Equilibrium Mortgage Design with Heterogeneous Borrowers*

Joshua Abel†

October 28, 2024

Abstract

Conventional mortgages penalize "unsophisticated" borrowers for their failure to refinance. This paper shows that the absence of mortgages that better serve such borrowers – for instance by including an automatic refinancing provision – can be rationalized as a strategy by lenders to extract rents. Conventional mortgages allow lenders to charge a higher effective price of credit to unsophisticated borrowers, who are less likely to shop around during mortgage origination, than to sophisticated borrowers. An automatic refinancing provision would eliminate this price discrimination and lower rents, explaining why it is not offered. Prepayment penalties and "points" also impact the price discrimination mechanism. The baseline calibration of the paper's model predicts that lenders should offer to pay points to borrowers at origination and not include prepayment penalties, consistent with what is observed in most market segments in recent decades. A policy of taxing borrowers for refinancing would benefit the unsophisticated by lowering interest rates and, potentially, leading to a change in mortgage designs that lenders offer.

*I am grateful for helpful feedback from Paul Willen, John Campbell, Edward Glaeser, Robert Triest, Andreas Fuster, Adam Guren, Carol Osler, Andrew Abel, Valentin Bolotnyy, Adrien Bilal, Michael Seiler, and seminar participants at Northeastern University and Brandeis University.

†Northeastern University. E-mail: j.abel@northeastern.edu

1 Introduction

This paper considers why financial 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.[1] 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.[2] In many other countries, the typical contract is an adjustable-rate mortgage (ARM).[3] Refinancing is important in these countries, such as Australia and the United Kingdom, 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.[4] A growing literature finds that this heterogeneous refinancing amounts to a substantial[5] and regressive transfer, given that refinancing propensities tend to be higher among borrowers with markers of a stronger financial position.[6] Recent work has also quantified substantial aggregate deadweight costs of

[1] A large recent literature has shown the positive outcomes that accrue to borrowers from refinancing their mortgages, such as lower defaults and increased spending. See e.g., Agarwal *et al.* (2022), Beraja *et al.* (2019), Bhutta and Keys (2016), Chen *et al.* (2019), Ehrlich and Perry (2015), Di Maggio *et al.* (2017), Di Maggio *et al.* (2020), Hurst and Stafford (2004), Fuster and Willen (2017), Zhu *et al.* (2015), Tracy and Wright (2016), Karamon *et al.* (2016), Abel and Fuster (2021). Benefits like these had long been assumed by policymakers. See e.g., Greenspan (2004).

[2] See, e.g. Andersen *et al.* (2020), Zhang (2023), Berger *et al.* (2023b).

[3] See, e.g., Lea (2010).

[4] See Fisher *et al.* (2024).

[5] Berger *et al.* (2023b) estimate a structural model and find that the most-sophisticated borrowers make payments that are roughly 15% lower than the least-sophisticated borrowers over their lifetime, in present value. Zhang (2023) also studies a structural model of the mortgage market in which borrowers have heterogeneous refinancing costs and finds a cross-subsidy of over $1,300 from unsophisticated to sophisticated types. Fisher *et al.* (2024) study the mortgage market in the UK, where the dominant mortgage is the ARM with a teaser rate. 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 halfway between the pre- and post-reset rates in the current regime. As a result, that alternative contract would help the unsophisticated borrowers, who make many payments at the high post-reset rate in the current regime, and hurt the sophisticated borrowers, who would no longer have access to the low pre-reset rate by regularly refinancing.

[6] For instance, Abel and Fuster (2021) show that borrowers with higher credit scores, more unused revolving credit, and lower loan-to-value ratios on their homes are more likely to refinance, as are borrowers in higher-income ZIP codes. In Danish data, Andersen *et al.* (2020) find that borrowers with higher income and education are more likely to refinance. Gerardi *et al.* (2023) find that White and Asian borrowers are more likely to refinance than Black and Hispanic borrowers in the US, a difference that persists even when controlling for credit score, income, and home equity. Agarwal *et al.* (2015), Agarwal *et al.* (2009), Johnson *et al.* (2018), Keys *et al.* (2016), Kiefer *et al.* (2023) and Deng and Quigley (2007) are among other papers that document and analyze the failure of different types of borrowers to refinance.

the mortgage origination process, with refinancing playing an important role.[7] 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.[8,9] But in equilibrium, such a feature is not offered. In fact, as pointed out by Zhang (2023), 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 the mortgage market typically assume perfect competition among lenders, 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 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 Unso-

[7] See Fuster *et al.* (forthcoming) and Zhang (2023).

