of Table 1.[29]

[29] See page 602 of Agarwal *et al.* (2013) and Table C.2 of Berger *et al.* (2024).

The second panel of Table 1 shows parameters that ensure an initial FRM interest rate of 6% as used in Agarwal *et al.* (2013). In the model of Section 2, $m_0^{FRM} = i_0 + \upsilon_{Soph} + \omega_{Soph}^{FRM}$. While ω_{Soph}^{FRM} is pinned down (at 0.0067)[30] by the parameters listed in the top panel, i_0 and υ_{Soph} are not. The values of i_0 and C_{Soph}^{Seach} shown in the table are not the unique combination to generate $m_0^{FRM} = 6\%$, but their precise breakdown does not substantively impact the results. The only impact that these parameters have on the results is that C_{Soph}^{Seach} does affect the calibration of the final parameter in the table, S_{Mixed}, discussed below. The combination with $C_{Soph}^{Search} = 0$ is chosen to keep the model as close to the standard Perfect Competition benchmark as possible.

Before discussing the calibrated model's predictions, the two parameters governing heterogeneous refinancing should be considered in greater detail. S_{Soph} and χ, come from Table C.2 of Berger *et al.* (2024), who estimate the shares of borrowers in five bins of attention (χ), where both the values of attention and the shares are flexible. The distribution is chosen to best match refinancing activity in GSE-backed mortgages from 2005 through 2017. To translate their distribution to this paper's two-type setting, the four bins with $\chi > 0$ are averaged to get $\chi = 0.56$ for the Sophisticated type, while the bin with $\chi = 0$, with 45% of the population, is this paper's Unsophisticated type.

Figure 1 validates that the two-type model with these parameters is appropriate by simulating borrower behavior from 2019-2022. This was created using publicly-available loan-level data from Fannie Mae, described more in Internet Appendix D. For each loan active in December 2018, $y_{i,Dec2018}$ was calculated as the change in market interest rates from loan origination until December 2018. Refinancing activity was then simulated feeding in the path of interest rates through 2022, updating y_{it} each month for each borrower, and drawing attention shocks for the 55% of borrowers who are Sophisticated and moving shocks for all borrowers. The black line represents the quarterly refinancing share from that simulation.

[30]This is comparable to the premium for the option to refinance discussed in Lea and Sanders (2011), which was 0.005.

The red line shows the empirical quarterly refinancing share backed out from the Fannie Mae data. To do so, terminations were tracked in the loan-level data to get a quarterly prepayment share, and then the refinancing share is imputed as the difference between that prepayment share and the quarterly sale share reported by Batzer *et al.* (2024), which is also based on GSE mortgages. The appendix shows that while this approach assumes that all non-sale terminations are refinances, ignoring other events like full amortizations, the series constructed in this way has a very strong correspondence with aggregate refinancing data reported by the Federal Housing Finance Agency.

Figure 1's two series, data and model, rise in parallel through 2020 as interest rates plummeted before and during the COVID-19 pandemic; they then fall in parallel as interest rates rose following the onset of inflation after the pandemic. As interest rates peaked in 2022, refinancing in the model evaporates to essentially zero, while the data series remains positive, though low. This highlights the main difference between the two series: the data series is systematically higher by about 1pp. This likely reflects two factors. First, as mentioned above, the data series may over-count refinances by including full amortizations. Second, real borrowers in the data refinance for reasons not considered by the model, such as an improvement in credit score that allows them to lower their interest rate by even more than y, or equity extraction. As the model only considers refinancing incentives driven by changes in market interest rates, it should not be expected to capture such activity. Aside from that persistent level shift, which is attributable to issues not central to this paper, the patterns align very closely.

Now consider the results of the calibration. Under the parameters in Table 1, $D^{FRM}(\bar{m}^{FRM}_{Soph}, \chi) = 0.011$. In words, a Sophisticated borrower is expected to bear \$0.011 in refinancing costs for every dollar borrowed on a FRM, meaning a lender could offer a SRM that leaves the borrower indifferent while extracting an additional \$0.011 in revenue. This is the downside of the FRM from the lender's perspective. The benefit of the FRM comes from Unsophisticated borrowers. Under the parameters described above, an Unsophisticated borrower pays

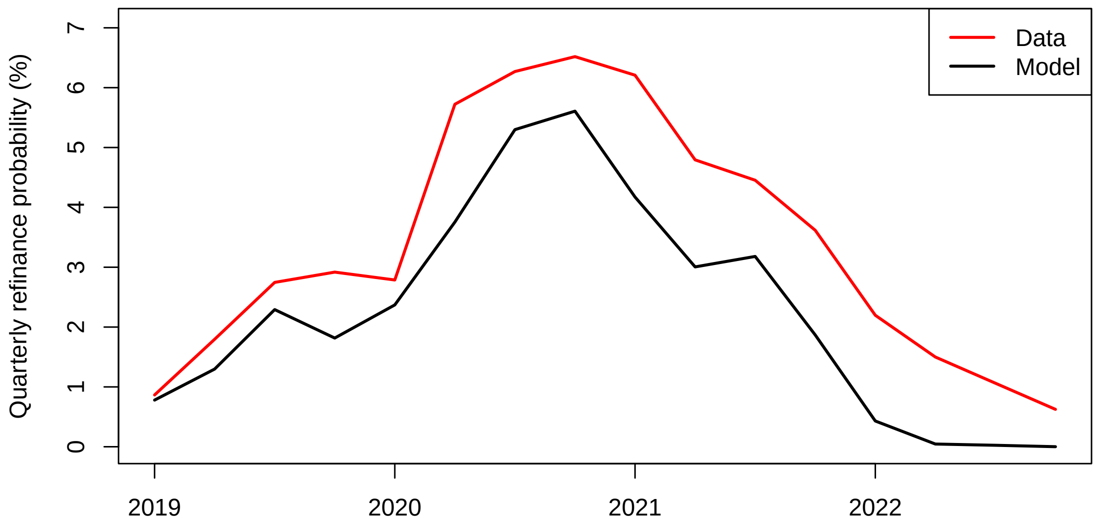


Figure 1: Comparison of model and data for refinancing activity for 2019-2022. "Data" represents the difference in quarterly prepayment shares in loan-level Fannie Mae data and sales shares from Batzer *et al.* (2024). "Model" shows a simulated series using Fannie Mae loans active in December 2018 and simulated through 2022 given the realized path of interest rates and the arrival of Poisson attention and moving shocks. Given the parameters, 45% of borrowers are assumed to have $\chi = 0$, while the remaining 55% have $\chi = 0.56$; given the other parameters, these borrowers have refinancing thresholds of $y^* = -0.0057$. Additional details of this analysis are available in Appendix D.

an additional \$0.037 on a FRM relative to a SRM. Given $S_{Soph} = 0.55$, Expression 14 holds: $(1-0.55)\cdot 0.037 > 0.55\cdot 0.011$. So, a FRM is more profitable than a SRM using an off-the-shelf calibration of the model.

One assumption made so far is that every Unsophisticated borrower – anyone who does not refinance – has high search costs during the origination process. This need not be the case, as some borrowers who do not pay attention to interest rates after origination (i.e. do not refinance) may nonetheless have low search costs when engaged in the process of buying a house and originating a mortgage. So suppose that a share $S_{Mixed} \cdot (1 - S_{Soph})$ of the population are such borrowers of "Mixed" sophistication with $\chi_{Mixed} = C^{Search}_{Mixed} = 0$ (low search costs, low attention). These borrowers will reject the FRM offer in the captive phase and then sign a FRM that generates zero profit and zero deadweight loss in the Bertrand phase.[31] With this added wrinkle, the FRM is more profitable than the SRM if and only if:

$$(1 - S_{Mixed}) \cdot (1 - S_{Soph}) \cdot \left(P^{FRM}(\bar{m}^{FRM}_{Soph}, 0) - P^{SRM}(\bar{m}^{SRM}_{Soph})\right) > S_{Soph} \cdot D^{FRM}(\chi), \quad (19)$$

which generalizes Expression 14 to allow $S_{Mixed} > 0$. Given the other parameters in the calibration, this will hold so long as $S_{Mixed} < 0.65$. In other words, even if lenders earn zero profit on 64% of non-Sophisticated borrowers because they negotiate low interest rates, the profit on the remaining 36% (who make up $36\% \cdot 45\% = 16\%$ of the entire population) is enough to make up for the deadweight loss caused by the 55% of the population that is Sophisticated (i.e. refinances), and therefore leads the FRM to be more profitable than the SRM. So 64% is an upper bound for S_{Mixed} such that the SRM is absent in equilibrium; moving forward, S_{Mixed} is set to 34%, as this leads to a profit margin of 0.62% observed in the industry, as indicated in Table 1. While S_{Mixed} is a free parameter to match a single moment, the calculations in this section still constitute a valid test of whether the model can explain the dominance of FRMs. The first calculation showed that with $S_{Mixed} = 0$, the FRM is more profitable than the SRM, which was not guaranteed. That was the primary test. But because that model's profitability is even greater than the observed 0.62%, $S_{Mixed} > 0$ is used to pull profit back down to that level.

In addition to Expression 14, the other requirement for the FRM-only equilibrium is that

[31] It will generate zero deadweight loss because these borrowers do not refinance. It will generate zero profit because it is being offered under conditions of Perfect Competition. Technically, Mixed borrowers could be offered a SRM during the Bertrand phase rather than a FRM, but it would not lead to a better payoff for them or the lender. Note that this is the first equilibrium discussed in which some offers are rejected. This occurs because lenders do not tailor the offers to Mixed borrowers because they are not profitable; those borrowers therefore reject the offers. I make the simplifying assumption that upon rejection, the borrower's type (in equilibrium, Mixed) is revealed. This assumption allows for lenders to offer the Mixed borrowers a zero-profit FRM in the Bertrand phase without the risk that it will attract Sophisticated borrowers, potentially scuttling the equilibrium.

$\bar{m}^{FRM}_{Uns} > \bar{m}^{FRM}_{Soph}$, i.e. $(\rho+\mu+\pi)\cdot(C^{Search}_{Uns} - C^{Search}_{Soph}) > \omega^{FRM}_{Soph}$. This ultimately hinges on there being sufficient dispersion in search costs between the borrower types. Internet Appendix E discusses the empirical literature on borrower search, which shows substantial dispersion among borrowers. Because the model is stylized with just two types, there is not a direct mapping from the data to the condition that $(\rho+\mu+\pi)\cdot(C^{Search}_{Uns} - C^{Search}_{Soph}) > \omega^{FRM}_{Soph}$. Having said that, given the calibrated value of $\omega^{FRM}_{Soph} = 0.0067$, the empirical dispersion in search costs does appear to be of a sufficient magnitude for the condition to hold.[32]

3.2 Robustness

Internet Appendix F presents results while allowing for mortgage amortization and the tax deductibility of mortgage payments and closing costs in ways proposed by Agarwal *et al.* (2013). The details are in that appendix, but suffice it to say here that those features do not have a large impact on the model's results.

Internet Appendix G presents results from a version of the model in which the cost of capital – and by extension mortgage interest rates – are mean-reverting. In particular, it uses an Ehrenfest chain parameterized to create an AR(1) process matching the mean, variance, and persistence of the interest rate process in Agarwal *et al.* (2022). With mean reversion, the wedge between the cost of capital and the mortgage interest rate is no longer constant. This means that the refinancing threshold, y^* is no longer a scalar: it depends on the level of the interest rate. The model finds the fixed point of refinancing rules and interest rates such that the rules are optimal given borrower inattention and the interest rates, and the interest rates generate costs for Sophisticated borrowers equal to their outside options. This is done separately for the FRM and the SRM. Taking the average across states in the stationary

[32] Since mortgage design in the baseline model is binary, search costs do not play any role in the model's predictions about mortgage design beyond determining whether $\bar{m}^{FRM}_{Uns} > \bar{m}^{FRM}_{Soph}$. That is why Table 1 does not provide a value for C^{Search}_{Uns}: since mortgage design in the baseline model is binary, the model does not distinguish between values of C^{Search}_{Uns} so long as the condition holds. Given that the empirical evidence discussed in the appendix suggests the condition is met, there is little further discussion of search costs in the analysis to follow.

distribution, the FRM extracts an additional \$0.024 per dollar lent to an Unsophisticated borrower and loses \$0.011 per dollar lent to a Sophisticated borrower, through the deadweight loss channel. Using the groups' respective weights, this leads to net profit of \$0.005. Those numbers were \$0.037, \$0.011, and \$0.011, respectively, in the random walk baseline presented above. Because $\$0.005 > 0$, this version of the model also predicts lenders should not offer the SRM. Interestingly, the results in the appendix show that as interest rates get low, the value of the FRM to lenders diminishes: this is because when borrowers start from a low rate, refinancing does not play a big role moving forward and so the FRM's role as a price discrimination tool declines accordingly (though so does the deadweight loss associated with it).

Internet Appendix H additionally considers a finite (30-year) horizon, risk aversion, and income risk. Income and the mortgage interest rate are modeled as AR(1) processes and calibrated following Agarwal *et al.* (2022), and household behavior is solved for numerically through backwards induction. Because of the size of the state space, the model take the interest rate process as given and does not solve for equilibrium interest rates; it is a partial equilibrium study of household behavior. It is used to shed light on the mortgage design question of this paper in the following way. For a given initial interest rate, m_0, we can find by Monte Carlo simulation $E[U^{FRM}_{Soph}(m_0)]$, the expected NPV of utility over 30 years for a Sophisticated FRM borrower, and $P^{FRM}(m_0, \chi_i)$, the expected NPV of mortgage payments received by lenders from that same contract, conditional on borrower attentiveness. I then solve for $\bar{m}^{SRM}(m_0)$, the interest rate on a SRM that would make the Sophisticated borrower indifferent between the FRM and SRM. Mortgage payments from such an SRM, $P^{SRM}(\bar{m}^{SRM}(m_0))$, can then be found via Monte Carlo simulation. In line with the logic of Expression 14, I then compare $(1 - S_{Soph}) \cdot \big(P^{FRM}(m_0, 0) - P^{SRM}(\bar{m}^{SRM}(m_0))\big)$ with $S_{Soph} \cdot \big(P^{SRM}(\bar{m}^{SRM}(m_0)) - P^{FRM}(m_0, \chi)\big)$, the former being lender losses on Unsophisticated borrowers from switching from FRM to SRM, and the latter being lender gains from

Sophisticated borrowers from such a switch.[33] If the former is larger than the latter, the lender prefers offering the FRM over the SRM.

The model finds a sizable increase in lender profit from the FRM relative to the SRM. Unlike in the baseline model, evaluating these expressions does depend on the initial interest rate, as well as the wealth and income of the household, so the result is harder to summarize compactly. As an example, if $m_0 = 6\%$ and a household has liquid wealth of roughly \$60 thousand and income of roughly \$90 thousand, the FRM generates an additional \$0.044 per dollar of principal from an Unsophisticated borrower (as opposed to \$0.037 in the risk neutral model) and \$0.015 less from a Sophisticated borrower (as opposed to \$0.011). Given 55% of borrowers being Sophisticated, we get $(1 - 0.55) \cdot 0.044 > 0.55 \cdot 0.016$, indeed confirming that the model with risk aversion also finds the FRM to be more profitable. More detail and discussion is provided in the appendix.

3.3 Comparative Statics

To better understand the model and its quantitative implications, this subsection shows how it behaves under different parameter values. Figure 2 shows how the model's key outcomes depend on the cost of refinancing (C^{Refi}). Understanding the role played by this cost is critical for the extensions explored in Sections 4 and 5. As seen in Figure 2a, a higher cost of refinancing lowers the FRM interest rate as it lowers Sophisticated borrowers' inclination to refinance. Importantly, Figure 2b shows that Unsophisticated borrowers' payments are isomorphic to the interest rate, as they will be paying the interest rate for the entire life of the mortgage, since they do not refinance. Therefore, higher refinancing costs benefit the Unsophisticated borrower by reducing the interest rate that Sophisticated borrowers are willing to accept.[34] Figure 2c shows how the deadweight loss associated with Sophisticated

[33] In the baseline model with risk neutrality focused on this paper, lenders' gains from Sophisticated borrowers are exactly equal to the expected deadweight cost of refinancing, as shown in Expression 14. That no longer holds in the model with risk aversion, as that breaks the equivalence of costs and (dis)utility.

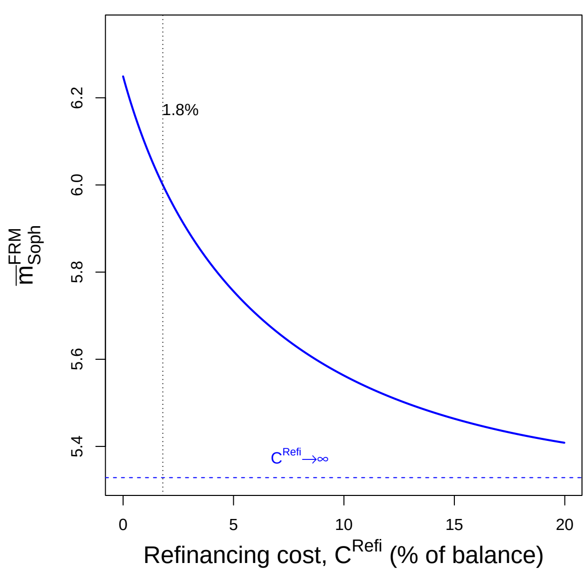
[34] Sophisticated borrowers are indifferent to where in the parameter space the model is, because their cost is constant – it is pinned down by their outside option of rejecting offers. This constant cost is in fact what

borrowers' refinancing varies. The relationship is non-monotonic because there are competing effects: conditional on the amount of refinancing that occurs, a higher refinancing cost of course leads to higher deadweight loss; however, higher refinancing costs lower the propensity of Sophisticated borrowers to refinance.

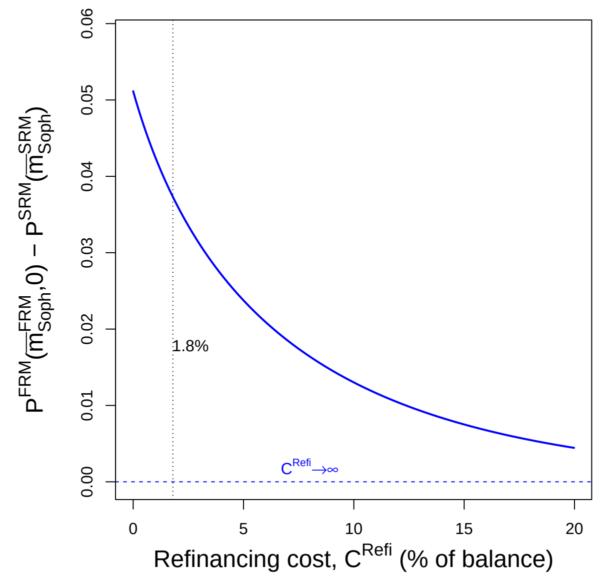
Figure 2d brings it all together, showing how lenders' rents depend on the refinancing cost. This combines two core effects: higher cost leads to a lower interest rate and therefore lower revenue from the Unsophisticated borrowers; the higher cost affects the deadeweight loss and therefore the amount that can be extracted from Sophisticated borrowers. While the ultimate effect on lender profit is ambiguous, in the neighborhood of 1.8%, there is the interesting result that lenders actually benefit when Sophisticated borrowers find refinancing to be low-cost. Lowering that cost simultaneously lowers deadweight loss and, more importantly, allows lenders to charge a higher interest rate and extract more heavily from Unsophisticated borrowers. Given the other parameters, the refinancing cost would have to more-than-double (from 1.8% to 4.0%) before the extractive benefits of the FRM would be outweighed by its deadweight loss and lenders would choose to offer the SRM instead.