[8] Berger *et al.* (2023b) 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 (2023) 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.

[9] In addition to the literature discussed above about deadweight refinancing effort and distributional issues, other papers have advocated for mortgages with such state-contingent payments due to macroeconomic risk. See e.g., Piskorski and Tchistyi (2011), Piskorski and Tchistyi (2020), Campbell (2013), Eberly and Krishnamurthy (2014), Campbell *et al.* (2021), Guren *et al.* (2021), Piskorksi and Seru (2018), and Bhagat (2021). Advocacy for automatically refinancing mortgages goes back to at least Flesaker and Ronn (1993), who gave the contract the provocative name "Falling Interest Rate Adjustable-Rate Mortgage (FIREARM)."

phisticated, who have high search costs during origination and do not consider refinancing. 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. 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.

Importantly, this price discrimination comes at a cost to lenders: the deadweight resource cost that active refinancing requires, which they (at least partially) bear. This creates a tradeoff for lenders when considering mortgage designs. While an automatic refinancing provision would eliminate their ability to price discriminate, it would cause an increase in 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 prepayment penalties would lower rents, and an automatic refinancing provision would lower rents even further. As discussed above, observed mortgage designs largely coincide with these predictions of the model: 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, although mortgages in France and Germany are notable exceptions.[10]

So, the paper provides a positive explanation for why the normative benefits of the prior

[10] See Lea (2010). Prepayment penalties were not always uncommon in the US; they largely disappeared only in the 1980s, though they reappeared in the subprime market of the 2000s. This is discussed below.

literature's proposed financial innovation are not realized. However, the analysis concludes by showing that a perhaps-counterintuitive policy could generate those normative benefits, even in the face of the obstacles highlighted here: tax borrowers for (actively) refinancing. Such a tax would chip away at the effectiveness of the price discrimination mechanism: it would increase the private burden of refinancing on Sophisticated borrowers, and by lowering the equilibrium interest rate it would reduce the ability of lenders to extract from Unsophisticated borrowers. A tax of sufficient size would render the price discrimination mechanism unprofitable and so could endogenously generate the automatic refinancing provision that the prior literature has advocated.

While positive issues of equilibrium mortgage design are less studied than normative issues of optimal mortgage design, 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 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. 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, among other reasons, 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 1960s and GSEs began engaging with them in the 1980s.[11] 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, it is clear that other financial innovations have overcome it. Indeed, mortgages with automatic refinancing provisions seem institutionally 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.

[11] See Lea and Sanders (2011).

Another proposed explanation comes from Berger *et al.* (2023b), 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. While this is true, it cannot simultaneously explain the effective absence of mortgages with prepayment penalties in most market segments, which as discussed above could lower the price of credit for unsophisticated borrowers with a *lower* initial interest rate.

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.[12] 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,[13] 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. Relative to the rest of the literature on the topic, the key difference in this paper's model 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.[14] Woodward and Hall (2010) and

[12] Campbell (2013) provides a helpful explanation of this mechanism.

[13] It would have a lower interest rate than a FRM because it does not contain a premium to compensate lenders for the option to refinance, a point also made about mortgages with prepayment penalties in Mayer *et al.* (2013). Beltratti *et al.* (2017), Elliehausen *et al.* (2008), and Ho and Pennington-Cross (2008) provide empirical evidence in favor of this discount associated with prepayment penalties.