Figure 3 shows how profit depends on some of the other key parameters. As shown in Figure 3a, interest rate volatility has an ambiguous effect: on the one hand, more volatility leads to more refinancing and more deadweight loss; on the other hand, it also allows for a higher interest rate and therefore more extraction from Unsophisticated borrowers. In the neighborhood of 0.0109, the latter effect dominates. This volatility would have to fall more than 50% to 0.0049 before the SRM would be more profitable than the FRM. Figure 3b shows that lender profit is increasing in χ at low levels of χ, meaning that lenders benefit when Sophisticated borrowers pay more attention to interest rates. This is because it allows them to charge a higher interest rate and extract more from Unsophisticated borrowers. However, at some point this relationship reverses, and the additional attention from Sophisticated borrowers generates sufficient deadweight loss to outweigh further gains from extracting

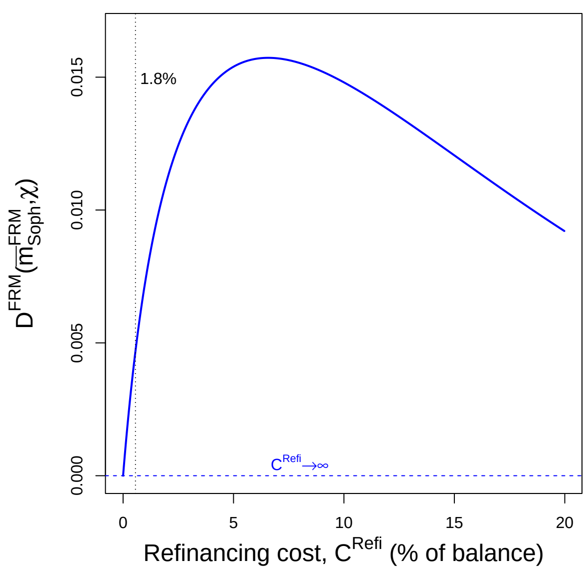
pins down the equilibrium interest rate.



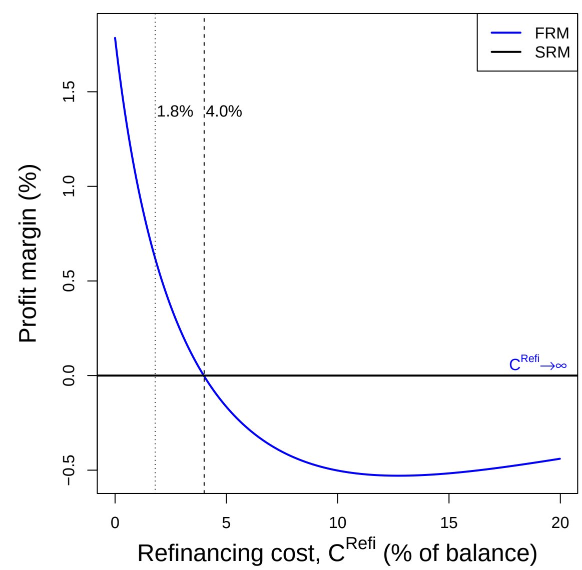
(a) FRM interest rate.



(b) Expected revenue extracted from Unsophisticated borrowers by FRM relative to SRM.



(c) Expected deadweight loss from Sophisticated borrowers refinancing FRM.



(d) Expected profit of FRM and SRM.

Figure 2: Comparative statics with respect to refinancing cost (C^{Refi}) of key model outputs using the calibration from Section 3.1. The baseline level of C^{Refi} (1.8%) is shown with a vertical line. The horizontal blue lines show the limiting values as $C^{Refi} \to \infty$. A second vertical line in Figure 2d shows the level of C^{Refi} at which the lender is indifferent between the FRM and SRM.

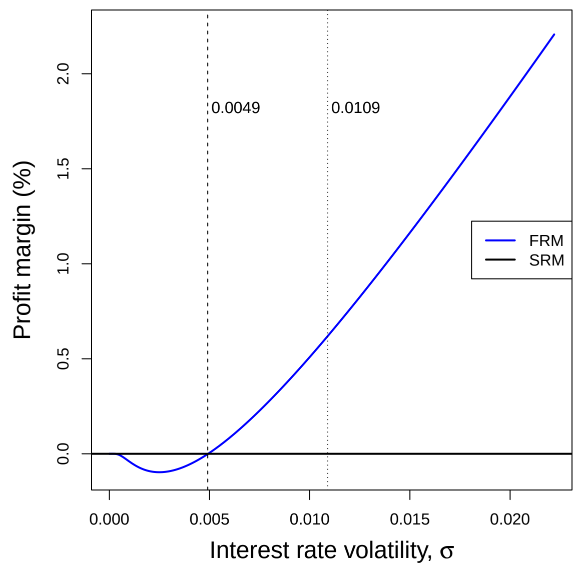
from Unsophisticated borrowers. Given the other parameters, there is no value of χ such that the SRM is preferable to the FRM.

Figures 3c and 3d show the comparative statics with respect to the two key population shares of the model. Relative to the SRM, the FRM generates benefits for the lenders from Unsophisticated borrowers and costs from Sophisticated borrowers. Therefore, the profit margin is strictly decreasing in S_{Soph}; given the other parameters in Table 1, the FRM is more profitable than the SRM if $S_{Soph} < 70\%$. In a similar vein, S_{Mixed} represents the share of non-Sophisticated borrowers who do not generate any profit because they have low search costs during origination. Therefore, profit is strictly decreasing in S_{Mixed}. As discussed earlier, the FRM continues to be more profitable than the SRM so long as $S_{Mixed} < 65\%$, and the observed profit margin of 0.62% is matched when $S_{Mixed} = 34\%$.

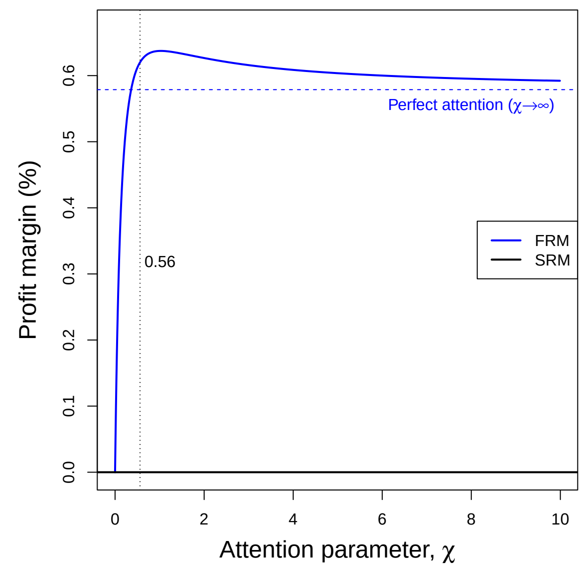
This exploration is not only revealing about how the model works in general, but it also shows that the conclusion that the FRM is better for lenders than the SRM is not particularly sensitive to the parameter choices, as it would require substantial changes in the parameters in order for the model to flip and predict that the SRM would be offered.

4 Expanding the Mortgage Space: Prepayment Penalties and Points

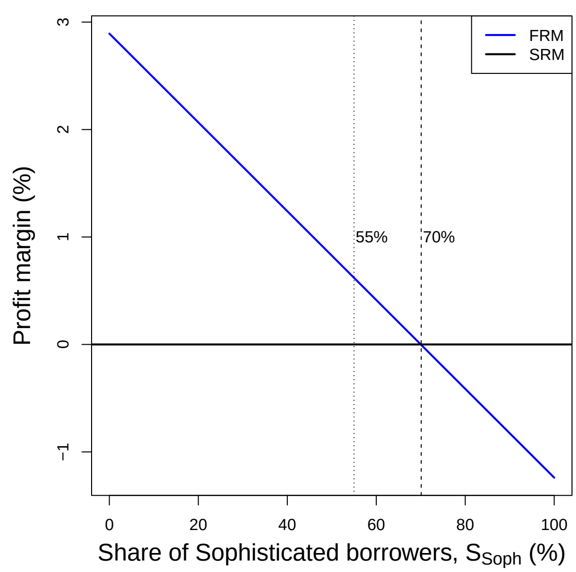
This section expands the set of features that lenders can include in a mortgage. In particular, it allows them to include prepayment penalties for a mortgage as well as offer a *menu* of interest rates, each accompanied by different levels of points (i.e. payments to or from the borrower) during mortgage origination. These features add subtlety to how lenders can navigate the paper's key tradeoff between efficiency and price discrimination, and they interact with the decision of whether to offer a SRM in important ways.



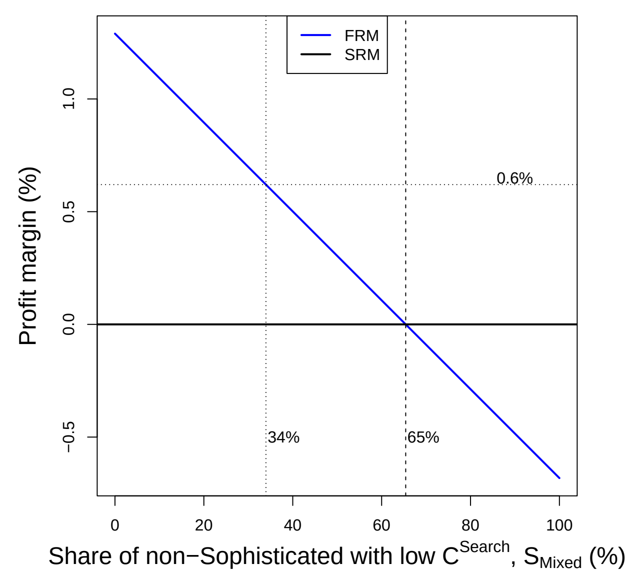
(a) Interest rate volatility.



(b) Attention parameter for Sophisticated borrowers.



(c) Sophisticated share of borrowers.



(d) Share of non-Sophisticated borrowers that have low search costs (i.e. Mixed borrowers).

Figure 3: Comparative statics of lender profit with respect to key model inputs using the calibration from Section 3.1. For each parameter, the baseline level from Table 1 is shown with a vertical line. The horizontal blue line in Figure 3b shows the limit as $\chi \to \infty$. With the exception of Figure 3b, a second vertical line shows the level of the given parameter at which the lender is indifferent between the FRM and SRM. The horizontal dashed line in Figure 3d shows how the calibrated value of $S_{Mixed} = 34\%$ was chosen, by generating a profit margin of 0.62%.

4.1 Prepayment Penalties

Consider a mortgage with a prepayment penalty, referred to here as a "PPM-p" (or "PPM," for short). A PPM-p is the same as a FRM, except the borrower must pay a penalty of p to the lender when she refinances it.[35] This PPM-p presents the lenders a continuum of choices, as they get to select their preferred $p \in [0, \infty)$, with the FRM and (as shown below) SRM being limiting cases.

4.1.1 FRM and SRM as Special Cases of the PPM-p

By definition, the FRM is equivalent to the PPM-0, which is a FRM with a prepayment penalty of 0. Less obvious is that the SRM is equivalent to $\lim\limits_{p\to\infty}$ PPM-p in this model. The two mortgages seem to be polar opposites, in that the SRM refinances without any effort, while the $\lim\limits_{p\to\infty}$ PPM-p can only be refinanced with effort and an arbitrarily large payment to the lender. However, they will generate the same *expected* payments, costs, and revenues for all parties.[36] To see why, note that the interest rate on the $\lim\limits_{p\to\infty}$ PPM-p will be $\lim\limits_{p\to\infty} m_0^{\text{PPM-}p} = i_0 + v_{Soph}$, which is the interest rate on a FRM with no premium for the option to refinance ($\omega_{Soph}^{FRM} = 0$). This will lead to expected (real) payments of $\frac{i_0+v_{Soph}+\mu}{\rho+\mu+\pi}$, which is precisely what the SRM generates.[37] This works out because the $\lim\limits_{p\to\infty}$ PPM-p has the SRM's two critical features: it generates no deadweight loss (because not even Sophisticated borrowers refinance) and it treats all borrowers equally (because, again, no one refinances).

So the lower limit of the PPM-p ($p = 0$) is a FRM and the upper limit ($p \to \infty$) is a SRM.

[35] Whether the penalty is also imposed when the borrower moves has no bearing on any outcomes of consequence. If the mortgage had only a "refinancing penalty," which were imposed upon a refinance but not a move, all outcomes would be the same. Imposing the penalty for moving simply leads to a lower interest rate and the same expected costs and revenues for all parties. Intuitively, this is because penalizing people for moving has no bearing on efficiency (as moves are exogenous) and no distributional impact (as all borrowers move at the same rate).

[36] This equivalence only holds in expectation. *Ex-post*, the path of payments will differ for the SRM — which will adjust as market rates change – and the $\lim\limits_{p\to\infty}$ PPM-p, which will maintain the interest rate agreed to at origination. If borrowers value mortgage payments differently at different times or different states of the world (e.g. due to risk aversion or liquidity constraints), that would break the equivalence. But in the risk-neutral, unconstrained setting of this paper, this convenient equivalence holds.

[37] Recall that $P^{SRM}(m_0^{SRM}) = \frac{m_0^{SRM}+\mu-1/\psi}{\rho+\mu+\pi}$ and $m_0^{SRM} = i_0 + v_{Soph} + 1/\psi$.

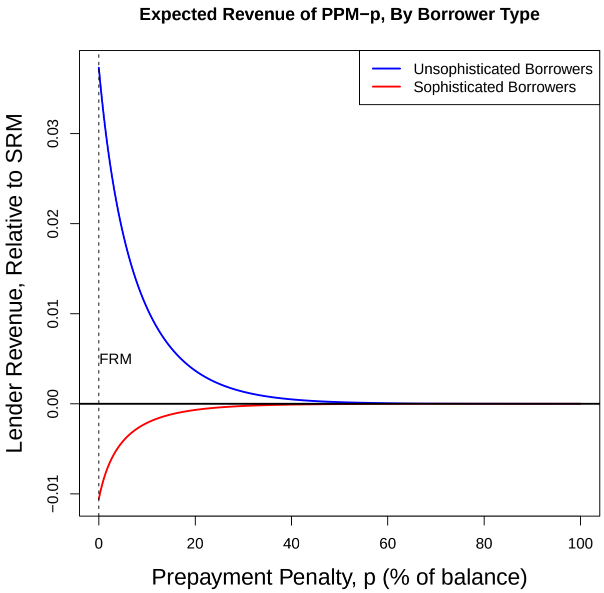
4.1.2 Equilibrium Mortgage Design When PPM-p Is Possible

A prepayment penalty affects mortgage design via the tradeoff explored throughout this paper. On the one hand, increasing the penalty, p, lowers deadweight loss by reducing refinancing activity, and as a result doing so allows the lender to extract more from the Sophisticated borrowers. On the other hand, the penalty is a burden on Sophisticated borrowers, so lenders have to lower the interest rate on the mortgage as they increase p. This leads to lower revenue on Unsophisticated borrowers, who then get a lower interest rate. Counter-intuitively, then, prepayment penalties are a partial solution to the problem caused by FRMs: the PPM lowers both inefficiency and the disparity between Sophisticated and Unsophisticated borrowers.

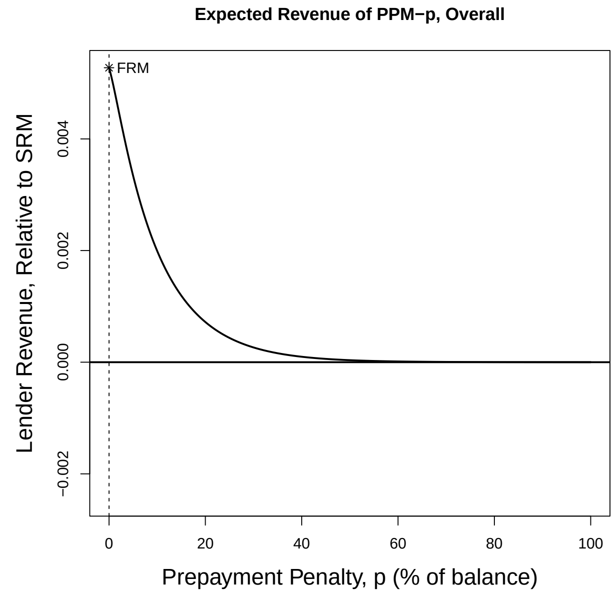
The PPM-p therefore generalizes all of the discussion above: switching from an FRM to a SRM decreased lender revenue from Unsophisticated borrowers but increased it from Sophisticated borrowers, and an increase in p on a PPM-p – which is a movement away from the FRM and toward the SRM – has the same impact. In a sense, the SRM is the metaphorical "hatchet" that cuts out all deadweight loss, but at the cost (from the lenders' perspective) of foregone revenue from Unsophisticated borrowers. The PPM-p can serve as the metaphorical "scalpel," finding a better tradeoff by eliminating some deadweight loss without forfeiting too much revenue on Unsophisticated borrowers.

4.1.3 Baseline Calibration

Figure 4 shows the quantitative results of expanding the model to allow any PPM-p to be offered under the baseline calibration explored in Section 3.1. Figure 4a shows the tradeoff discussed above: as the penalty increases, the lender captures less from Unsophisticated borrowers (due to the lower interest rate) but more from Sophisticated borrowers (due to reduced deadweight loss). As the penalty gets large, refinancing dries up and the revenue from – and costs to – each type of borrower converge to the level generated by the SRM.



(a) Expected FRM revenue from a Sophisticated and Unsophisticated borrower relative to SRM, as a function of prepayment penalty, p.



(b) Expected overall FRM revenue relative to SRM, as a function of prepayment penalty, p.

Figure 4: Analysis of prepayment penalty using the calibration from Section 3.1. A borrower must pay p every time she refinances or moves.

Figure 4b shows that given this tradeoff, the profit-maximizing PPM-p is the PPM-0, or in other words the FRM. Therefore, the model continues to predict that the FRM should be the dominant mortgage form, even when we allow not just a SRM as an alternative, but the more general PPM-p as well. This is consistent with the very small share of mortgages with prepayment penalties observed in recent decades, a point expanded on below.[38]

[38]The Dodd-Frank Act's restrictions on prepayment penalties of course provide another explanation for why prepayment penalties are not a common feature of present-day mortgages. However, those restrictions were only implemented in 2014, but prepayment penalties declined from being a common feature of residential mortgages in the 1980s to being very uncommon even before the Dodd-Frank Act's implementation. This is discussed in more depth below. One recent setting in which prepayment penalties *were* common – in contrast to the general trend – was the subprime market in the early 2000s. As discussed in Section 1, Mayer *et al.* (2013) argue that this was a form of risk-sharing: the penalties lowered the equilibrium interest rate, and then borrowers who experienced a positive innovation in their creditworthiness paid the prepayment penalty to refinance into a non-subprime mortgage.

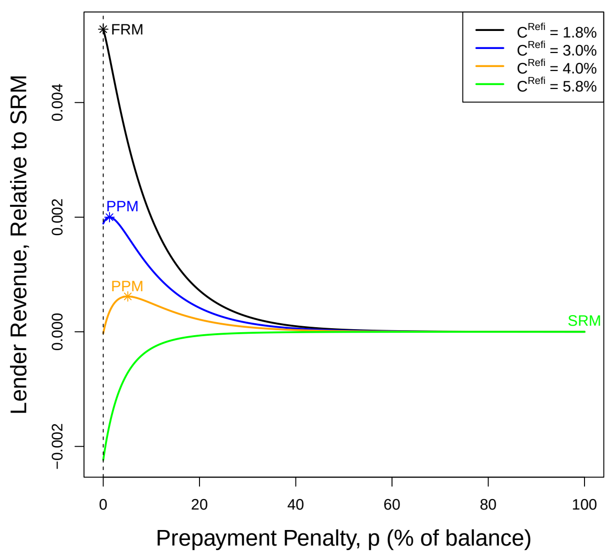
4.1.4 Comparative Statics

While the possibility of a PPM-p does not affect equilibrium mortgage design under the baseline calibration, it is interesting to explore how it affects the comparative statics of the model away from those baseline parameters. This will become more relevant in Section 5 when policy is discussed.

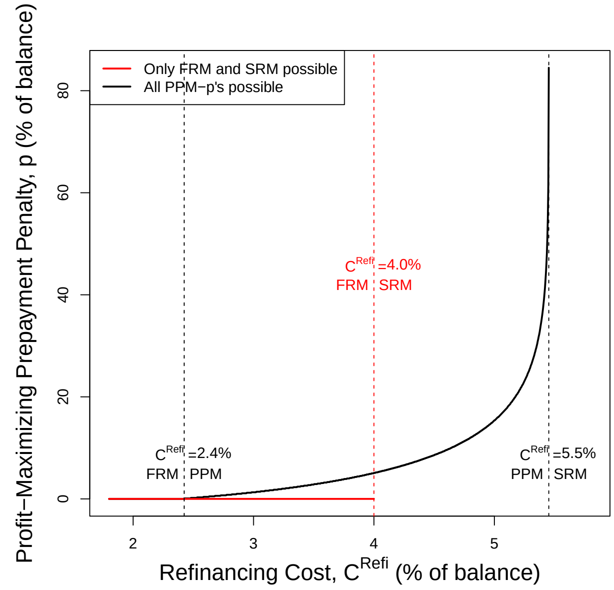
Recall that under the baseline parameters from Table 1, the model predicted that the SRM would be excluded if and only if $C^{Refi} < 4.0\%$. Figure 5 shows how this changes when the PPM is possible. Figure 5a shows lender revenue for PPMs for select levels of C^{Refi}. When $C^{Refi} = 1.8\%$ as in the baseline calibration, the optimal PPM-p has $p = 0$, as discussed above. The other lines show the progression of the mortgage from a FRM ($p = 0$) to a SRM ($p \to \infty$): as C^{Refi} increases, the deadweight burden of refinancing increases and so lenders can do better by suppressing refinancing activity with prepayment penalties. Interestingly, $C^{Refi} = 4.0\%$ no longer generates a SRM, as it did before. As the figure shows, while the FRM is not more profitable than the SRM in that case, there is a PPM that *is* better than the SRM, and so the SRM will not be offered.

Figure 5b generalizes this analysis. The red line shows that when only the FRM and SRM are possible, lenders transition from the FRM to the SRM when C_{Refi} reaches 4.0%, as discussed previously. The black curve shows how the possibility of a general PPM-p smooths the transition: the lenders abandon the FRM if $C^{Refi} > 2.4\%$ and impose prepayment penalties. Only when $C_{Refi} > 5.5\%$ is it profit-maximizing to implement the SRM.

As mentioned above, prepayment penalties have been quite rare in the prime mortgage sector in recent decades. Canner *et al.* (2002) note that prepayment penalties once were quite common but largely disappeared in the late 1980s, coinciding with falling interest rates and improving technology that lowered transaction costs, all of which encouraged refinancing activity. The results in Figure 5b add nuance to that narrative by allowing prepayment penalties to be something that is endogenously produced by the market rather than exoge-



(a) Expected FRM revenue relative to SRM as a function of prepayment penalty, p, for select levels of C^{Refi}.



(b) Equilibrium mortgage design as a function of C^{Refi}.

Figure 5: Analysis of prepayment penalty assuming different refinancing costs, C^{Refi}, using the calibration from Section 3.1. A borrower must pay p every time she refinances (or moves). The first black vertical line in Figure 5b shows the transition from a FRM to a PPM-p when C^{Refi} passes 2.4%; the second black vertical line shows the transition from PPM-p to SRM when C^{Refi} passes 5.5%. The red vertical line shows the transition from FRM to SRM when C^{Refi} passes 4.0%, assuming no other PPM-p is allowed. Based on the prior literature, the baseline calibration has $C^{Refi} = 1.8\%$.

nously imposed. The model can interpret the disappearance of prepayment penalties in the following way: lower-cost refinancing led sophisticated borrowers to value the refinancing option more highly, allowing lenders to charge a higher premium for it; as a result, they removed prepayment penalties, which had diminished the value of the option, all as a way to ratchet up the equilibrium interest rate, which allowed for greater extraction of unsophisticated borrowers. This is summarized in Figure 5b, as decreasing C^{Refi} leads to lower prepayment penalties and then, eventually, their disappearance (i.e. down-left movement along the black curve).

4.2 Points

Another way that lenders can expand the mortgage space to price discriminate is with points, a transfer between the borrower and lender at the time of closing accompanied by a corresponding change to the mortgage's interest rate. To model this phenomenon, suppose lenders can simultaneously offer two FRMs that differ along two dimensions: the interest rate, and an upfront payment. There is a "simple offer," $(m^{FRM-0}, 0)$, which is equivalent to the FRM discussed so far, and an offer with an additional transfer when the mortgage is closed, (m^{FRM-C}, C^{Close}). In particular, for the second offer, the borrower pays an additional C^{Close} to the lender when the mortgage is originated. Importantly, C^{Close} can be negative, in which case the lender is paying the borrower at closing.

Points can be part of equilibrium mortgage design because lenders can use them to extract additional profit. Note that in the "vanilla" FRM-only equilibrium described in Proposition 1 (i.e. when points are not allowed), the lender has almost surely not extracted all possible surplus from the Unsophisticated borrower.[39] Lenders can therefore increase profit with the following approach. They can target the simple mortgage with no points to the Unsophisticated borrower, setting $m^{FRM-0} = \bar{m}^{FRM}_{Uns}$, which will increase the surplus extracted from Unsophisticated borrowers by charging them their reservation interest rate, rather than $\bar{m}^{FRM}_{Soph} < \bar{m}^{FRM}_{Uns}$. Sophisticated borrowers will not accept this expensive mortgage,[40] so the lender can offer a separate FRM with $C^{Close} < 0$ and a high interest rate, m^{FRM-C}, that Sophisticated borrowers will accept (due to their relatively high refinancing likelihood) but Unsophisticated borrowers will not (since they will not refinance). Essentially, the lender is paying Sophisticated borrowers in order to increase contract interest rates, which heightens extraction from Unsophisticated borrowers.

Figure 6 shows how this plays out in the model calibrated above. The mechanics of the

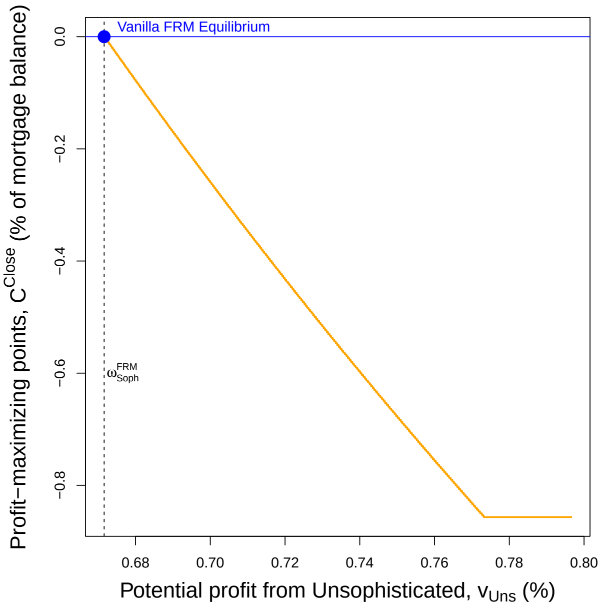
[39]Proposition 1 stipulates that C^{Search}_{Uns} is sufficiently high that $v_{Uns} \geq \omega^{FRM}_{Soph}$. Given this condition, the Unsophisticated borrower is paying her reservation interest rate only in the knife-edge case that $v_{Uns} = \omega^{FRM}_{Soph}$– otherwise, she is paying less.

[40]As throughout most of the paper, I continue to assume $\bar{m}^{FRM}_{Soph} < \bar{m}^{FRM}_{Uns}$.

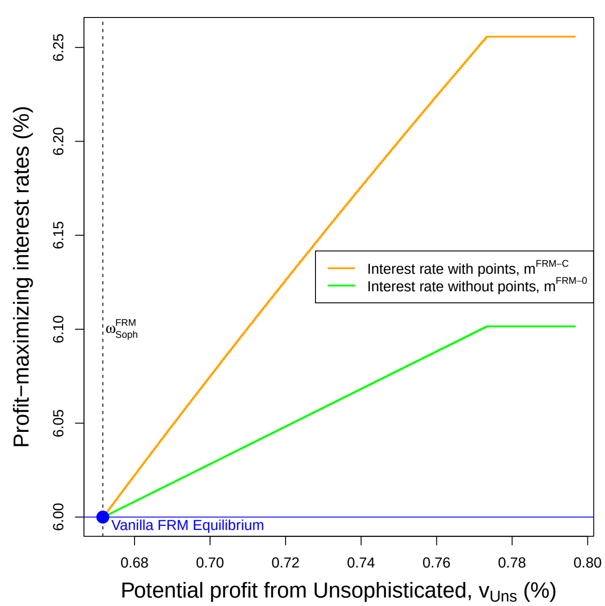
mortgage offers are shown in Figures 6a-6b. When $v_{Uns} = \omega_{Soph}^{FRM}$, the Unsophisticated borrower is paying her reservation interest rate and points cannot be used effectively. However, as v_{Uns} increases due to higher C_{Uns}^{Search}, the lender can extract more surplus by offering two mortgages with higher interest rates: one targeted to the Unsophisticated borrower with no points; and another targeted to the Sophisticated borrower with negative points and an even higher interest rate. As shown in Figure 6c, this allows the lender to modestly increase its profit margin relative to the baseline model with the "vanilla" FRM equilibrium described in Proposition 1. Note that the impact of points reaches a limit, even as v_{Uns} continues to grow. This occurs because the points system effectively subsidizes refinancing by Sophisticated borrowers and so increases deadweight loss, putting downward pressure on lender profits. Under the baseline calibration, further extraction is not practical once $C^{Close} = -0.85\%$, as it becomes too costly due to the refinancing costs it induces. Interestingly, the result that points of -0.85% of the mortgage balance are associated with an increase in the interest rate of roughly 15bp, as shown in Figure 6, is consistent with the real "rate sheet" shown in Figure 1 of Zhang (2024).[41]

The equilibrium here is similar to Stanton and Wallace (1998) and Zhang (2024), who show that a menu of interest rates and points can result from heterogeneous prepayment speeds. As Zhang (2024) has demonstrated, points exacerbate the efficiency and distributional impact of the FRM, leading to higher cost for unsophisticated borrowers and more refinancing by sophisticated ones. The core difference in this paper is that this redistribution from unsophisticated borrowers goes to lenders rather than sophisticated borrowers, as in the above papers, since lender profits in my model are not pinned to zero by Perfect Competition. More importantly for the purposes of this paper, this analysis shows that the introduction of points can further bolster the FRM as the mortgage design of choice for lenders over the

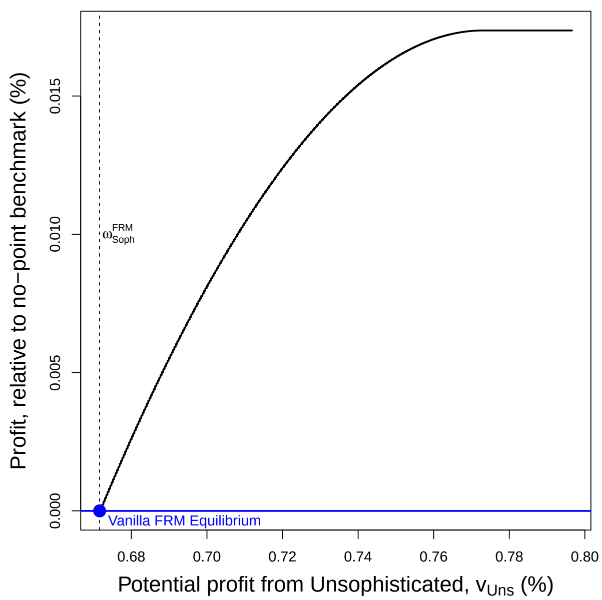
[41] The rate sheet he shows has three dimensions: interest rate, points, and lock period (i.e. how long the borrower can hold that interest rate before closing). In the rate sheet shown by Zhang (2024), the interest rate difference for mortgages with points of -0.569% and +0.238% (i.e. a difference of -0.807%) and a lock period of 15 days is 12.5bp. The magnitudes are nearly identical for the 30-day and 45-day lock periods.



(a) Equilibrium points as v_{Uns} varies.



(b) Equilibrium interest rates as v_{Uns} varies.



(c) Profit margin as v_{Uns} varies.

Figure 6: Analysis of points using the calibration from Section 3.1. Figures 6a-6b show the points and interest rates on a menu of two mortgages for different values of Unsophisticated borrowers' search costs, as parameterized by v_{Uns}: one mortgage has interest rate m^{FRM-C} and points of C^{Close}; the other has interest rate m^{FRM-0} with points equal to zero. Figure 6c shows the profit level associated with this approach. Values of v_{Uns} greater than ω_{Soph}^{FRM} are shown so that the conditions of Proposition 1 are met.

SRM. The seed of price discrimination within the vanilla FRM equilibrium can be amplified by points. Importantly, because the SRM treats all borrowers equally, all borrowers would have the same preference ordering across a menu of SRMs with different combinations of interest rates and points. As a result, introducing such a menu would not enable lenders to price discriminate any better than the single-SRM approach does, which is to say not at all.

5 Policy

Much of the existing literature studying the failure of unsophisticated borrowers to refinance FRMs discussed above has recommended the adoption of alternative mortgage structures, similar to SRMs, to address their overpayment for mortgage credit. This paper argues that the FRM dominates because lenders are able to increase their profit by offering it, and so we cannot expect alternative mortgage contracts to appear without being coaxed. This section considers how some interventions would affect mortgage design.

5.1 Refinancing Tax

To aid unsophisticated borrowers, the model of this paper recommends a policy that may be somewhat surprising: tax borrowers who actively refinance their mortgages.[42] (To be clear, if the interest rate falls automatically, as with an SRM, that would not be taxed.)[43] There would be three primary benefits of this policy.

First, it would reduce refinancing by Sophisticated borrowers and thus reduce deadweight loss. Refinancing is a form of rent-seeking, so curtailing it leads to an efficiency gain. A naive reaction could be that while there may be an efficiency gain, it would be undesirable from a distributional perspective because it prevents borrowers from recovering payments from lenders. However, this ignores the fact that reduced refinancing lowers the equilibrium FRM

[42] Such a tax is not unheard of: New York State has a "recording fee" 0.5% of the balance for mortgages, including refinances. Counties in the state can and do impose additional recording fees on borrowers.

[43] While the tax by itself might seem politically unpopular, it could be paired with a subsidy paid to all households to make it revenue neutral and increase its palatability.

interest rate. Sophisticated borrowers are in fact indifferent, as it is their indifference between the FRM and rejecting all offers that pins down the mortgage interest rate. Meanwhile, lenders capture the efficiency gain and the government collects tax revenue on the refinancing that still occurs, the incidence of which falls on the lenders.

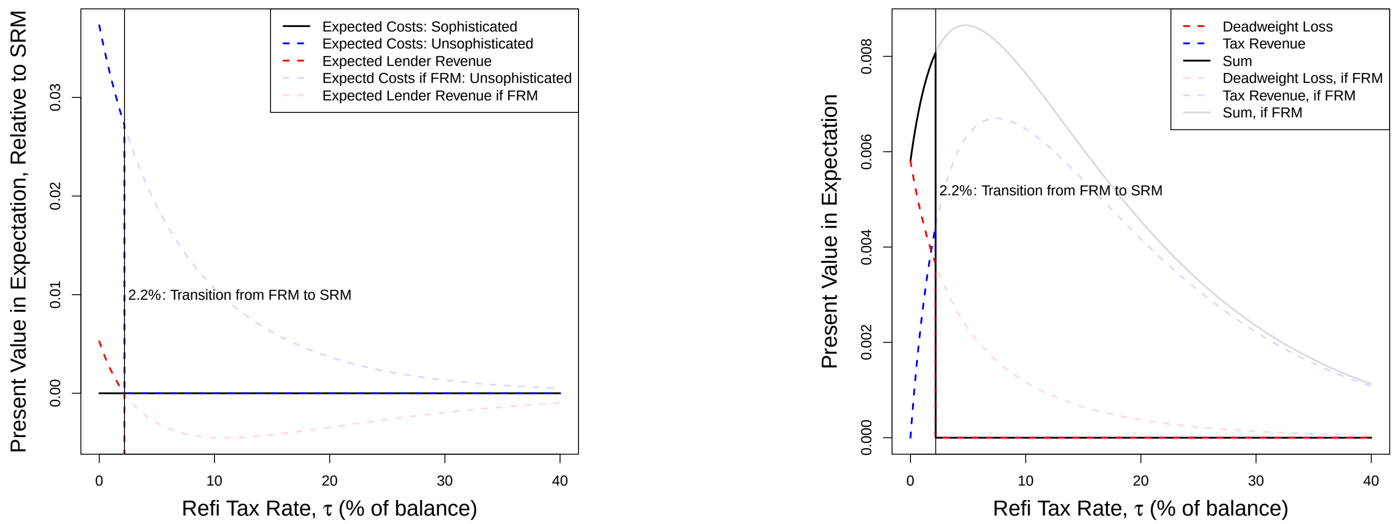
Second, the reduction in the interest rate is a windfall for Unsophisticated borrowers, who unambiguously benefit from the tax. They were not going to refinance anyway, so they are not directly impacted by the tax but gain indirectly due to the reduced interest rate. In all, then, the tax is essentially a transfer from lenders to Unsophisticated borrowers, as it lowers Sophisticated borrowers' tolerance for high interest rates and therefore reduces lenders' ability to use the FRM to extract from Unsophisticated borrowers. In addition to this transfer, there are efficiency gains that are captured by the government and lenders.