[14] Nelson (2023) also finds that it is critical to consider lender rents in the market for credit cards. He finds that a policy limiting interest rate increases on credit cards created substantial consumer surplus, largely resulting from reduced markups charged by lenders. The importance of lenders' ability to extract rents is also being emphasized more in studies of monetary policy transmission. See e.g., Scharfstein and Sunderam (2016), Wang *et al.* (2022), and Enkhbold (2023). Relatedly, Bjonnes *et al.* (2024) documents

Woodward 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.* (2024) 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 also bear other markers that suggest they are more 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 UK. 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 mortgages that punish unsophisticated borrowers dominate while more equitable and efficient ones are not observed. In fact, as discussed in Section 2, the model assumes risk-neutrality for tractability, and as a result, it perceives no difference between an automatically-refinancing mortgage, a mortgage that *cannot* be refinanced, and a "vanilla" ARM with no teaser rate. All three mortgages share two key features: they have no deadweight loss (due to the lack of active refinancing), and they treat all borrowers equally. As a result, the model does not technically take a stand on which would be the most likely alternative to conventional mortgages. I focus on the FRM with an automatic refinancing provision because US borrowers seem to appreciate the risk-mitigation and guaranteed cap on payments that FRMs provide, particularly when

extensive variation in execution costs for OTC trades, with client sophistication being the primary driver of that variation.

inflation is stable.[15] But given other institutional features or preferences, other efficient and equitable designs may result.

The paper proceeds as follows. Section 2 presents the model and its key qualitative results, the most important being that an equilibrium can exist with no mortgages with automatic refinancing provisions, but only under certain conditions. Section 3 shows that using off-the-shelf parameter values to calibrate the model, it correctly predicts that lenders do not offer the automatic refinancing provision (i.e. those "certain conditions" do hold). Section 4 extends the model to allow lenders to add additional features to the mortgage: prepayment penalties and points. Section 5 discusses the benefits of a refinancing tax, and Section 6 provides a concluding discussion of the paper's findings.

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.* (2023b) 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.

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 and have discount 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 United States, 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. In setting up the model, this section focuses on the FRM; the SRM is discussed subsequently.

Assume an inflation rate of π and that the real short-term interest rate, r, follows a Brownian motion:[16]

[15] See, e.g., Campbell and Cocco (2003).

[16] Agarwal *et al.* (2013) assume the real FRM interest rate itself exogenously follows Brownian motion. In my model, mortgage rates are endogenous but, as discussed at length in Section 2.2, are above the short-term interest rate by a constant wedge. Therefore, in my model, it is the short-term interest rate that varies

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

where dz represents standard Brownian increments. Given this setup, the nominal short-term interest rate, $i_t = r_t + \pi$, is a random walk.[17]

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 over time), but more crucially it ensures that m_t^{FRM} follows a Brownian motion, as discussed in Section 2.2.[18]

exogenously, which drives movements in mortgage rates. Agarwal *et al.* (2013) note that all results of their model go through under the assumption that the short-term riskless rate follows a random walk and that the mortgage rate is a constant wedge above that level, which is the approach taken in this paper.

[17] Agarwal *et al.* (2013) discuss at length that the assumption of Brownian motion is not innocuous, as some researchers have argued that nominal interest rates exhibit mean reversion, though others have disputed that. As in Agarwal *et al.* (2013), I set up a model in which interest rates follow a random walk rather than exhibit mean reversion for analytical tractability. See Berger *et al.* (2023b) for analysis of a similar problem with an interest rate that exhibits mean reversion.

[18] The assumption of an interest-only mortgage differs subtly from the model of Agarwal *et al.* (2013). 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, as I show in Section 3.

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.[19] 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 (2023).[20] In particular, she is only aware of her refinancing incentive when she receives an "attention shock," which arrives at a hazard rate of χ. Under this assumption, I follow Berger *et al.* (2023b) to solve for the optimal refinancing rule in this setting.

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.* (2023b), the Appendix shows that $y^*(\chi)$, the threshold value such that a borrower with attention parameter χ who is paying attention at time t will refinance if and only if $y < y^*(\chi)$, results from a system of three equations in three unknowns (the other two unknowns being coefficients in the borrower's value function) which 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 present discounted value of expected costs, at the time of origination, for a borrower with attention parameter χ who signs her FRM contract with 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

[19] As the exposition focuses on a mortgage of a single size, I abstract from considering the distinction between fixed costs of refinancing and costs that vary with the mortgage size. As Agarwal *et al.* (2013) point out, the presence of fixed refinancing costs makes refinancing a more appealing option for borrowers with large mortgages, as the growth in the costs of refinancing is outstripped by the growth in its benefits, as mortgage size increases. Additionally, Kiefer *et al.* (2023) point out that there is substantial variation in refinancing costs based on state- and sub-state-level policies. In part motivated by this, the quantitative analysis in Section 3.3 show sensitivities to C^{Refi}.

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

mortgage payments, and $D^{FRM}(\chi)$ accounts for deadweight refinancing costs that are captured by no one.[21] 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. The Appendix 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:

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

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

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 almost seamlessly. 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:

$$m_t^{SRM} = i_t + \Delta^{SRM}. \quad (6)$$

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 present discounted value of expected (real) costs, at the time of origination, for a borrower who signs her SRM with interest rate m_0^{SRM}, we have:

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

where $P^{SRM}(m_0^{SRM})$ is the expected payment to the lender.[22] SRM outcomes differ from the FRM in two critical ways. First, there is no deadweight loss, since borrowers are not

[21] 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 the Appendix.

[22] 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.

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.[23] Assume fraction S_{Soph} are "Sophisticated" borrowers, who refinance 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 not refinance their mortgages.[24]

Assume there are at least two 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. Unsophisticated borrowers have higher search costs: $C^{Search}_{Uns} \geq C^{Search}_{Soph} \geq 0$.[25] 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

[23]Since mortgage contracts can be conditioned on observables, it is not a substantive assumption to assume that all borrowers have the same loan size. It simply means that the analysis solves for the mortgage design for this particular loan size. There could be many different groups in a cohort, each with a different loan size; the model would just be solved separately for each one. There *is* a substantive assumption which is that the mortgage size is exogenous, meaning borrowers do not decide whether nor how much to borrow in response to the mortgage offers they receive. I maintain this assumption for tractability.

[24]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.

[25]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). A model in which, rather than revealing all offers at once, searching allowed a borrower to view one additional offer at a time could generate the same results.

must maximize profits by offering the same mortgage(s) to all borrowers.[26] 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 profit margins from them. Costly refinancing will be central to that mechanism.

I will make one more assumption that is not important for the key conceptual results of the model but is convenient when performing the quantitative evaluation of the model in Section 3. 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. The core qualitative results of the model do not depend on this assumption, but the assumption is helpful quantitatively because it allows the model to generate a no-SRM equilibrium even if Sophisticated borrowers have low (or even zero) search costs. In the absence of this assumption, lenders might effectively ignore Sophisticated borrowers if their search costs are low, obviating any incentive to price discriminate.

In summary, each borrower is matched to a lender who makes a mortgage offer in a captive phase; the borrower can accept a mortgage in this phase, or reject and pay a cost to move to a Bertrand phase in which they receive offers from all lenders.

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, for the life of the mortgage, i.e. until a moving shock forces the repayment of the mortgage. Denoting by i_0 the nominal short-term interest rate at the time

[26] Related to the discussion of loan size in the footnote above, 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.

of origination, the (real) expected present value 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.[27] 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 capturing 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:

$$\frac{i_0+\mu}{\rho+\mu+\pi} + C_j^{Search} = K^{SRM}(\bar{m}_j^{SRM}) = K^{FRM}(\bar{m}_j^{FRM}, \chi_j), \tag{8}$$

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

The reservation interest rates on a FRM are therefore:

$$\bar{m}_j^{FRM} = i_0 + v_j + \omega_j^{FRM}, \tag{9}$$

where v_j accounts for expected profit[28] and ω_j^{FRM} is a premium for the option to refinance.[29] Similarly, the reservation interest rates on a SRM are:

$$\bar{m}_j^{SRM} = i_0 + v_j + \omega^{SRM}. \tag{10}$$

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.[30]

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.[31] However, it is ambiguous which borrower type will have a higher reservation interest rate on an FRM. Intuitively, Unsophis-

[27] 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.

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

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

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

[31] Formally, $\bar{m}_{Uns}^{SRM} - \bar{m}_{Soph}^{SRM} = v_{Uns} - v_{Soph} = (\rho+\mu+\pi) \cdot (C_{Uns}^{Search} - C_{Soph}^{Search}) \geq 0$, since $C_{Uns}^{Search} \geq C_{Soph}^{Search}$.

ticated borrowers have higher search costs, which imply a higher reservation interest rate, as their outside option of searching is costlier.[32] 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.[33] 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.