Finally, a tax of sufficient size can make refinancing so burdensome for the private sector that the equilibrium mortgage flips from a FRM to a SRM. As shown below, however, the possibility of prepayment penalties complicates this transition.

5.1.1 Refinancing Tax when PPM-p Is Not Possible

To begin, this subsection returns to assuming that only the FRM and SRM are possible; the implications of allowing other PPMs is discussed in the next subsection. Figure 7 demonstrates the impact of a refinancing tax using the calibration presented in Section 3.1. The tax is implemented by requiring the borrower to pay τ to the government each time she refinances. Figure 7a shows the main results. Imposing the tax lowers the costs of Unsophisticated borrowers and the profit of lenders as a result of the decline in the interest rate, which occurs due to Sophisticated borrowers' decreased willingness to refinance.[44] Figure 7b shows that the tax does lower deadweight loss, but this is initially swamped by the rise in govern-

[44]Sophisticated borrowers are indifferent to the level of the tax because it is their outside option of rejecting the offers and going to the Bertrand phase that pins down the mortgage interest rate. Therefore, the heightened burden of refinancing is exactly offset by the decline in the mortgage interest rate, which is why their welfare measure is constant.



(a) Borrower and lender outcomes.

(b) Deadweight loss and tax revenue.

Figure 7: Analysis of a refi tax using the calibration discussed in Section 3.1. A borrower must pay τ every time she refinances.

ment revenue, which further hurts lenders' profit. When the tax rate reaches $\tau^* = 2.2\%$, the FRM's ability to extract from Unsophisticated borrowers has become so weakened that it no longer generates higher profit than does the SRM. At this point, the model generates a change in mortgage design.

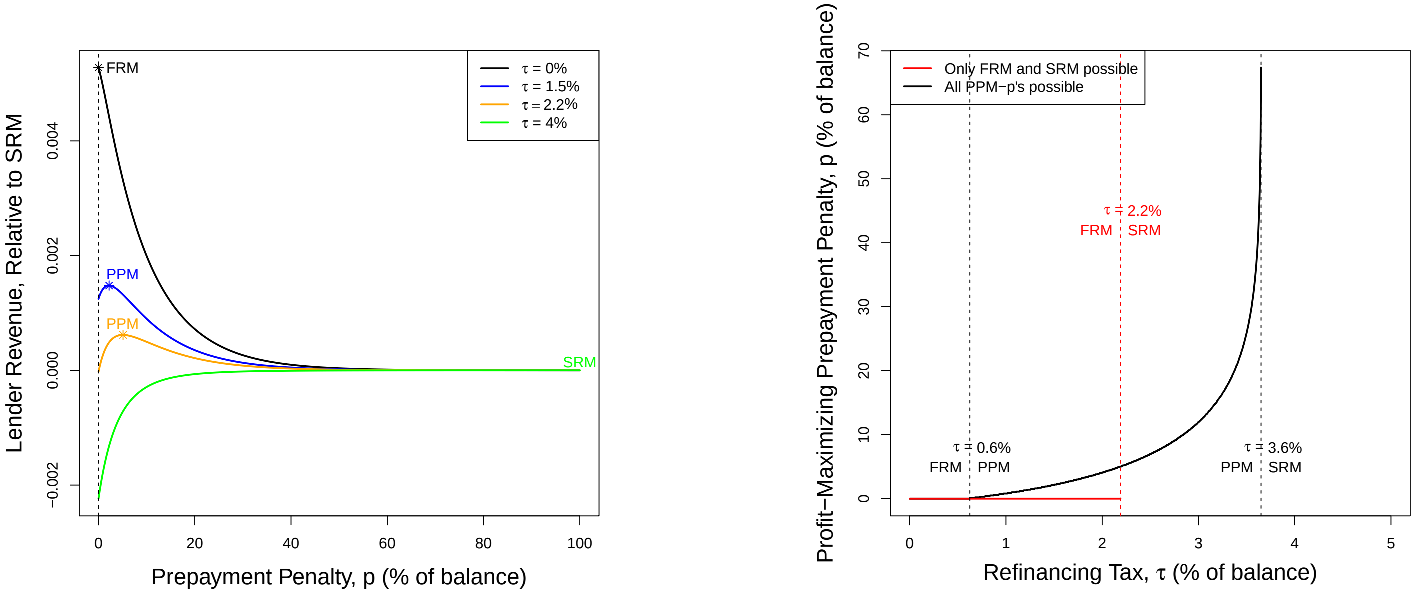
Given that we often think of Unsophisticated borrowers as not refinancing enough, it may be surprising that the correct policy is to tax refinancing. However, the real problem is that *Sophisticated borrowers refinance too much*, which allows lenders to extract extra profit from Unsophisticated borrowers, a price discrimination mechanism that requires generating deadweight loss. The refinancing tax remedies both of these problems, benefiting Unsophisticated borrowers and increasing overall social surplus.

It is fairly intuitive how a refinancing tax could lead to the desirable financial innovation predicted here: the introduction of SRMs. If borrowers are taxed for actively refinancing but not when their interest rate adjusts automatically, this will give an impetus for the market to avoid the tax by offering mortgages that refinance automatically. This form of tax avoidance is exactly what we want to see: a mortgage that refinances on its own and does not create disparities based on borrowers' financial sophistication. No longer able to effectively extract different payouts from different types of borrowers, lenders will be incentivized to offer the desirable (i.e. equitable and efficient) mortgage.

Suppose the policymaker set the tax to, say, 1.0% and so the market maintained the FRM; would this be a bad policy? It depends on the policymaker's goal. In terms of the main issues considered in this paper – eliminating inefficiency from refinancing effort and lowering costs for Unsophisticated borrowers – the policy would offer an improvement over the no-tax status quo. By lower refinancing, it would lower deadweight loss and allow for a lower equilibrium interest rate, to the benefit of Unsophisticated borrowers. On the other hand, if the policymaker was seeking to transition to the SRM for macroprudential reasons, so all borrowers would have mortgage payments that respond to monetary policy, then the policy would have backfired, as there would be less payment relief when interest rates decline, since there would be less refinancing. So while all of these goals are promoted by a SRM, a refinancing tax with no change in mortgage design puts the goals in opposition: a heavier tax helps from the perspectives of equality and efficiency but is harmful from a macroprudential perspective.

5.1.2 Refinancing Tax when PPM-p Is Possible

The introduction of the PPM-p affects the model's prediction of how policy would impact mortgage design. Figure 8a shows lender revenue as a function of the prepayment penalty for different levels of the refinancing tax. The orange line shows lender revenue when the tax is 2.2%, which as discussed right above, is the point at which the lender switched from offering the FRM to offering a SRM. However, now that a PPM-p is possible, the lenders would not offer the SRM; while the FRM would not deliver higher profit than the SRM, a PPM would. As a result, a policy that was intended to bring about a transition from FRM



(a) Expected FRM revenue relative to SRM as a function of prepayment penalty, p, for select levels of τ.

(b) Equilibrium mortgage design as a function of τ.

Figure 8: Analysis of prepayment penalty assuming different refinancing taxes, using the calibration from Section 3.1. A borrower must pay p to the lender every time she refinances (or moves), as well as τ to the government. The first black vertical line in Figure 8b shows the transition from a FRM to a PPM-p when τ passes 0.6%; the second black vertical line shows the transition from PPM-p to SRM when τ passes 3.6%. The red vertical line shows the transition from FRM to SRM when τ passes 2.2%, assuming no other PPM-p is allowed.

to SRM could instead lead to the adoption of a PPM. Figure 8b shows that – rather than a sharp transition from FRM to SRM at a tax of $\tau = 2.2\%$ – increasing the tax leads to a smooth transition away from the FRM. At a tax above 0.6%, the lenders would impose a prepayment penalty, which would increase until the tax rate exceeds 3.6%, at which point the lender would find the SRM to be profit-maximizing.

Note that the same issue discussed at the end of Section 5.1.1 is only reinforced here: if the tax leads to the introduction of a PPM-p rather than an SRM, the prepayment penalties will increase efficiency and equity, reinforcing the direct effect of the tax, while similarly reinforcing the negative macroprudential impact via further reduction in refinancing activity.

Finally, this analysis suggests a nuance to the impact of restrictions on prepayment penalties, like those enacted in the Dodd-Frank Act. The issue ultimately hinges on what would be offered in lieu of prepayment penalties. Consider the case of a refinancing tax less than 2.2% in this model's calibration. In that environment, a prepayment penalty ban would be harmful to both efficiency and equity because the equilibrium mortgage would be a FRM, which has more refinancing and higher interest rates. (As discussed above, the ban would still be helpful to the transmission of monetary policy, though.) On the other hand, if the tax were 3%, the ban on prepayment penalties would lead to an equilibrium SRM, promoting all goals: efficiency, equity, and macroprudence. So while the ban can be harmful if enacted alone, it can be beneficial if paired with a refinancing tax in a certain range.

5.2 Financial Education and Technology

A number of papers on this topic, such as Keys *et al.* (2016) and Bajo and Barbi (2018), have advocated for improving the population's financial literacy. The model of the present paper demonstrates that while such an intervention could yield the desired benefits, there is the possibility for undesired consequences.

Central to the evaluation of this policy is *who* is becoming more educated. To the extent that Sophisticated borrowers benefit from the education, which we can think of as a decrease in C^{Refi} or an increase in the attention parameter χ, this will actually hurt Unsophisticated borrowers. Sophisticated borrowers would value the refinancing option more and so lenders could charge a higher interest rate, increasing the extraction from Unsophisticated borrowers. Indeed, this is the reason that, in the neighborhood of the baseline calibration, lenders actually *benefit* from a decrease in C^{Refi} or an increase in χ, as shown in Figures 2d and 3b.

On the other hand, to the extent that such education reaches Unsophisticated borrowers and improves their attention, this will obviously benefit them. Interestingly, Figure 3c shows a feature of this policy that is not obvious, which emerges from allowing mortgage design to be endogenous. If we model increasing financial literacy among the Unsophisticated as an increase in S_{Soph}, then the policy can have benefits even to people who do not increase their own financial knowledge. In particular, under the baseline calibration, if the Sophisticated share were to rise from 55% to 70%, there would be an endogenous change in mortgage design towards the SRM, yielding benefits even for the 30% of borrowers who remain Unsophisticated. This suggests that perhaps financial education can have a bigger impact and be more cost-effective than it would otherwise appear if the FRM-dominant regime is taken as a given.[45]

Note that this also has implications for how we think about the positive and normative effect of the rise of "fintech" firms on the mortgage market. Fintech firms like RocketMortgage aggressively advertise mortgage refinancing opportunities and help borrowers go through the process. As with financial education, the model suggests this can help bring about the SRM and improve equity to the extent that it increases the share of sophisticated borrowers (increase in S_{Soph}), but the effects can be perverse – entrenching the FRM and harming equity – to the extent that it simply makes it easier for sophisticated borrowers to refinance (decrease in C^{Refi} or increase in χ). Given that many borrowers continue to ignore their refinancing incentive, despite the presence of such firms and their predecessors like LendingTree, the model provides a pessimistic view of the impact of these developments.

6 Conclusion

If one were to ask the proverbial "person on the street" why lenders do not offer a mortgage like the SRM that refinances automatically, they might think the answer is obvious: "Because lenders do not want borrowers to refinance!"

An economist would likely provide two critiques of this simple answer. First of all, because lenders compete for business, it does not matter what they want; they must provide what borrowers want, so long as it covers their costs. Indeed, as discussed in Section 2.3, in the limiting case of Perfect Competition, all borrowers will end up with the desirable SRM. The

[45] I thank my discussant, Noah Kouchekinia, for pointing this out.

second critique is that lenders do not mind in principle giving a mortgage that refinances automatically, as they can simply set a higher interest rate to compensate for the increased prepayment risk that they face.

This paper has presented a model that neutralizes those critiques and restores the explanation of the person on the street. By deviating from Perfect Competition (via search costs), the model allows for the possibility that lenders really can benefit from borrowers' failure to refinance, and borrower heterogeneity makes it so that the FRM is the effective way to do that, since the SRM treats all borrowers equally and so cannot price discriminate. Put differently, an economist might expect a lender's extraction from unsophisticated borrowers to be disciplined either by a competitor or by the need to appeal to sophisticated borrowers. Search costs limit the ability of competitors to interfere, while sophisticated borrowers are satisfied with the option to refinance – and hence we get observed mortgage designs.

This paper has not shown that borrower confusion or institutional inertia have no role to play in mortgage design. They likely do. But the paper has shown that an over-looked explanation – one that only depends on well-documented frictions in the mortgage origination and refinancing processes and in which borrowers and lenders behave optimally subject to those frictions – is quantitatively meaningful, likely enough so to prevent the offering of SRMs even in the absence of other explanations. Furthermore, as discussed in Section 1, while much of the exposition focused on the choice between FRMs and SRMs, the choice is really between mortgages that exploit unsophisticated borrowers and those that do not. In that way, a similar mechanism can explain the causes and consequences of the prevalence of ARMs with teaser rates in the UK, as studied by Fisher *et al.* (2024). The teaser rates succeed in extracting from the unsophisticated borrowers who fail to refinance upon their expiration. Arguably, that example is even more compelling, because while institutional inertia can potentially explain why lenders *do not add* automatic refinance provisions like in the SRM, it will have a harder time explaining why such ARMs *do add* a teaser rate provision. Such a story could potentially be at work in other settings where consumers must waste effort to get savings, as with credit card reward programs. This paper makes the point that rather than being a transfer from sophisticated to unsophisticated households, the products may be crafted by lenders to transfer from the latter to themselves.

Bibliography

Abel, J. and Fuster, A. (2021). How Do Mortgage Refinances Affect Debt, Default, and Spending? Evidence from HARP. *American Economic Journal: Macroeconomics*, **13** (2), 254–291.

Agarwal, S., Amromin, G., Chomsisengphet, S., Landvoigt, T., Piskorski, T., Seru, A. and Yao, V. (2022). Mortgage refinancing, consumer spending, and competition: Evidence from the home affordable refinance program. *The Review of Economic Studies*.

—, Ben-David, I. and Yao, V. (2017). Systemic Mistakes in the Mortgage Market and Lack of Financial Sophistication. *Journal of Financial Economics*, **123** (1), 42–58.

—, Chang, Y. and Yavas, A. (2012). Adverse selection in mortgage securitization. *Journal of Financial Economics*, **105** (3), 640 – 660.

—, Driscoll, J. C., Gabaix, X. and Laibson, D. (2009). The Age of Reason: Financial Decisions over the Life Cycle and Implications for Regulation. *Brookings Paper on Economic Activity*, **2**, 51–101.

—, — and Laibson, D. I. (2013). Optimal Mortgage Refinancing: A Closed-Form Solution. *Journal of Money, Credit, and Banking*, **45** (4), 591–622.

—, Grigsby, J., Hortacsu, A., Matvos, G., Seru, A. and Yao, V. (2024). Searching for Approval. *Econometrica*, **92** (4), 1195–231.

—, Rosen, R. J. and Yao, V. (2015). Why do borrowers make mortgage refinancing mistakes? *Management Science*, **62** (12), 3494–3509.

Andersen, S., Campbell, J. Y., Nielsen, K. N. and Ramadorai, T. (2020). Sources of inaction in household finance: Evidence from the danish mortgage market. *American Economic Review*, **110** (10), 3184–3230.

Bajo, E. and Barbi, M. (2018). Financial illiteracy and mortgage refinancing decisions. *Journal of Banking and Finance*, **94**, 279–296.

Batzer, R., Coste, J., Doerner, W. and Seiler, M. (2024). *The Lock-In Effect of Rising Mortgage Rates*. FHFA Staff Working Paper Series.

Benetton, M. (2021). Leverage regulation and market structure: A structural model of the u.k. mortgage market. *The Journal of Finance*, **76** (6), 2997–3053.

Berger, D., Milbradt, K., Tourre, F. and Vavra, J. (2024). *Refinancing Frictions, Mortgage Pricing and Redistribution*. Working Paper.

—, —, — and — (2025). Optimal mortgage refinancing with inattention. *American Economic Review: Insights*.

Bhagat, K. (2021). *Extending the Benefits of Mortgage Refinancing: The Case for the Auto-Refi Mortgage*. Working Paper.

Bhutta, N., Fuster, A. and Hizmo, A. (2025). Paying Too Much? Borrower Sophistication and Overpayment in the US Mortgage Market. *The Journal of Finance*.

Bjonnes, G., Kathitziotis, N. and Osler, C. (2024). *Cross-client Variation in OTC Execution Costs*. Working Paper.

Buchak, G., Matvos, G., Piskorski, T. and Seru, A. (2024). Beyond the balance sheet model of banking: Implications for bank regulation and monetary policy. *Journal of Political Economy*, **132** (2), 616–693.

Calvo, G. (1983). Staggered Prices in a Utility-Maximizing Framework. *Journal of Monetary Economics*, **12** (3), 383–398.

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

— (2013). Mortgage market design. *Review of Finance*, **17** (1), 1–33.

—, Nuno, C. and Cocco, J. F. (2021). Structuring Mortgages for Macroeconomic Stability. *The Journal of Finance*, **76** (5), 2525–76.

— and Ramadorai, T. (2025). *Fixed: Why Personal Finance Is Broken and How to Make It Work For Everyone*. Princeton: Princeton University Press.

Canner, G., Dynan, K. and Passmore, W. (2002). *Mortgage Refinancing in 2001 and Early 2002*. Federal Reserve Bulletin, Federal Reserve Board.

Deng, Y. and Quigley, J. M. (2007). *Irrational Borrowers and the Pricing of Residential Mortgages*. Working Paper.

Eberly, J. and Krishnamurthy, A. (2014). Efficient Credit Policies in a Housing Debt Crisis. *Brookings Papers on Economic Activity,*, **Fall 2014**.

Egan, M. (2019). Brokers versus retail investors: Conflicting interests and dominated products. *The Journal of Finance*, **74** (3), 1217–1260.

—, Ge, S. and Tang, J. (forthcoming). The market for financial adviser misconduct. *Review of Financial Studies*.

—, Hortacsu, A. and Matvos, G. (2017). Deposit Competition and Financial Fragility: Evidence from the US Banking Sector. *American Economic Review*, **107** (1), 169–216.

—, Matvos, G. and Seru, A. (2019). The market for financial adviser misconduct. *Journal of Political Economy*, **127** (1), 233–295.

Enkhbold, A. (2026). Monetary policy transmission, bank market power, and wholesale funding reliance. *Journal of Banking and Finance*, **187**.