2.2.3 Lenders' Profit-maximizing Mortgage Offer(s)

The main conceptual result of the paper occurs in the case when $\bar{m}_{Uns}^{FRM} > \bar{m}_{Soph}^{FRM}$, as this condition allows for the possibility that that no SRM will be offered in equilibrium, and the FRM will dominate. 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 $\bar{m}_{Soph}^{SRM}$, and profit will be:

$$V_{SRM} = P^{SRM}(\bar{m}_{Soph}^{SRM}) - \frac{i_0 + \mu}{\rho + \mu + \pi}. \tag{11}$$

The first term is the value of the expected payments from borrowers, which are equal between Sophisticated and Unsophisticated 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 $\bar{m}_{Soph}^{FRM}$, and profit will be:

$$V_{FRM} = S_{Soph} \cdot P^{FRM}(\bar{m}_{Soph}^{FRM}, \chi) + (1 - S_{Soph}) \cdot P^{FRM}(\bar{m}_{Soph}^{FRM}, 0) - \frac{i_0 + \mu}{\rho + \mu + \pi}. \tag{12}$$

The first term represents value generated from Sophisticated borrowers, the second term represents value generated 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.[34]

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

[32]This is the same logic as for the SRM described in the previous footnote: $v_{Uns} > v_{Soph}$.

[33]Formally, $\omega_{Uns}^{FRM} < \omega_{Soph}^{FRM}$.

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

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 if the following condition holds:*

$$\underbrace{(1 - S_{Soph}) \cdot \left(P^{FRM}(\bar{m}^{FRM}_{Soph}, 0) - P^{SRM}(\bar{m}^{SRM}_{Soph})\right)}_{\substack{\textit{Additional revenue extracted from} \\ \textit{Unsophisticated borrowers by FRM}}} > \underbrace{S_{soph} \cdot D^{FRM}(\chi)}_{\substack{\textit{Reduced revenue extracted from} \\ \textit{Sophisticated borrowers by FRM}}} . \tag{13}$$

If Expression 13 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 the Appendix, but the intuition is as follows. Expression 13 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 can capture by sufficiently 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, because the lenders could realize that as revenue if the Sophisticated borrowers were spared that effort with the SRM.

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 13 did not hold, this effect would dominate and lenders would do better by offering the efficiency-maximizing SRM.[35]

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

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

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

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

This, too, is proven in the Appendix, 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 $\bar{m}_{Uns}^{FRM} < \bar{m}_{Soph}^{FRM}$ and $\bar{m}_{Uns}^{SRM} > \bar{m}_{Soph}^{SRM}$. As a result, the lender can get the Unsophisticated borrower to agree to a SRM at her reservation interest rate ($\bar{m}_{Uns}^{SRM}$) 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 they 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 14), it has the drawback of

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

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 due to its value in price-discriminating; when the expression does not hold, lenders should simply offer an SRM.

2.3 Comparison to Standard Assumption of Perfect Competition

Proposition 1 is a novel result, as it 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"), a point also recognized by Berger *et al.* (2023b).

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

$$S_{Soph} \cdot P^{FRM}(\hat{m}_{PC}^{FRM}, \chi) + (1 - S_{Soph}) \cdot P^{FRM}(\hat{m}_{PC}^{FRM}, 0) = \frac{i_0 + \mu}{\rho + \mu + \pi}. \tag{15}$$

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.[36]

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:

$$P^{SRM}(m_{PC}^{SRM}) = \frac{i_0 + \mu}{\rho + \mu + \pi}. \tag{16}$$

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 15. This causes adverse selection in the FRM market, as lenders will be left with

[36] Lenders receive less than the opportunity cost of capital from Sophisticated borrowers. Intuitively, Unsophisticated borrowers cross-subsidize Sophisticated borrowers by refinancing less. This is a point stressed by Zhang (2023) and Berger *et al.* (2023b), among others.

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:

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

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.

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. In trying to rationalize this, Berger *et al.* (2023b) point out that the the SRM would have a higher interest rate, so that might cause borrowers to exceed acceptable payment-to-income ratios, or highly-constrained borrowers may prefer the lower-interest rate FRM with its lower initial payments. Campbell (2013) argues that Unsophisticated borrowers may not understand the benefits the SRM would afford them.