Fisher, J., Gavazza, A., Liu, L., Ramadorai, T. and Tripathy, J. (2024). Refinancing cross-subsidies in the mortgage market. *Journal of Financial Economics*, **158**.

Fuster, A., Lo, S. and Willen, P. (2024). The Time-Varying Price of Financial Intermediation in the Mortgage Market. *The Journal of Finance*, **79** (4), 2553–2602.

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

Gerardi, K., Willen, P. S. and Zhang, D. H. (2023). Mortgage Prepayment, Race, and Monetary Policy. *Journal of Financial Economics*, **147** (3), 498–524.

Guren, A. M., Krishnamurthy, A. and McQuade, T. J. (2021). Mortgage Design in an Equilibrium Model of the Housing Market. *The Journal of Finance*, **76** (1).

Gurun, U. G., Matvos, G. and Seru, A. (2016). Advertising Expensive Mortgages. *The Journal of Finance*, **71** (5).

Hartman-Glaser, B. and Hebert, B. (2020). The insurance is the lemon: Failing to index contracts. *The Journal of Finance*, **75** (1), 463–506.

Jiang, E. (2023). Financing Competitors: Shadow Banks' Funding and Mortgage Market Competition. *Review of Financial Studies*, **36** (10), 3861–3905.

Johnson, E. J., Meier, S. and Toubia, O. (2018). What's the Catch? Suspicion of Bank Motives and Sluggish Refinancing. *Review of Financial Studies*, **32** (2), 467–495.

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

Kiefer, L., Kiefer, H. and Mayock, T. (2025). Transaction costs, mortgage rate heterogeneity, and the failure to refinance. *Real Estate Economics*, **53** (4), 701–735.

Lea, M. (2010). *International Comparison of Mortgage Product Offerings*. Special Report, Mortgage Bankers Association.

— and Sanders, A. (2011). Government Policy and the Fixed-Rate Mortgage. *Annual Review of Financial Economics*, **3**, 223–34.

Maturana, G. and Nickerson, J. (2019). Teachers Teaching Teachers: The Role of Workplace Peer Effects on Financial Decisions. *The Review of Financial Studies*, **32** (10), 3920–3957.

Mayer, C., Piskorski, T. and Tchistyi, A. (2013). The inefficiency of refinancing: Why prepayment penalties are good for risky borrowers. *Journal of Financial Economics*, **107**, 694–714.

MBA (2022). *IMB Production Profits Fell in 2021 from Record 2020*. Annual Mortgage Bankers Performance Report 2021, Mortgage Bankers Association.

Miles, D. (2004). *The U.K. Mortgage Market: Take a Longer-Term View, Interim Report: Information, Incentives, and Pricing*. HM Treasury Report.

Nelson, S. (2025). Private information and price regulation in the us credit card market. *Econometrica*.

Piskorksi, T. and Seru, A. (2018). Mortgage Market Design: Lessons from the Great Recession. *Brookings Papers on Economic Activity,*, **Spring 2018**.

Piskorski, T. and Tchistyi, A. (2011). Stochastic house appreciation and optimal mortgage lending. *The Review of Financial Studies*, **24** (5), 1407–46.

— and — (2020). *An Equilibrium Model of Housing and Mortgage Markets with State-Contingent Lending Contracts*. SSRN Working Paper.

Scharfstein, D. and Sunderam, A. (2017). *Market Power in Mortgage Lending and the Transmission of Monetary Policy*. Working Paper, Harvard University.

Stanton, R. and Wallace, N. (1998). Mortgage choice: What's the point? *Real Estate Economics*, **26** (2), 173–205.

Wang, Y., Whited, T. M., Wu, Y. and Xiao, K. (2022). Bank Market Power and Monetary Policy Transmission: Evidence from a Structural Estimation. *The Journal of Finance*, **77**, 2093–2141.

Woodward, S. E. and Hall, R. E. (2010). Consumer confusion in the mortgage market: Evidence of less than a perfectly transparent and competitive market. *American Economic Review: Papers and Proceedings*, **100** (2), 511–515.

— and — (2012). Diagnosing Consumer Confusion and Sub-optimal Shopping Effort: Theory and Mortgage-Market Evidence. *American Economic Review*, **102** (7), 3249–3276.

Zhang, D. (2024). *Closing Costs, Refinancing, and Inefficients in the Mortgage Market*. Working Paper.

Internet Appendix

A Two-Period Model

This section presents a simplified version of the model in Section 2 to highlight the key mechanisms. Rather than modeling the process of shopping for a mortgage followed by a continuous-time optimal stopping refinancing problem with attention shocks, this model exogenously assumes lender market power (subject to a participation constraint) in an initial homebuying period, followed by a single period in which refinancing might occur. This allows the following three key features of the model to be more prominent: 1) lender market power, subject to borrowers' participation constraints; 2) unobservable heterogeneous refinancing propensity among borrowers; 3) negative correlation between a borrower's refinancing propensity and the cost of its outside option. These are the ingredients that lead to the paper's key results on mortgage design: market powers can allow lenders to extract more from borrowers with worse outside options with the FRM, because those borrowers will fail to refinance.

A.1 Model Setup

Consider a lender who, in period 0, is matched to a homebuyer looking for a mortgage. The lender will offer a FRM at interest rate m^{FRM}, a SRM at interest rate m^{SRM}, or both. The homebuyer can accept an offer or reject all offers, resulting in an outside option of expected cost $1 + \kappa$. The borrower makes their choice to minimize their expected cost, so while the lender does have market power, they also face participation constraints for the borrowers.

Mortgages work in the following way. In period 1, a borrower pays back the mortgage with interest; there is no payment in period 0. If the borrower does not refinance, their payment is therefore $(1 + m^{FRM}) \cdot M$ or $(1 + m^{SRM}) \cdot M$, depending on the mortgage they chose. (M is the mortgage size, which henceforth will be normalized to $M = 1$.) However, borrowers can refinance in period 1 to change the initial interest rate they agreed to. If they have a FRM, this process has a resource cost of C^{Refi} per dollar of mortgage balance; if they have a SRM, it has cost zero. The alternative interest rate available in period 1 depends on the realization of lenders' cost of capital: $i \in \{i^L, i^H\}$, with $i^L < i^H$. Restrictions on i^L and i^H are discussed below, but for now simply assume that a borrower with a FRM wants to refinance if and only if $i = i^L$. Let π be the probability that $i = i^L$. For simplicity, assume that in period 1, competition among lenders allows borrowers who refinance to get a new interest rate of i, meaning the lender would earn zero profit.

When deciding what mortgage(s) to offer in period 0, the lender knows that there is probability S_{Soph} that the borrower is Sophisticated and probability $1 - S_{Soph}$ that they are Unsophisticated, though the borrower's type is not revealed. There are two differences between the types. First, Unsophisticated borrowers have worse outside options: $\kappa_{Soph} < \kappa_{Uns}$. Second, an Unsophisticated borrower does not refinance a FRM, even if $i = i^L$.

Finally, assume that the lender must get business from Sophisticated types. In the main body of the paper, this was expressed through a "name recognition" assumption that Unsophisticated borrowers will get matched to a lender only if it also attracts Sophisticated borrowers. This assumption rules out perverse equilibria in which Sophisticated borrowers are effectively ignored and lenders target only the Unsophisticated type, effectively erasing any heterogeneity in the setting.

A.2 Costs and Revenues

Define $K^{FRM}_{Soph}(m^{FRM})$ to be the expected cost faced by a Sophisticated borrower who agrees to a FRM with interest rate m^{FRM} in period 0. Then:

$$K^{FRM}_{Soph}(m^{FRM}) = 1 + (1 - \pi) \cdot m^{FRM} + \pi \cdot (i^L + C^{Refi}). \tag{20}$$

In this setting, m^{FRM} is a cap on what the borrower can be forced to pay, and there is value to the refinancing option because it allows for the possibility of paying $i^L + C^{Refi} < m^{FRM}$. In contrast, the Unsophisticated borrower does not refinance, meaning:

$$K^{FRM}_{Uns}(m^{FRM}) = 1 + m^{FRM}. \tag{21}$$

Unlike the FRM, the SRM refinances for all borrowers when $i = i^L$ and without cost, so:

$$K^{SRM}(m^{SRM}) = 1 + (1 - \pi) \cdot m^{SRM} + \pi \cdot i^L. \tag{22}$$

The lender's expected revenue differs from the borrower's expected cost only due to the resource cost of refinancing, which the borrower bears but is not captured by the lender. Defining $P^d_j(m)$ to be the NPV of payments received by the lender from a mortgage of type d to a borrower of type j with an interest rate of m, we have:

$$P^{FRM}_{Soph}(m^{FRM}) = K^{FRM}_{Soph}(m^{FRM}) - \pi \cdot C^{Refi} = 1 + (1 - \pi) \cdot m^{FRM} + \pi \cdot i^L; \tag{23}$$

$$P^{FRM}_{Uns}(m^{FRM}) = K^{FRM}_{Uns}(m^{FRM}) = 1 + m^{FRM}; \tag{24}$$

$$P^{SRM}(m^{SRM}) = K^{SRM}(m^{SRM}) = 1 + (1 - \pi) \cdot m^{SRM} + \pi \cdot i^L. \tag{25}$$

A.3 Reservation Interest Rates

Define $\bar{m}^{FRM}_i$ to be the highest interest rate that would make borrower of type i indifferent between the FRM and its outside option; $\bar{m}^{SRM}_i$ is defined analogously for the SRM. Therefore:

$$\bar{m}_{Soph}^{FRM} = \frac{\kappa_{Soph} - \pi \cdot (i^L + C^{Refi})}{1 - \pi}; \tag{26}$$

$$\bar{m}_{Uns}^{FRM} = \kappa_{Uns}; \tag{27}$$

$$\bar{m}_j^{SRM} = \frac{\kappa_j - \pi \cdot i^L}{1 - \pi}. \tag{28}$$

It is guaranteed that $\bar{m}_{Soph}^{SRM} < \bar{m}_{Uns}^{SRM}$; because the SRM leads to the same cost for both types at a given interest rate, the Unsophisticated borrowers' inferior outside option makes them more willing to accept a higher interest rate. On the other hand, it is ambiguous which of $\bar{m}_{Soph}^{FRM}$ and $\bar{m}_{Uns}^{FRM}$ will be bigger; while the inferior outside option pushes up Unsophisticated borrowers' reservation interest rate, the additional option value the FRM provides to Sophisticated borrowers (relative to Unsophisticated) pushes up Sophisticated borrowers' reservation interest rate.

With the reservation interest rates established, we can pin down the restriction on the distribution of i that is required. As discussed below, if the FRM is offered it will have an interest rate of $\bar{m}_{Soph}^{FRM}$. Given that, in order for a Sophisticated household with a FRM to refinance if and only if $i = i^L$, it must be the case that $i^L < \kappa_{Soph} + C^{Refi} < i^H$. That will be assumed throughout.

A.4 Equilibrium Mortgage Design

Consider the case in which $\bar{m}_{Soph}^{FRM} < \bar{m}_{Uns}^{FRM}$. In that case, if the lender offers only the FRM, it will do so at $\bar{m}_{Soph}^{FRM}$ and get expected revenue:

$$V^{FRM} = 1 + S_{Soph} \cdot \left((1 - \pi) \cdot \bar{m}_{Soph}^{FRM} + \pi \cdot i^L \right) + (1 - S_{Soph}) \cdot \bar{m}_{Soph}^{FRM}. \tag{29}$$

If instead the lender offers only the SRM, it will do so at $\bar{m}^{SRM}_{Soph}$ and get expected revenue:

$$V^{SRM} = 1 + (1 - \pi) \cdot \bar{m}^{SRM}_{Soph} + \pi \cdot i^L. \tag{30}$$

Comparing the two strategies, $V^{FRM} - V^{SRM}$ if and only if:

$$(1 - S_{Soph}) \cdot (\bar{m}^{FRM}_{Soph} - \kappa_{Soph}) > S_{Soph} \cdot \pi \cdot C^{Refi}. \tag{31}$$

This result is equivalent to Proposition 1 in the main body of the paper. The lefthand side of the inequality measures extraction from Unsophisticated types that the FRM generates that the SRM cannot. The margin that the SRM can extract is capped by κ_{Soph}: since the lender cannot distinguish the types, the Unsophisticated types benefit from the inexpensive outside option of the Sophisticated types. Offering the FRM instead allows them to get more, since $\bar{m}^{FRM}_{Soph} > \kappa_{Soph}$ – due to the Sophisticated types' ability to refinance the rate down – and the Unsophisticated type will simply pay that sticker price, since they will not refinance it. The righthand side shows the lost revenue from Sophisticated borrowers: the lender has to compensate the Sophisticated borrower for the resource cost of refinancing ($S_{Soph} \cdot \pi \cdot C^{Refi}$ in expectation), since that would be zero with the SRM. As in the main body of the paper, this shows that the FRM is a costly price discrimination tool, and it will be offered if and only if the price discrimination's extraction from Unsophisticated borrowers outweighs the costs imposed by active refinancing on the Sophisticated borrowers. Note that there is no profitable way to separate the types in this case: since the Unsophisticated borrower is paying $1 + \bar{m}^{FRM}_{Soph}$ and none of it is deadweight loss, the only way for the lender to earn more from the Unsophisticated type would be to raise their costs above $1 + \bar{m}^{FRM}_{Soph}$, which of course would not be acceptable to the borrower.

Now consider the case in which $\bar{m}^{FRM}_{Soph} > \bar{m}^{FRM}_{Uns}$. The lender can push both types to the limits imposed by their outside options by offering an FRM at $\bar{m}^{FRM}_{Soph}$ and an SRM at $\bar{m}^{SRM}_{Uns}$.

Sophisticated borrowers will not choose the SRM because its cost is above their outside option, and the same is true for Unsophisticated borrowers with the FRM. Essentially, the Sophisticated borrowers do not want to pay the high profit margin associated with the expensive SRM targeted at the Unsophisticated types, and the Unsophisticated types do not want to pay the high refinancing premium on the FRM, since they will not use it. This strategy yields revenue of:

$$V^{FRM\&SRM} = 1 + \pi \cdot i^L + (1-\pi) \cdot \Big(S_{Soph} \cdot \bar{m}^{FRM}_{Soph} + (1 - S_{Soph}) \cdot \bar{m}^{SRM}_{Uns} \Big). \tag{32}$$

An alternative, which does not charge everyone their reservation interest rate but maximizes efficiency, is to just offer the SRM at $\bar{m}^{SRM}_{Soph}$, which yields the same revenue as in the $\bar{m}^{FRM}_{Soph} < \bar{m}^{FRM}_{Uns}$ case. The former strategy is more profitable than the latter one if and only if:

$$(1 - S_{Soph}) \cdot (\kappa_{Uns} - \kappa_{Soph}) > S_{Soph} \cdot \pi \cdot C^{Refi}. \tag{33}$$

This is the same result as Proposition 2 in the main body of the paper. Once again, the mortgage design issue hinges on the benefit (to the lender) of increasing extraction from the Unsophisticated borrower ($\kappa_{Uns} > \kappa_{Soph}$) relative to the resource cost of the active refinancing by Sophisticated borrowers. So while the details of the situation are different (separating borrowers into different mortgages rather than pooling them in the FRM), the decision to use the FRM is governed by the same deep economics of trading off price discrimination with efficiency gains.

Finally, note that an FRM-only equilibrium cannot be profit-maximizing in this case since it would generate the same revenue from Sophisticated borrowers as the separating approach but less revenue from Unsophisticated types, since the separating approach extracts their full willingness-to-pay.

B Expected Payments to Lenders and Refinancing Costs

Define $K^{FRM}(m_0^{FRM}, \chi)$ to the expected costs at the time of origination of a borrower who pays attention to interest rates with a hazard rate of χ and agreed to an interest-only FRM (of size normalized to \$1) with interest rate m_0^{FRM}. The borrower can pay C^{Refi} in order to replace her current interest rate (m_0^{FRM}) with the market interest rate (m_t^{FRM}) – i.e. refinance. She discounts time at rate ρ, the inflation rate is π, she moves and pays back the balance at exogenous rate μ, and $dm_t^{FRM} = \sigma \cdot dz_t$, where z_t is a standard Brownian Motion.

To begin, it is helpful to solve for $K^{FRM}(m_0^{FRM}, 0)$, the expected costs of a borrower who is completely inattentive and never refinances (i.e. receives no benefit from the refinancing option). This borrower's costs come from two sources: interest payments, which are discounted at rate $\rho + \mu + \pi$; and the repayment of principal (\$1) at the time of a move, which has PDF $\mu \cdot exp(-\mu \cdot t)$; this payment is discounted at rate $\rho + \pi$. Therefore:

$$\begin{aligned} K^{FRM}(m_0^{FRM}, 0) &= \int_0^\infty m_0^{FRM} \cdot exp(-(\rho + \mu + \pi) \cdot t) \cdot dt + \int_0^\infty \mu \cdot exp(-\mu \cdot t) \cdot exp(-(\rho + \pi) \cdot t) \cdot dt \\ &= \frac{m_0^{FRM}}{\rho + \mu + \pi} + \frac{\mu}{\rho + \mu + \pi} \\ &= \frac{m_0^{FRM} + \mu}{\rho + \mu + \pi}. \end{aligned} \tag{34}$$

Now, define $y_t = m_t^{FRM} - m_0^{FRM}$. Denoting by $-\Omega^{FRM}(y_t, \chi)$ the option value of a borrower with attention parameter χ when the interest rate gap is y_t, we have:

$$K^{FRM}(m_0^{FRM}, \chi) = K^{FRM}(m_0^{FRM}, 0) + \Omega^{FRM}(0, \chi). \tag{35}$$

To solve for the option value, note:

$$\Omega^{FRM}(y,\chi) = \min_{y^*(\chi)} E\Big[\int_0^\infty exp(-(\rho+\mu+\pi)\cdot t)\cdot\big(C^{Refi} + \frac{y_{t-}}{\rho+\mu+\pi}\big)\cdot A_t\cdot dt\Big]$$
$$\text{s.t. } dy_t = \sigma\cdot dz_t - y_{t-}\cdot A_t, \tag{36}$$

where $y^*(\chi)$ is the optimal threshold such that a borrower refinances when paying attention and $y_t < y^*(\chi)$, A_t is an indicator variable marking that a refinance occurs at time t, and y_{t-} is the interest rat gap at time t, approaching from the left.

$\Omega^{FRM}(y,\chi)$ satisfies the Hamilton-Jacobi-Bellman equation:

$$\begin{aligned}
\rho\cdot\Omega^{FRM}(y,\chi) &= -(\mu+\pi)\cdot\Omega^{FRM}(y,\chi) + \frac{E[d\Omega^{FRM}(y,\chi)]}{dt}\\
(\rho+\mu+\pi)\cdot\Omega^{FRM}(y,\chi) &= \frac{\sigma^2}{2}\cdot\frac{\partial^2\Omega^{FRM}(y,\chi)}{\partial y^2} + \chi\cdot\min\big(0,\Omega^{FRM}(0,\chi) + C^{Refi} + \frac{y}{\rho+\mu+\pi} - \Omega^{FRM}(y,\chi)\big)\\
(\rho+\mu+\pi+\chi)\cdot\Omega^{FRM}(y,\chi) &= \frac{\sigma^2}{2}\cdot\frac{\partial^2\Omega^{FRM}(y,\chi)}{\partial y^2} + \chi\cdot\min\big(\Omega^{FRM}(y,\chi),\Omega^{FRM}(0,\chi) + C^{Refi} + \frac{y}{\rho+\mu+\pi}\big).
\end{aligned} \tag{37}$$

This differential equation is solved by:

$$\Omega^{FRM}(y,\chi) = \begin{cases} J_\psi\cdot exp\big(-\psi\cdot(y-y^*(\chi))\big) & \text{if } y\geq y^*(\chi)\\ J_\eta\cdot exp\big(\eta\cdot(y-y^*(\chi))\big) + \frac{\chi}{\rho+\mu+\pi+\chi}\cdot\big(J_\psi\cdot exp(\psi\cdot y^*(\chi)) + C^{Refi} + \frac{y}{\rho+\mu+\pi}\big) & \text{if } y\leq y^*(\chi)\end{cases} \tag{38}$$

Plugging this into Equation 37 yields $\psi = \frac{\sqrt{2\cdot(\rho+\mu+\pi)}}{\sigma}$ and $\eta = \frac{\sqrt{2\cdot(\rho+\mu+\pi+\chi)}}{\sigma}$.

The "Value Matching" and "Smooth Pasting" conditions that characterize $y^*(\chi)$ are:

$$J_\psi = J_\eta + \frac{\chi}{\rho+\mu+\pi+\chi}\cdot\big(J_\psi\cdot exp(\psi\cdot y^*(\chi)) + C^{Refi} + \frac{y^*(\chi)}{\rho+\mu+\pi}\big) \tag{39}$$

and

$$-\psi \cdot J_\psi = \eta \cdot J_\eta + \frac{\chi}{(\rho+\mu+\pi)\cdot(\rho+\mu+\pi+\chi)}. \tag{40}$$

This provides two equations in three unknowns (J_ψ, J_η, and $y^*(\chi)$). To solve the problem, finally note that when $y = y^*(\chi)$ and the borrower is paying attention, she must be indifferent between refinancing and not. As a result:

$$\begin{aligned} \Omega^{FRM}(y^*(\chi),\chi) &= \Omega^{FRM}(0,\chi) + C^{Refi} + \frac{y^*(\chi)}{\rho+\mu+\pi} \\ J_\psi &= J_\psi \cdot exp(\psi \cdot y^*(\chi)) + C^{Refi} + \frac{y^*(\chi)}{\rho+\mu+\pi}. \end{aligned} \tag{41}$$

Equations 39-41 are three equations in three unknowns, and as shown by Berger *et al.* (2025), there is a unique solution with $y^*(\chi) < 0$.

This allows us to solve for $\Omega^{FRM}(0,\chi)$:

$$\begin{aligned} \Omega^{FRM}(0,\chi) &= J_\psi \cdot exp(\psi \cdot y^*(\chi)) \\ &= \frac{\chi \cdot \left(\eta \cdot \left(\frac{y^*(\chi)}{\rho+\mu+\pi} + C^{Refi}\right) - \frac{1}{\rho+\mu+\pi}\right)}{(\rho+\mu+\pi+\chi)\cdot(\eta+\psi)\cdot exp(-\psi\cdot y^*(\chi)) - \chi\cdot\eta}. \end{aligned} \tag{42}$$

As in the main text, define $P^{FRM}(m_0^{FRM},\chi)$ and $D^{FRM}(\chi)$ to be, respectively, the payments to the lender and the deadweight refinancing costs that a borrower with attention parameter χ expects to pay at the time of origination of a FRM with interest rate m_0^{FRM}. Then:

$$K^{FRM}(m_0^{FRM},\chi) = P^{FRM}(m_0^{FRM},\chi) + D^{FRM}(\chi), \tag{43}$$

where, based on Equations 34, 35, and 42:

$$P^{FRM}(m_0^{FRM},\chi) = \frac{m_0^{FRM}+\mu}{\rho+\mu+\pi} + \frac{\chi\cdot\left(\eta\cdot\frac{y^*(\chi)}{\rho+\mu+\pi} - \frac{1}{\rho+\mu+\pi}\right)}{(\rho+\mu+\pi+\chi)\cdot(\eta+\psi)\cdot exp(-\psi\cdot y^*(\chi)) - \chi\cdot\eta} \quad (44)$$

and

$$D^{FRM}(\chi) = \frac{\chi\cdot\eta\cdot C^{Refi}}{(\rho+\mu+\pi+\chi)\cdot(\eta+\psi)\cdot exp(-\psi\cdot y^*(\chi)) - \chi\cdot\eta}. \quad (45)$$

As $y^*(\chi)$ is produced as the solution to Equations 39-41, Equations 44 and 45 allow us to calculate the payments to the lender and the deadweight loss expected by borrowers as a function of the model's primitives.

The case of the SRM can be evaluated by letting $\chi\to\infty$ and setting $C^{Refi}=0$.

C Proofs of Propositions 1 and 2

C.1 Proposition 1

Assume, as stipulated in Proposition 1, that $\bar{m}_{Uns}^{FRM} > \bar{m}_{Soph}^{FRM}$.

Recalling that $P^{SRM}(m) = \Big(m+\mu-1/\psi\Big)\cdot\frac{1}{\rho+\mu+\pi}$, then under the assumption that m_t^{SRM} follows Brownian motion (justified subsequently), a SRM offered with initial interest rate of (i_t+1/ψ) will cover the lender's opportunity cost of capital:

$$P^{SRM}(i_t+1/\psi) = \frac{i_t+\mu}{\rho+\mu+\pi}.$$

Because $\bar{m}_{Uns}^{SRM} > \bar{m}_{Soph}^{SRM} = i_t + v_{Soph} + 1/\psi$ and because $P^{SRM}(m)$ is an increasing function of m, this implies that – at worst – lenders can offer an SRM with interest rate $\bar{m}_{Soph}^{SRM}$ that all borrowers will accept (because $\bar{m}_{Uns}^{SRM} > \bar{m}_{Soph}^{SRM}$) and will generate non-negative profit relative to the opportunity cost of capital, since $v_{Soph}\geq 0$. Such a choice will lead to expected profit of:

$$\Pi_{SRM} = P^{SRM}(\bar{m}^{SRM}_{Soph}) - \frac{i_t + \mu}{\rho + \mu + \pi} \geq 0.$$

Alternatively, the lender could offer only an FRM at $\bar{m}^{FRM}_{Soph}$, which all borrowers will accept, because $\bar{m}^{FRM}_{Uns} > \bar{m}^{FRM}_{Soph}$. That would lead to expected profit of:

$$\Pi_{FRM} = S_{Soph} \cdot P^{FRM}(\bar{m}^{FRM}_{Soph}, \chi) + (1 - S_{Soph}) \cdot P^{FRM}(\bar{m}^{FRM}_{Soph}, 0) - \frac{i_0 + \mu}{\rho + \mu + \pi}.$$

Therefore:

$$\begin{aligned}
\Pi_{FRM} - \Pi_{SRM} &= S_{Soph} \cdot \big(P^{FRM}(\bar{m}^{FRM}_{Soph}, \chi) - P^{SRM}(\bar{m}^{SRM}_{Soph})\big) \\
&+ (1 - S_{Soph}) \cdot \big(P^{FRM}(\bar{m}^{FRM}_{Soph}, 0) - P^{SRM}(\bar{m}^{SRM}_{Soph})\big) \\
&= S_{Soph} \cdot \big(K^{FRM}(\bar{m}^{FRM}_{Soph}, \chi) - D^{FRM}(\chi) - K^{SRM}(\bar{m}^{SRM}_{Soph})\big) \\
&+ (1 - S_{Soph}) \cdot \big(P^{FRM}(\bar{m}^{FRM}_{Soph}, 0) - P^{SRM}(\bar{m}^{SRM}_{Soph})\big) \\
&= -S_{Soph} \cdot D^{FRM}(\chi) \\
&+ (1 - S_{Soph}) \cdot \big(P^{FRM}(\bar{m}^{FRM}_{Soph}, 0) - P^{SRM}(\bar{m}^{SRM}_{Soph})\big),
\end{aligned}$$

where the final step follows from the fact that $\bar{m}^{FRM}_{Soph}$ and $\bar{m}^{SRM}_{Soph}$ are characterized by giving the same overall costs to the Sophisticated borrower (which is equal to the costs of rejecting offers in the capitve phase and getting competitive offers in the Bertrand phase).

Therefore, the profit from offering only the FRM is greater than that of only offering the SRM if and only if:

$$(1 - S_{Soph}) \cdot \big(P^{FRM}(\bar{m}^{FRM}_{Soph}, 0) - P^{SRM}(\bar{m}^{SRM}_{Soph})\big) > S_{Soph} \cdot D^{FRM}(\chi),$$

which is Expression 14 shown in Proposition 1.

Furthermore, note that profit cannot be increased by offering both the SRM and the FRM in order to separate the borrowers ex-ante. To start, note that such a scheme would involve having Sophisticated borrowers choose the FRM and Unsophisticated borrowers choose the SRM, since Unsophisticated borrowers have more to gain from the SRM due to their lack of attention. To see why this will not occur, first suppose that Expression 14 holds, so the candidate strategy is to offer only the FRM at $\bar{m}^{FRM}_{Soph}$. In order to attract the Unsophisticated borrower to the SRM, they would need to offer it with interest rate m^{SRM} such that $P^{SRM}(m^{SRM}) < P^{FRM}(\bar{m}^{FRM}_{Soph}, 0)$. But this would – by construction – extract less revenue from the Unsophisticated borrower and therefore decrease profits.

Similar logic justifies why the lender will not separate the borrowers ex-ante if Expression 14 does not hold, in which case the candidate strategy is to offer only the SRM at $\bar{m}^{SRM}_{Soph}$. In that case, offering a FRM that the Sophisticated borrowers would accept requires its interest rate to be at most $\bar{m}^{FRM}_{Soph}$. To get the Unsophisticated borrower to then accept a SRM, it would have to be offered with interest rate m^{SRM} such that $P^{SRM}(m^{SRM}) < P^{FRM}(\bar{m}^{FRM}_{Soph}, 0)$, as above. Such a scenario would result in, at most, the same revenue as the FRM-only approach, which amounts to a decrease in profits relative to the SRM-only approach because Expression 14 does not hold.

In symmetric equilibria in which all lenders offer only a FRM at $\bar{m}^{FRM}_{Soph}$ when Expression 14 holds and only a SRM at $\bar{m}^{SRM}_{Soph}$ when Expression 14 does not hold, it also would not make sense for lenders to deviate and make offers that are rejected by Sophisticated borrowers. In that case, they would lose not only the Sophisticated borrowers, but Unsophisticated borrowers as well, due to the assumption that Unsophisticated borrowers will get matched to the lender with the highest market share, if such a lender exists. Since the lender is making non-negative profit, as shown above, it would not want to lose all of its business.

So if Expression 14 holds, the FRM will be observed, with interest rate $m^{FRM}_t = i_t + \Delta^{FRM}$, where $\Delta^{FRM} = v_{Soph} + \omega^{FRM}_{Soph}$ is a constant wedge. Therefore, since i_t follows a Brownian motion, m^{FRM}_t will follow a Brownian motion, as maintained throughout the paper.

Similarly, if Expression 14 does not hold, the SRM will be observed, with interest rate $m^{SRM}_t = i_t + \Delta^{SRM}$, where $\Delta^{SRM} = v_{Soph} + \omega^{SRM}$ is a constant wedge. Therefore, since i_t follows a Brownian motion, m^{SRM}_t will follow a Brownian motion, as maintained throughout the paper.

C.2 Proposition 2

Assume, as stipulated in Proposition 2, that $\bar{m}^{FRM}_{Uns} < \bar{m}^{FRM}_{Soph}$.

Recalling that $P^{SRM}(m) = \Big(m + \mu - 1/\psi\Big) \cdot \frac{1}{\rho+\mu+\pi}$, then under the assumption that m^{SRM}_t follows Brownian motion (justified subsequently), a SRM offered with initial interest rate of $(i_t + 1/\psi)$ will cover the lender's opportunity cost of capital:

$$P^{SRM}(i_t + 1/\psi) = \frac{i_t + \mu}{\rho + \mu + \pi}.$$

Because $\bar{m}^{SRM}_{Uns} > \bar{m}^{SRM}_{Soph} = i_t + v_{Soph} + 1/\psi$ and because $P^{SRM}(m)$ is an increasing function of m, this implies that – at worst – lenders can offer an SRM with interest rate $\bar{m}^{SRM}_{Soph}$ that all borrowers will accept (because $\bar{m}^{SRM}_{Uns} > \bar{m}^{SRM}_{Soph}$) and will generate non-negative profit relative to the opportunity cost of capital, since $v_{Soph} \geq 0$. Such a choice will lead to expected profit of:

$$\Pi_{SRM} = P^{SRM}(\bar{m}^{SRM}_{Soph}) - \frac{i_t + \mu}{\rho + \mu + \pi} \geq 0.$$

Alternatively, the lender could offer an FRM at $\bar{m}^{FRM}_{Soph}$ and an SRM at $\bar{m}^{SRM}_{Uns}$. The Sophisticated borrower will find the FRM acceptable but the SRM unacceptable (because $\bar{m}^{SRM}_{Uns} > \bar{m}^{SRM}_{Soph}$); the Unsophisticated borrower will find the SRM acceptable but the FRM unacceptable (because $\bar{m}^{FRM}_{Uns} < \bar{m}^{FRM}_{Soph}$, as stipulated in Proposition 2). That would lead to expected profit of:

$$\Pi_{FRM\&SRM} = S_{Soph} \cdot P^{FRM}(\bar{m}^{FRM}_{Soph}, \chi) + (1 - S_{Soph}) \cdot P^{SRM}(\bar{m}^{SRM}_{Uns}) - \frac{i_0 + \mu}{\rho + \mu + \pi}.$$

Therefore:

$$\begin{aligned}
\Pi_{FRM\&SRM} - \Pi_{SRM} &= S_{Soph} \cdot \big(P^{FRM}(\bar{m}^{FRM}_{Soph}, \chi) - P^{SRM}(\bar{m}^{SRM}_{Soph})\big) \\
&+ (1 - S_{Soph}) \cdot \big(P^{SRM}(\bar{m}^{SRM}_{Uns}) - P^{SRM}(\bar{m}^{SRM}_{Soph})\big) \\
&= S_{Soph} \cdot \big(K^{FRM}(\bar{m}^{FRM}_{Soph}, \chi) - D^{FRM}(\chi) - K^{SRM}(\bar{m}^{SRM}_{Soph})\big) \\
&+ (1 - S_{Soph}) \cdot \big(P^{SRM}(\bar{m}^{SRM}_{Uns}) - P^{SRM}(\bar{m}^{SRM}_{Soph})\big) \\
&= -S_{Soph} \cdot D^{FRM}(\chi) \\
&+ (1 - S_{Soph}) \cdot \big(P^{SRM}(\bar{m}^{SRM}_{Uns}) - P^{SRM}(\bar{m}^{SRM}_{Soph})\big),
\end{aligned}$$

where the final step follows from the fact that $\bar{m}^{FRM}_{Soph}$ and $\bar{m}^{SRM}_{Soph}$ are characterized by giving the same overall costs to the Sophisticated borrower (which is equal to the costs of rejecting offers in the capitve phase and getting competitive offers in the Bertrand phase).

Therefore, the profit from offering both the FRM and the SRM is greater than that of only offering the SRM if and only if:

$$(1 - S_{Soph}) \cdot \big(P^{SRM}(\bar{m}^{SRM}_{Uns}) - P^{SRM}(\bar{m}^{SRM}_{Soph})\big) > S_{Soph} \cdot D^{FRM}(\chi),$$

which is Expression 15 shown in Proposition 2.

If Expression 15 holds and so the candidate equilibrium strategy is to offer both the FRM and SRM, the change in profit from switching to the FRM-only strategy is:

$$
\begin{aligned}
\Pi_{FRM} - \Pi_{FRM\&SRM} &= S_{Soph} \cdot \left(P^{FRM}(\bar{m}_{Uns}^{FRM}, \chi) - P^{FRM}(\bar{m}_{Soph}^{FRM}, \chi)\right) \\
&+ (1 - S_{Soph}) \cdot \left(P^{FRM}(\bar{m}_{Uns}^{FRM}, 0) - P^{SRM}(\bar{m}_{Uns}^{SRM})\right) \\
&= S_{Soph} \cdot \left(P^{FRM}(\bar{m}_{Uns}^{FRM}, \chi) - P^{FRM}(\bar{m}_{Soph}^{FRM}, \chi)\right) \\
&< 0,
\end{aligned}
$$

where the second equality comes from the definition of the reservation interest rates and the fact that Unsophisticated borrowers have zero deadweight loss; and the inequality comes from the fact that $\bar{m}_{Uns}^{FRM} < \bar{m}_{Soph}^{FRM}$. Therefore, if Expression 15 holds, an all-FRM equilibrium is not possible.

If Expression 15 does not hold and so the candidate equilibrium strategy is to offer only the SRM at $\bar{m}_{Soph}^{SRM}$, the change in profit from switching to the FRM-only strategy is:

$$
\begin{aligned}
\Pi_{FRM} - \Pi_{SRM} &= S_{Soph} \cdot \left(P^{FRM}(\bar{m}_{Uns}^{FRM}, \chi) - P^{SRM}(\bar{m}_{Soph}^{SRM})\right) \\
&+ (1 - S_{Soph}) \cdot \left(P^{FRM}(\bar{m}_{Uns}^{FRM}, 0) - P^{SRM}(\bar{m}_{Soph}^{SRM})\right) \\
&< -S_{Soph} \cdot D^{FRM}(\chi) + (1 - S_{Soph}) \cdot \left(P^{FRM}(\bar{m}_{Uns}^{FRM}, 0) - P^{SRM}(\bar{m}_{Soph}^{SRM})\right) \\
&= -S_{Soph} \cdot D^{FRM}(\chi) + (1 - S_{Soph}) \cdot \left(P^{SRM}(\bar{m}_{Uns}^{SRM}) - P^{SRM}(\bar{m}_{Soph}^{SRM})\right) \\
&< 0,
\end{aligned}
$$

where the first inequality comes from definition of reservation interest rates and the fact that $\bar{m}_{Uns}^{FRM} < \bar{m}_{Soph}^{FRM}$; the second equality comes from the definition of the reservation interest rates and the fact that Unsophisticated borrowers have zero deadweight loss; and the final inequality comes from the fact that Expression 15 does not hold. Therefore, if Expression 15 holds, an all-FRM equilibrium is not possible.