While these explanations have some appeal, they cannot simultaneously explain the absence of other types of mortgages, such as one that cannot be refinanced.[37] While seemingly the opposite of a SRM, it in fact would achieve essentially the same results, allowing Unsophisticated borrowers to pay the cost of capital by surrendering the option to refinance, which they were not going to use anyway. For a borrower who would not refinance, this should be no harder to understand than a FRM, and it would come with a *lower* interest rate, making it the most appealing to constrained borrowers and achieving a lower payment-to-income ratio. Yet, like the SRM, such a mortgage is not observed.

From the perspective of the model in this paper, the problem with the Perfect Competition assumption is that it 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

[37] Such a mortgage has similarities to a mortgage with prepayment penalties, which is observed. This is analyzed in depth in Section 4.1, but suffice it to say here that a mortgage with a finite prepayment penalty has critical differences relative to a mortgage that cannot be refinanced. Most importantly, it still generates heterogeneous payments among borrowers, and it also generates deadweight refinancing costs, making it more like the FRM than the SRM.

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.* (2023b)
χ	Attention for Sophisticated	0.56	Berger *et al.* (2023b)
C^{Search}_{Soph}	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

a minor deviation deviation from Perfect Competition; lenders are prepared to compete fiercely, but (at least some) borrowers might have to pay a cost to initiate that process. This gives lenders a measure of power to extract a profit, 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.

3 Quantitative Evaluation

Proposition 1 showed that the SRM can be absent from the market if Expression 13 holds and $\bar{m}^{FRM}_{Uns} > \bar{m}^{FRM}_{Soph}$. This section begins by showing that Expression 13 holds when the model is calibrated using parameter values from the literature. The question of whether $\bar{m}^{FRM}_{Uns} > \bar{m}^{FRM}_{Soph}$ hinges on $C^{Search}_{Uns} - C^{Search}_{Soph}$, so the section then surveys evidence from the literature on dispersion in borrowers' effective search costs. The section concludes by performing comparative statics of mortgage design on the model's primitives.

3.1 Expression 13 Under Baseline Calibrations

To assess whether Expression 13 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.* (2023b), as shown in the top panel of Table 1.[38]

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, $m_0^{FRM} = i_0 + v_{Soph} + \omega^{FRM}_{Soph}$. While ω^{FRM}_{Soph} is pinned down (at 0.0067)[39] by the parameters listed in the top panel, i_0 and v_{Soph} are

[38]See page 602 of Agarwal *et al.* (2013) and Table C.1 of Berger *et al.* (2023b).

[39]This is comparable to the premium for the option to refinance discussed in Lea and Sanders (2011),

not. The same statement holds for an SRM.[40] The values of i_0 and C^{Seach}_{Soph} shown in the table are not the unique combination to generate $m^{FRM}_0 = 6\%$, but their precise breakdown does not substantively impact the results. The only impact that these parameters have on the results is that C^{Seach}_{Soph} does affect the calibration of the final parameter in the table, S_{Mixed}, discussed below. The combination with $C^{Search}_{Soph} = 0$ is chosen to keep the model as close to the standard Perfect Competition benchmark as possible.

Under these parameter assumptions, $D^{FRM}(\bar{m}^{FRM}_{Soph}, \chi) = 0.011$. 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 an additional \$0.037 on a FRM relative to a SRM. Given $S_{Soph} = 0.55$, Expression 13 holds: $(1 - 0.55) \cdot 0.037 > 0.55 \cdot 0.011$. So, a FRM is more profitable than a SRM.

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.[41] With this added wrinkle, the FRM is more profitable than the SRM if and only if:

$$(1 - S_{Mixed}) \cdot (1 - S_{Soph}) \cdot \left(P^{FRM}(\bar{m}^{FRM}_{Soph}, 0) - P^{SRM}(\bar{m}^{SRM}_{Soph})\right) > S_{Soph} \cdot D^{FRM}(\chi), \quad (18)$$

which was 0.005.

[40] Recall that $v_{Soph} = C^{Search}_{Soph} \cdot (\rho + \mu + \pi)$, and $\omega^{FRM}_{Soph} = \frac{\chi \cdot \left(1 - \eta \cdot (y^*(\chi) - (\rho+\mu+\pi) \cdot C^{Refi})\right)}{(\rho+\mu+\pi+\chi) \cdot (\eta+\psi) \cdot exp(-\psi \cdot y^*(\chi)) - \chi \cdot \eta}$. For the SRM, $\omega^{SRM} = 1/\psi$.