In symmetric equilibria in which all lenders offer a FRM at $\bar{m}_{Soph}^{FRM}$ and a SRM at $\bar{m}_{Uns}^{SRM}$

when Expression 15 holds and only a SRM at $\bar{m}^{SRM}_{Soph}$ when Expression 15 does not hold, it also would not make sense for lenders to deviate and make offers that are rejected by Sophisticated borrowers. In that case, they would lose not only the Sophisticated borrowers, but Unsophisticated borrowers as well, due to the assumption that Unsophisticated borrowers will get matched to the lender with the highest market share, if such a lender exists. Since the lender is making non-negative profit, as shown above, it would not want to lose all of its business.

So if Expression 15 holds, the FRM will be observed, with interest rate $m^{FRM}_t = i_t + \Delta^{FRM}$, where $\Delta^{FRM} = v_{Soph} + \omega^{FRM}_{Soph}$ is a constant wedge. Therefore, since i_t follows a Brownian motion, m^{FRM}_t will follow a Brownian motion, as maintained throughout the paper. Simultaneously, the SRM will be observed, with interest rate $m^{SRM}_t = i_t + \Delta^{SRM}$, where $\Delta^{SRM} = v_{Uns} + \omega^{SRM}$ is a constant wedge. Therefore, since i_t follows a Brownian motion, m^{SRM}_t will follow a Brownian motion, as maintained throughout the paper.

Similarly, if Expression 15 does not hold, the SRM will be observed, with interest rate $m^{SRM}_t = i_t + \Delta^{SRM}$, where $\Delta^{SRM} = v_{Soph} + \omega^{SRM}$ is a constant wedge. Therefore, since i_t follows a Brownian motion, m^{SRM}_t will follow a Brownian motion, as maintained throughout the paper.

D Loan-Level Data and Simulations

D.1 Data

Empirical refinancing activity is measured using publicly available loan-level origination and performance data from Fannie Mae's "Single-Family Fixed Rate Mortgage" dataset, which covers loans acquired by Fannie Mae since 2000, with originations going back to 1999.[46]

[46]This dataset tracks 30-year FRM Fannie Mae mortgages that are fully-amortizing and fully-documented. Fannie Mae makes additional restrictions that remove high-risk mortgages, such as removing loans with origination loan-to-value ratios greater than 97%. While some of these sample selection criteria are more restrictive than is ideal for this paper, note that the data is used here to measure refinancing activity, rather than mortgage default, making the loss of higher-risk loans less trou-

The loan-level data is used to construct a quarterly time series of mortgage terminations by identifying loans that terminate in the data with statuses "Prepaid," "Third Party Sale," "Short Sale," or "REO."

From there, the quarterly refinancing probability is calculated as the share that prepaid their mortgage (from the microdata) minus the share that sold their home as reported in Batzer *et al.* (2024).[47]. This makes the strong assumption that all mortgage terminations result either from sales or refinances, which is not true. Nonetheless, Figure 9 shows that the series of refinancing probabilities constructed in this way has a strong correlation with aggregate refinancing activity of GSE mortgages as reported by the Federal Housing Finance Agency (FHFA).[48]

The analysis relies on monthly data on mortgage interest rates[49] from 1999 through 2022, as shown in Figure 10.

D.2 Simulations

This subsection describes how the model was used to simulate refinancing behavior from 2019-2022, as shown in Figure 1. The simulation uses 13,023,863 loans from the Fannie Mae microdata described above that were originated no later than December 2018 and terminated no earlier than January 2019. For each such monthly origination cohort, c, from January 1999 through December 2018, its value of $y_{c,Dec2018}$ in December 2018 was calculated as

bling. The data can be found at https://capitalmarkets.fanniemae.com/credit-risk-transfer/single-family-credit-risk-transfer/fannie-mae-single-family-loan-performance-data.

[47]Using microdata on mortgages guaranteed by the GSEs, these authors linked mortgage terminations to deeds records to identify which are sales. Their series displays a very strong correlation with aggregate home sales data as reported by the National Association of Realtors (NAR), as shown in Figure 2 of their paper. I am grateful to the authors of Batzer *et al.* (2024) for making auxiliary materials from their analysis available. These can be found at https://www.fhfa.gov/research/papers/wp2403.

[48]In particular, two sets of reports produced by FHFA can be used to construct a time series of refinance volume of GSE mortgages: Refinance Reports, which provide the relevant data from 2009Q1 through 2019Q1; and Foreclosure Prevention, Refinance, and FPM Reports, which took the place of the Refinance Reports after March 2019. They are all available at https://www.fhfa.gov/reports?topic=fannie-mae-and-freddie-mac-reports.

[49]In particular, the average interest rate on a 30-year FRM as reported by Freddie Mac's Primary Mortgage Market Survey is used. This is available from the Federal Reserve Economic Database (FRED) under the name "MORTGAGE30US."

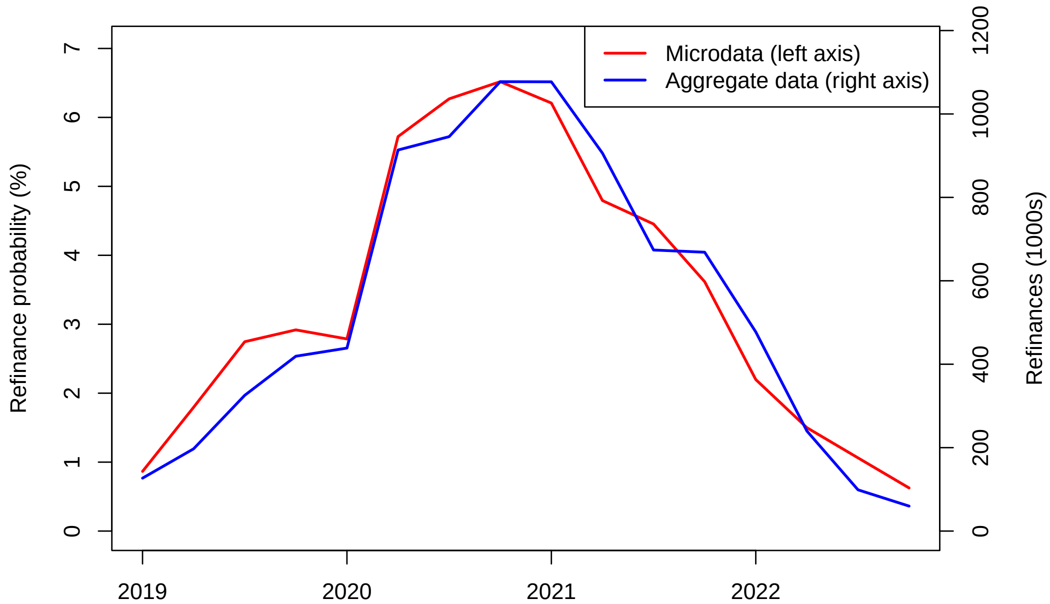


Figure 9: Validation of similarity between the probability of refinancing – calculated by combining microdata from Fannie Mae with the probability of sale from Batzer *et al.* (2024) – and aggregate refinancing data of GSE mortgages reported by the FHFA.

$m_{Dec2018} - m_{c_0}$, where $m_{Dec2018}$ is the mortgage interest rate in December 2018 as measured by the interest rate series described above, and m_{c_0} is the value of that series in the month of that cohort's origination.

Each cohort's refinancing behavior was then simulated analytically for 2019-2022, using the realized path of interest rates over that time period. This was done in the following way.

1. In each month, $y_{ct} = m_t - m_{c_0}$ was calculated and compared to y^*. Given the parameters used in the calibration, $y^* = -0.0057$.

2. The share of the cohort's remaining loans moving and refinancing in that month were computed using the probabilities below, with $\delta = 1/12$ to reflect the monthly frequency of the simulation:[50]

[50]The probability that neither of the two Poisson shocks – to attention and moving – occurs in a given month is $e^{-(\chi+\mu)\cdot\delta}$. If at least one of the shocks does arrive, the probability that the first shock that month is to attention is $\chi/(\chi+\mu)$.

Figure 10: Nominal mortgage interest rate. The vertical line marks the beginning of the simulation period.

$$S^{Refi} = \begin{cases} 0 & \text{if } y > y^* \\ \frac{\chi}{\chi+\mu} \cdot \left(1 - e^{-(\chi+\mu)\cdot\delta}\right) = 4.54\% & \text{if } y \leq y^* \end{cases}; \tag{46}$$

$$S^{Move} = \begin{cases} 1 - e^{-\mu\cdot\delta} = 0.83\% & \text{if } y > y^* \\ \frac{\mu}{\chi+\mu} \cdot \left(1 - e^{-(\chi+\mu)\cdot\delta}\right) = 0.81\% & \text{if } y \leq y^* \end{cases}. \tag{47}$$

3. Moves and refinances then form part of a new cohort in the month of that event. For instance, a new "January 2019" cohort forms, mixing together all movers and refinancers in that month initially from the January 1999-December 2018 cohorts. This cohort carries $m_{c_0} = m_{Jan2019}$ through the rest of the simulation.

4. The process is iterated for every month from January 2019 through December 2022.

5. The Sophisticated refinance share in any month is calculated as the weighted average of the cohorts' individual refinancing shares, where the weights are relative to the cohort

size.

6. The refinancing share reported in Figure 1 results from aggregating to the quarterly level and multiplying by 0.55, to account for the 45% of Unsophisticated borrowers, who do not ever pay attention.

E Evidence on Search Costs and Implications for Mortgage Design

In the model of Section 2, the existence of an equilibrium with no SRM requires $\bar{m}^{FRM}_{Uns} > \bar{m}^{FRM}_{Soph}$, which amounts to $v_{Uns} - v_{Soph} > \omega^{FRM}_{Soph} = 0.0067$, or 67 basis points ("bp"). Intuitively, Unsophisticated borrowers' search costs have to be high enough (relative to Sophisticated borrowers') that they are better off paying for the FRM's refinancing option (at cost ω^{FRM}_{Soph} in the interest rate)[51] even though they will not use it, rather than conducting a costly search for a better mortgage. The model predicts that if the condition is not met, lenders could increase profit by offering an SRM in addition to the FRM. A growing empirical literature is establishing that there indeed is substantial variation in search costs among borrowers, as this condition requires.

Woodward and Hall (2012) analyze compensation to mortgage brokers from a sample of 1,525 Federal Housing Administration mortgages originated in a short time window. They use quantile regressions to flexibly estimate distributions of broker compensation based on observables of the borrowers and find substantial variation, with some borrowers paying very high costs. They find a gap of about 3 percentage points ("pp") in the broker compensation paid by the 10^{th} and 90^{th} percentile borrowers, which is equivalent to roughly 75bp in the mortgage interest rate.[52] The authors interpret this variation as evidence that borrowers have greatly different effective costs of shopping across mortgage brokers. Their analysis shows that the distribution of costs shifts rightward for borrowers in Census tracts with low

[51] As noted earlier, Lea (2010) finds that the refinancing premimum is somewhat lower, at 50bp.

[52] See Table A-1 of Bhutta *et al.* (2025).

education levels, suggesting that financial sophistication and access is a key determinant of this cost. In that vein, they argue that this effective cost almost surely does not purely represent true effort costs but also borrower confusion.

Turning to findings on interest rates directly, Agarwal *et al.* (2024) look at loans originated from 2001-2011 and guaranteed by the GSEs and report that even conditioning on time, market, and borrower observables, the 90^{th}-10^{th} percentile gap in interest rates is 90bp. They point out that some borrowers may be willing to pay a higher interest rate not only due to low sophistication, per se, but also perhaps because they have a higher probability of having an application rejected and so rationally scale back their search. They find evidence that this is indeed responsible for some of the variation in interest rates borrowers receive. For the purposes of the analysis in the present paper, that distinction is immaterial, as either source of reduced search behavior provides lenders the ability to extract rents. Bhutta *et al.* (2025) use rich data on mortgage originations and offer sheets to evaluate the extent to which borrowers pay more than the price lenders are willing to transact at. They find that the 90^{th}-10^{th} percentile gap in interest rates agreed to by borrowers with the same observable characteristics (including points) is 55bp. Based on supplemental survey data, they show that more sophisticated borrowers search more and indeed get lower interest rates. They also find no evidence that higher interest rates are associated with better borrower experiences during the origination process, suggesting unsophisticated borrowers' higher interest rates result from lenders' rents. Gurun *et al.* (2016) study ARMs and show that conditioning on market, observables, and even initial interest rate, there is substantial variation in the reset rates that ARM borrowers receive: the 95^{th}-5^{th} percentile gap is 2.8pp. They show that this effect is more pronounced in products targeted at less sophisticated borrowers and in regions with more mortgage advertising, both suggestive of lenders extracting rents.

None of the empirical results discussed above is a direct analogue to $v_{Uns} - v_{Soph}$ in the model, largely because the model is quite stylized, with two types of borrowers and a simple mortgage shopping setting. Nonetheless, the evidence is indicative of variation in search costs of a magnitude on the order of what the model needs to generate an equilibrium in which no SRM is offered. A related but distinct point is that Unsophisticated borrowers may overestimate their future propensity to refinance and so underestimate the cost of a FRM. Effectively, this would mean $\omega_{Uns}^{FRM} > 0$, providing an alternative way for Unsophisticated borrowers to have a high reservation interest rate, beyond just search costs. In words, they are willing to accept a FRM not only because it is expensive for them to shop around, but also because they do not realize that they will not subsequently refinance it downward. In the extreme case where they believe themselves to have the same attention as a Sophisticated borrower, we would have $\omega_{Uns}^{FRM} = \omega_{Soph}^{FRM}$, guaranteeing $\bar{m}_{Uns}^{FRM} > \bar{m}_{Soph}^{FRM}$. While this is the extreme case, it makes the point that overconfidence by Unsophisticated borrowers can lessen the variation in search costs required for $\bar{m}_{Uns}^{FRM} > \bar{m}_{Soph}^{FRM}$ to hold. In a sense, the cost of learning and understanding the potential benefits of the search process – whether that be learning what other offers could exist or learning that one is not as sophisticated as she may believe – may play an important role beyond more easily identified "search costs" in causing $\bar{m}_{Uns}^{FRM} > \bar{m}_{Soph}^{FRM}$.

F Amortization and Tax Deductibility

As mentioned in Section 2, my assumption of an interest-only mortgage differs from the model of Agarwal *et al.* (2013) in a subtle way. They also assume that the mortgage balance stays constant until the arrival of an exogenous shock that forces full repayment. However, they assume that in addition to moving shocks, there is an additional type of "repayment event" that forces the borrower to pay off the mortgage, and they calibrate the arrival rate of this repayment shock to mimic the amortization of the mortgage. Effectively, then, they have assumed an interest-only mortgage in which the arrival rate of moving shocks depends on the mortgage rate, as that dictates the rate of amortization. I cannot follow them in making this assumption because if the arrival of moving shocks depended on the mortgage rate, Δ^{FRM} from Equation 2 would not be constant, meaning m_t^{FRM} would not follow a random walk, and the tractability of the setup would be lost. As such, I must formally assume the mortgage to be interest-only, rather than modeling the separate source of (mortgage rate-dependent) prepayments as Agarwal *et al.* (2013) do. Reassuringly, amortization is quite small relative to the repayments due to exogenous moving shocks in the calibrated model, and the quantitative results of my model are not sensitive to correspondingly small changes in μ, the arrival rate of moving shocks. As such, the assumption that mortgages are interest-only does not substantively impact the findings of the paper.

To show this, I follow Agarwal *et al.* (2013) in allowing the real value of the mortgage balance to decline at a constant rate, but I assume that this is rate is independent of the mortgage interest rate. For completeness, I also incorporate their approach to modeling tax deductibility of mortgage payments and closing costs into the refinancing problem.

As described in Appendix A of Agarwal *et al.* (2013), tax deductibility of mortgage payments and the partial deductibility of closing costs on a refinance can be modeled by defining a new cost of refinancing. Let ζ be the household's marginal tax rate, let T be the term of the refinanced mortgage, and let θ be the hazard rate of moving *or* refinancing; under the baseline calibration with no tax deductibility, we have $\theta \approx 0.21$,[53] though this is an approximation because the true hazard varies over time. Then, define:

$$C^{Refi}_{Deduc} = \frac{1}{1-\zeta} \cdot \left(C^{Refi} \cdot \left[1 - \frac{\zeta}{\theta+\rho+\pi} \cdot \left(\left[\frac{1-exp\big(-(\theta+\rho+\pi)\cdot T\big)}{T} \right] \cdot \left[\frac{\rho+\pi}{\theta+\rho+\pi} \right] + \theta \right) \right] \right). \tag{48}$$

Using this as the cost of refinancing (as opposed to simply C^{Refi}) incorporates both the fact that refinancing lowers the borrower's tax deduction by lowering mortgage payments to the lender (which produces the leading $\frac{1}{1-\zeta}$ term, which increases the effective cost of refinancing)

[53] To find an approximation of the hazard rate of refinancing, note that PDV of expected refinancing costs is $D^{FRM}(\chi)$. Assuming the refinancing hazard arrives at a constant rate θ_r, we have $D^{FRM}(\chi) \approx \int_0^\infty \theta_r \cdot C^{Refi} \cdot exp(-(\rho+\mu+\pi)\cdot t)dt$, so $\theta_r \approx (\rho+\mu+\pi) \cdot \frac{D^{FRM}(\chi)}{C^{Refi}}$. Under the baseline calibration, $\theta_r \approx 0.11$. Since the hazard of a move is $\mu = 0.10$, $\theta \approx 0.21$.

as well as the fact that the ability to spread a deduction of closing costs ($\zeta \cdot C^{Refi}$) over the remaining course of the mortgage (T years), lowering the effective cost of refinancing as captured by the bracketed term multiplying C^{Refi}. This C^{Refi}_{Deduc} can then simply replace C^{Refi}, and the results of the model go through. Following Agarwal *et al.* (2013), $\zeta = 0.28$ and $T = 30$.

With regards to amortization, Agarwal *et al.* (2013) assume that – in addition to inflation and the hazard of a moving shock – the expected real value of the mortgage balance declines at a rate of $\frac{m_0^{FRM}}{exp(\Gamma \cdot m_0^{FRM})-1}$, where Γ is the term of the contract, to approximate amortization. Since my amortization cannot technically be interest rate-dependent, I check the sensitivity of the results to allowing an additional decline the decline in mortgage value of $\frac{0.06}{exp(30 \cdot 0.06)-1} \approx 0.01$, in a way that is exogenous and not interest rate-dependent. All other parameters take on the same values as used throughout the paper, shown in Table 1.

When these adjustments are incorporated, the key outcomes are $D^{FRM}(\chi) = 0.010$ and $P^{FRM}(\bar{m}^{FRM}_{Soph}, 0) - P^{SRM}(\bar{m}^{SRM}_{Soph}) = 0.031$, as opposed to 0.010 and 0.037, respectively, in the main body of the paper. As a result, Expression 14 continues to hold: $(1-0.55) \cdot 0.031 > 0.55 \cdot 0.010$.

G Equilibrium Model with Mean-Reverting Interest Rate

To check the robustness of the core result (Expression 14 holding) to mean reversion in the interest rate, this appendix describes a solution approach and results from a model in which interest rates follow an AR(1) process, rather than a random walk.

G.1 Lenders' Cost of Capital

As in the body of the paper, let i_t be the instantaneous rate of return on lenders' outside option. It follows a continuous-time Markov chain on a finite set of states $i_1 < i_2 < ... < i_K$. Specifically, assume a continuous-time Ehrenfest chain with evenly-spaced states:

$$i_k = \mu_i + (2 \cdot k - K - 1) \cdot \frac{\sigma_i}{\sqrt{K-1}}, \tag{49}$$

where μ_i and σ_i are the mean and standard deviation of the i process, respectively. An Ehrenfest process can transition to adjacent regimes. The hazard rates of these transitions are:

$$f_{k,k-1} = c \cdot (k-1), \tag{50}$$

$$f_{k,k+1} = c \cdot (K-k), \tag{51}$$

where $c = -\frac{1}{2} \cdot ln(\rho_i)$, where ρ_i is the annual persistence of i.

Denote by κ_k the expected present cost to lenders of foregoing the short-term interest rate for the life of the mortgage (i.e. until the moving shock arrives), beginning in state k. Denoting $\delta \equiv \rho + \mu + \pi$ as the required rate of return, we have:

$$\underbrace{\delta \cdot \kappa_k}_{\text{required return}} = \underbrace{i_k + \mu}_{\text{flow}} + \underbrace{\sum_{l \neq k} f_{k,l} \cdot (\kappa_l - \kappa_k)}_{\text{expected capital gain}}. \tag{52}$$

In the random walk limit as $\rho_i \to 1$, this collapses to the $\frac{i_t+\mu}{\delta}$ used in the body of the paper.

G.2 Household Problem

Define $K^{FRM}(k, j; \chi)$ to be the expected value of costs starting in state k for a borrower with attention parameter χ that took out their most recent mortgage in state j. Therefore:

$$
\begin{aligned}
\delta \cdot K^{FRM}(k,j;\chi) = \underbrace{m_j^{FRM} + \mu}_{\text{flow}} + \underbrace{\sum_{l \neq k} f_{k,l} \cdot \left(K^{FRM}(l,j;\chi) - K^{FRM}(k,j;\chi)\right)}_{\text{continuation if no action}} \\
+ \underbrace{\chi \cdot \min\left\{K^{FRM}(k,k;\chi) + C^{Refi} - K^{FRM}(k,j;\chi), 0\right\}}_{\text{adjustment for optimal refinancing when attentive}}.
\end{aligned} \tag{53}
$$

To decompose this total cost into payments to the lender and deadweight resource costs, first define the refinancing rule that indicates whether an attentive borrower refinances:

$$
R(k,j;\chi) = \begin{cases} 1 & \text{if } K^{FRM}(k,j;\chi) > C^{Refi} + K^{FRM}(k,k;\chi) \\ 0 & \text{otherwise} \end{cases}. \tag{54}
$$

Therefore, the deadweight component of the total cost satisfies:

$$
\begin{aligned}
\delta \cdot D^{FRM}(k,j;\chi) = \sum_{l \neq k} f_{k,l} \cdot \left(D^{FRM}(l,j;\chi) - D^{FRM}(k,j;\chi)\right) \\
+ \chi \cdot R(k,j;\chi) \cdot \left[D^{FRM}(k,k;\chi) + C^{Refi} - D^{FRM}(k,j;\chi)\right].
\end{aligned} \tag{55}
$$

The lender receives $P^{FRM}(k,j;\chi) = K^{FRM}(k,j;\chi) - D^{FRM}(k,j;\chi)$. Note that for Unsophisticated borrowers with $\chi = 0$, there is no refinancing and no deadweight loss, resulting in:

$$
\delta \cdot P^{FRM}(k,j;0) = m_j^{FRM} + \mu. \tag{56}
$$

The SRM is more straightforward because it is type-independent, has no deadweight loss, and does not require an active refinancing decision. It satisfies:

$$\delta \cdot K^{SRM}(k,j) = m_j^{SRM} + \mu + \sum_{l \neq k} f_{k,l} \cdot \left(K^{SRM}\big(l, \min\{l,j\}\big) - K^{SRM}(k,j)\right), \tag{57}$$

and $P^{SRM}(k,j) = K^{SRM}(k,j)$.

G.3 Market Equilibrium

Following the calibration in the body of the paper, assume $C_{Soph}^{Search} = 0$, so a Sophisticated borrower's outside option has cost κ_k. Therefore, in equilibrium, m_k^{FRM} and m_k^{SRM} are pinned down by:

$$K^{FRM}(k,k;\chi_{Soph}) = \kappa_k = K^{SRM}(k,k). \tag{58}$$

Under the random walk used in the body of the paper, those conditions pin down the interest rates in closed form (conditional on $y^*(\chi_{Soph})$), but that is not the case with mean reversion in i. Because $m-i$ is no longer constant, refinancing no longer follows a scalar threshold rule and instead depends on the equilibrium determination of that gap. Therefore, the vectors of refinancing rules, $\{y^*(k,j;\chi_{Soph})\}$, and interest rate spreads, $\{m_k^{FRM} - i_k\}$, are solved for jointly using a fixed point algorithm. In the case of the SRM, the refinancing rule is mechanical and does not need to be solved for, though $\{m_k^{SRM} - i_k\}$ does.

The algorithm to find the FRM equilibrium proceeds as a nested loop. For an initial guess of $\{m_k^{FRM} - i_k\}$, refinancing thresholds are found via policy function iteration. Then, $\{m_k^{FRM} - i_k\}$ is adjusted so that $K^{FRM}(k,k;\chi_{Soph}) = \kappa_k$ in each state. This process is then repeated until $\{m_k^{FRM} - i_k\}$ converges.

As validation that this approach is sound, I analyzed a chain with perfect persistence – a driftless random walk – and calibrated it with $\sigma = 0.0109$ from Table 1. Setting $K = 41$, the results are extremely similar to what is found from the closed-form solutions in the body of the paper: an effectively-constant wedge of $m_k^{FRM} - i_k = 0.0065$ (compared to 0.0067), extraction from an Unsophisticated borrower of 0.036 (compared to 0.037), and deadweight loss on a Sophisticated borrower of 0.010 (compared to 0.011).

G.4 Calibration Results

Table 2 show the parameters used in the calibration. The top panel carries over parameters used in the body of the paper for the random walk version of the model; the bottom panel uses the autoregressive interest rate process estimated by Agarwal *et al.* (2022). Their analysis uses the one-year Treasury rate at constant maturity, a suitable stand-in for i in the present context. The model is solved with $K = 25$.

Symbol	Meaning	Value
ρ	Discount rate	0.05
π	Inflation rate	0.03
μ	Moving hazard	0.10
χ	Attention hazard (Sophisticated)	0.56
C^{Refi}	Refinancing cost (fraction of balance)	0.018
S_{Soph}	Share of Sophisticated borrowers	0.55
C^{Search}_{Soph}	Sophisticated search rent (normalization)	0
μ_i	Cost of capital: unconditional mean	0.049
ρ_i	Cost of capital: annual persistence	0.834
σ_i	Cost of capital: unconditional std. dev.	0.0255

Table 2: Model parameters.

Results are summarized in Table 3 for the states that occur at least 5% of the time in the unconditional distribution. When the cost of capital is low, there is little subsequent refinancing and so the role of the FRM in generating price discrimination is limited, as seen in the very low values of profit in the final column. At higher levels of the interest rate, refinancing is more important and the benefit to lenders of extracting from Unsophisticated borrowers (column 2) grows faster than the deadweight cost (column 3), and so the net profit grows. Evaluated at the stationary average, the net profit from the FRM is less than half what was evaluated with random walk, but critically it is still positive, showing that the model continues to predict that lenders will offer the FRM over the SRM.

Cost of capital	$P^{FRM}(\bar{m}^{FRM}, 0) - P^{SRM}(m^{SRM})$	$D^{FRM}(m^{FRM})$	Net profit from FRM
1.8%	0.007	0.005	0.000
2.8%	0.010	0.007	0.001
3.9%	0.014	0.009	0.002
4.9%	0.020	0.011	0.003
5.9%	0.027	0.013	0.005
7.0%	0.035	0.015	0.008
8.0%	0.045	0.017	0.011
Average	0.024	0.011	0.005
Random walk	0.037	0.011	0.011

Table 3: Core metrics of the model at initial costs of capital that occur with at least 5% in the stationary distribution. For each of those values, the table shows the excess profit from Unsophisticated borrowers extracted by the FRM relative to the SRM, the losses on Sophisticated borrowers from the FRM due to deadweight loss, and finally the net profit from the FRM, which weights the previous metrics by (1-S_{Soph}) and S_{Soph}, respectively. The penultimate row shows the probability-weighted average of those metrics across initial states. The bottom row recreates the analogous metrics from the main text from the model that assumes a random walk.

H Partial Equilibrium Model with Finite Horizon, Risk Aversion, and Mean-Reverting Interest Rate

To check the robustness of the core result (Expression 14 holding) to realistic features like income risk, this appendix describes a numerically-solved finite-horizon model in which a household with CRRA utility is subject to income shocks and an AR(1) interest rate process. Much of the structure and many parameters are based on Appendices C and D from Agarwal *et al.* (2022).

H.1 State Space

The household maximizes expected lifetime utility over a finite horizon of $T = 30$ annual periods. The household make a bequest upon exit due to an exogenous moving shock that arrives with probability p_{move} or survival to the terminal period.

The dynamic program is solved over four state variables, $S_t = \{W_t, Y_t, m_t, m_0\}$, in addition to time, $t = \{1, 2, ..., 30\}$:

1. Real Liquid Wealth (W_t): Real assets available at the start of period t;
2. Real Income (Y_t): Real labor income, which follows an AR(1) process;
3. Market Mortgage Rate (m_t): The current (nominal) interest rate available in the mortgage market, which follows an AR(1) process;
4. Contract Mortgage Rate (m_0): The nominal interest rate currently locked into the household's mortgage.

For simplicity, the nominal principal balance is assumed to be constant (i.e. interest-only mortgage). Following the main text, $B_0 = \$250,000$.

H.2 Mortgage Types

There are two mortgages as in the main text: FRM and SRM.

Under the FRM, the household must be attentive in order to consider refinancing. Attention occurs with probability p_{attn}. If attentive, the borrower chooses the optimal consumption (or next-period wealth, W_{t+1}) associated with either refinancing or not. Refinancing incurs a cost C^{Refi} of the nominal principal. If not attentive, the household simply chooses the optimal consumption and wealth assuming no refinancing.

Under the SRM, the mortgage automatically and costlessly refinances whenever $m_t < m_0$. The optimal decision is derived by maximizing over consumption, conditional on the mechanical refinancing outcome.

H.3 Budget Constraint

The evolution of wealth depends on consumption, mortgage payments, refinancing costs, and income. Note that because mortgage payments and refinancing costs are based on the nominal mortgage balance, they are deflated by $(1 + \pi)$ each period to be cast in real terms. Inflation is assumed to be constant. Households are assumed to be able to save at a real risk-free rate of r_f. The budget constraint therefore is:

$$W_{t+1} = (1 + r_f) \cdot W_t + Y_t - C_t - (1 - I_{Refi,t}) \cdot \left(\frac{B_0 \cdot m_0}{(1+\pi)^{t-1}} \right) \\ I_{Refi,t} \cdot \left(\frac{B_0 \cdot (m_t + C^{Refi} \cdot I_{FRM})}{(1+\pi)^{t-1}} \right), \tag{59}$$

where $I_{Refi,t}$ is an indicator of a refinance occurring in period t and I_{FRM} is an indicator for the mortgage being a FRM rather than SRM.

H.4 Value Function

Upon entering period t, value function is:

$$V_t^j(S_t) = p_{move} \cdot U_{bequest}(W_t) + (1 - p_{move}) \cdot \tilde{V}_t^j(S_t), \tag{60}$$

where $j \in \{FRM, SRM\}$ indexes mortgage type. If no move occurs, the household's flow and continuation values depend on whether it refinances and how much it consumes (or alternatively, how much wealth it passes to the next period). The flow utility function $u(C)$ is assumed to be CRRA:

$$u(C) = \frac{C^{1-\gamma}}{1-\gamma}. \tag{61}$$

If a move does occur, the household gets bequest utility:

$$b(W) = \alpha \cdot \frac{W^{1-\gamma}}{1-\gamma} \tag{62}$$

In the case of no move, a borrower with a SRM chooses optimal consumption given the mechanical refinancing rule, so:

$$\tilde{V}_t^{SRM}(S_t) = \max_{C_t} \left\{ u(C_t) + \beta \cdot E_t[V_{t+1}^{SRM}(S_{t+1})] \right\}, \tag{63}$$

subject to Equation 59.

For an FRM borrower, the attention shock is relevant. Defining $\tilde{V}^{FRM}(S_t|I_{Refi})$ to be the value of having an FRM in period t and state S_t, conditional on the refinancing decision, $\tilde{V}^{FRM}(S_t|0)$ and $\tilde{V}^{FRM}(S_t|1)$ are defined analogously to Equation 63, and then we have:

$$\tilde{V}^{FRM}_{Attn}(S_t) = \max\left\{\tilde{V}^{FRM}(S_t|0), \tilde{V}^{FRM}(S_t|1)\right\} \tag{64}$$

and

$$\tilde{V}^{FRM}_{No\ attn}(S_t) = \tilde{V}^{FRM}(S_t|0) \tag{65}$$

We then have:

$$\tilde{V}^{FRM}(S_t) = p_{attn} \cdot \tilde{V}^{FRM}_{Attn}(S_t) + (1 - p_{attn}) \cdot \tilde{V}^{FRM}_{No\ attn}(S_t). \tag{66}$$

H.5 Computational Solution

The stochastic processes for Y_t and m_t are modeled as AR(1) processes and discretized using the Tauchen method.

Wealth is a continuous state variable and is effectively a choice variable given Equation 59. Since the optimal next-period wealth choice generally does not align with a grid point, the model solves the problem by computing the expected continuation value $E_t[V_{t+1}(S_{t+1})]$ as the integral across wealth's grid points and then employing linear interpolation to approximate the continuation value on a much finer grid of wealth, allowing for a more precise choice of consumption.

The value functions are then solved for via backwards induction.

H.6 Model Parameters

Table 4 lists the parameter values used in the quantitative model.

Parameter	Meaning	Value	Source
$T_{horizon}$	# of periods (years)	30	
γ	Risk aversion	5	Agarwal *et al.* (2022)
α	Bequest weight	3.0	Agarwal *et al.* (2022)
ρ	Discount rate	0.05	Agarwal *et al.* (2013)
μ_y	Unconditional mean of Y	\$75k	
ρ_y	Income persistence (Log)	0.90	Agarwal *et al.* (2022)
σ_y	Income innovation std. dev. (Log)	0.21	Agarwal *et al.* (2022)
μ_m	Unconditional mean of m	0.049	Agarwal *et al.* (2022)[54]
ρ_m	Mortgage rate persistence	0.834	Agarwal *et al.* (2022)
σ_m	Mortgage rate std. dev.	0.0255	Agarwal *et al.* (2022)
B_0	Nominal Mortgage Principal Balance	250,000	
C^{Refi}	Refinancing Cost (fraction of B_{nom})	0.018	Agarwal *et al.* (2013) (for \$250k mtg)
π	Inflation rate	0.03	Agarwal *et al.* (2013)
r_f	Real risk-free return on wealth	0.019	$m - \pi$
p_{move}	Moving probability	0.095	Agarwal *et al.* (2013)[55]
p_{attn}	Attention probability	0.43	Berger *et al.* (2024)[56]
S_{Soph}	Share of Sophisticated Borrowers	0.55	Berger *et al.* (2024)

Table 4: Model Parameters

[54] They report a real rate of 1.9%, to which I add 3pp of inflation.
[55] They use a continuous-time hazard rate of 0.1, which is equivalent to this 0.095 in discrete annual periods.
[56] They find an average continuous-time hazard rate of 0.56, which is equivalent to this 0.43 in discrete annual periods.

H.7 Results

The backward induction process generates time-dependent policy functions (optimal consumption C, and refinancing decisions I_{Refi}) for every state defined by (W, Y, m, m_0). These policies are then used in Monte Carlo simulations to evaluate the profitability of the FRM and SRM contracts. Specifically, shocks are simulated to Y, and m, as well as to attention and moving. For a given path, consumption and refinancing are simulated using the policy functions. To allow for decisions that do not fall on the discretized Wealth grid, policy decisions are determined by linearly interpolating the policy arrays over wealth level. The simulation records the expected accumulated discounted utility and the expected discounted stream of mortgage payments.

To compare lender profitability across mortgage types, I first identify the expected utility level achieved by a Sophisticated borrower under the FRM, assuming some initial interest rate. Referring to that expected utility level as $E[U^{FRM}_{Soph}(m_0)]$, the simulation also finds $P^{FRM}(m_0, \chi)$ and $P^{FRM}(m_0, 0)$, the payments received by the lender from Sophisticated and Unsophisticated borrowers, respectively.[57,58] I then find the SRM interest rate that yields the same expected utility for the Sophisticated borrower, $\bar{m}^{SRM}(m_0)$ and then rerun the simulations a final time to get $P^{SRM}(\bar{m}^{SRM}(m_0))$, payments receive by the lender under the SRM.

Table 5 shows the results for an initial liquid wealth of \$62k and income of \$88k. Across a range of initial interest rates, the FRM's extraction from Unsophisticated borrowers outweighs the losses on Sophisticated borrowers, leading to excess profit from the FRM relative to the SRM. Notably, the magnitude is quite comparable to the result from the risk-neutral model for moderate interest rate levels, though the margin does shrink considerably for the very low $m_0 = 3\%$.

[57] Using the notation of the main text, χ is attention parameter that maps one-to-one to p_{attn}. Sophisticated borrowers with $\chi = 0.56$ pay attention in 43% of years.

[58] Upon moving or reaching term, the lender receives back the principal balance. This is not paid out of the household's liquid wealth but rather is assumed to be covered by the proceeds from selling the home.

m_0	$P^{FRM}(m_0, 0) - P^{SRM}(\bar{m}^{SRM}(m_0))$	$P^{SRM}(\bar{m}^{SRM}(m_0)) - P^{FRM}(m_0, \chi)$	Net profit from FRM
3.0%	0.011	0.000	0.005
4.0%	0.028	0.000	0.013
4.9%	0.034	0.008	0.011
6.0%	0.044	0.015	0.012
Baseline model	*0.037*	*0.011*	*0.011*

Table 5: Core metrics of the model at different initial interest rates, assuming initial wealth of \$62k and initial income of \$88k. For each of the initial interest rates shown, the table shows the excess profit from Unsophisticated borrowers extracted by the FRM relative to the SRM, the losses on Sophisticated borrowers from the FRM, and finally the net profit from the FRM, which weights the previous metrics by (1-S_{Soph}) and S_{Soph}, respectively. The interest rates shown are those on the discretized interest rate grid closest to 3%, 4%, 5%, and 6%. The bottom row recreates the analogous metrics from the main text from the model that assumes risk-neutrality and a random walk.