[41] 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.

which generalizes Expression 13 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.[42]

I check the robustness of this result to two issues discussed at length in Agarwal *et al.* (2013): tax deductibility of mortgage payments and closing costs, and amortization of the mortgage balance. The details are shown in the Appendix, but at a high level, the issue of tax deductibility impacts the effective cost of refinancing, as lowering mortgage payments affects a household's tax bill. 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. As discussed in Section 2.1, I cannot assume that the decline in the real value of the mortgage depends on the interest rate, as that would ruin the result that the equilibrium interest rate follows a Brownian motion. Nonetheless, I can check the sensitivity of the results to this assumption by testing how they change if the decline in mortgage value were higher by $\frac{0.06}{exp(30 \cdot 0.06)-1} = 0.01$, in a way that is exogenous and not interest rate-dependent. When these adjustments are incorporated, the critical value of S_{Mixed} falls from 65% to 60%. In other words, so long as at least 40% (rather than 35%) of non-Sophisticated borrowers have high search costs, the model still predicts that the FRM is more profitable than the SRM. So these changes do not have a large impact on the model's results.

In summary, an off-the-shelf calibration of the model presented in Section 2 indeed suggests that the price discrimination that the FRM affords lenders outweighs the inefficiency it causes, endogenously leading lenders to choose against offering the SRM.

[42]Note that while I treated S_{Mixed} as a free parameter to match the profit level of 0.62%, there was no guarantee that such a positive value would exist, nor even that the FRM would generate positive profit at all. So while S_{Mixed} was chosen to match the data, this calibration still provides a valid test of the model, as it was possible for the data to reject the model, which it did not.

3.2 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), 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.[43] 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 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.* (2024) use rich data on mortgage originations and offer sheets to evaluate the extent that borrowers pay more than the price lenders are willing to transact at. They find that the

[43] See Table A-1 of Bhutta *et al.* (2024).

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 $\bar{m}_{Uns}^{FRM} > \bar{m}_{Soph}^{FRM}$.

Note that as this paper proceeds, there will be little further discussion of search costs. That is because, conditional on $\bar{m}_{Uns}^{FRM} > \bar{m}_{Soph}^{FRM}$ – i.e. search costs for Unsophisticated borrowers being high enough – they do not play any further role in the model's predictions about mortgage design. That is why Table 1 does not provide a value for C_{Uns}^{Search} and why, as discussed above, the level of C_{Soph}^{Search} does not play an important role in determining mortgage design. The subsequent analysis assumes C_{Uns}^{Search} is high enough to ensure $\bar{m}_{Uns}^{FRM} > \bar{m}_{Soph}^{FRM}$, unless otherwise stated.

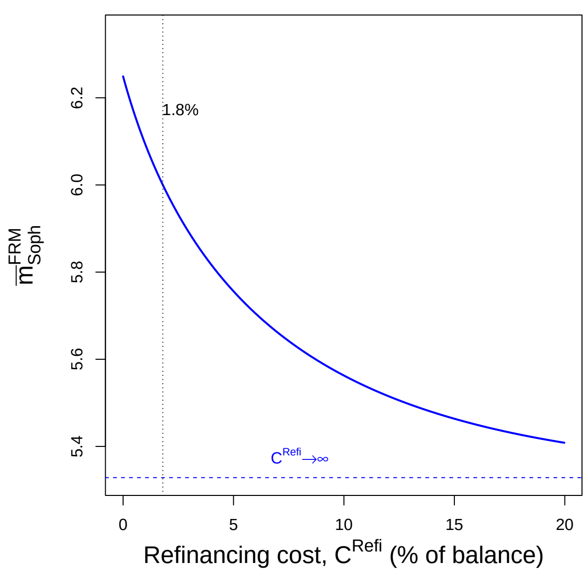
3.3 Comparative Statics

To better understand the model and its quantitative implications, this subsection shows how it behaves under different parameter values. Figure 1 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 1a, a higher cost of refinancing lowers the FRM interest rate as it lowers Sophisticated borrowers' inclination to refinance. Importantly, Figure 1b 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, as they do not refinance. Therefore, higher refinancing costs benefit the Unsophisticated borrower by reducing the interest rate that Sophisticated borrowers are willing to accept.[44] Figure 1c shows how the deadweight loss associated with Sophisticated 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.

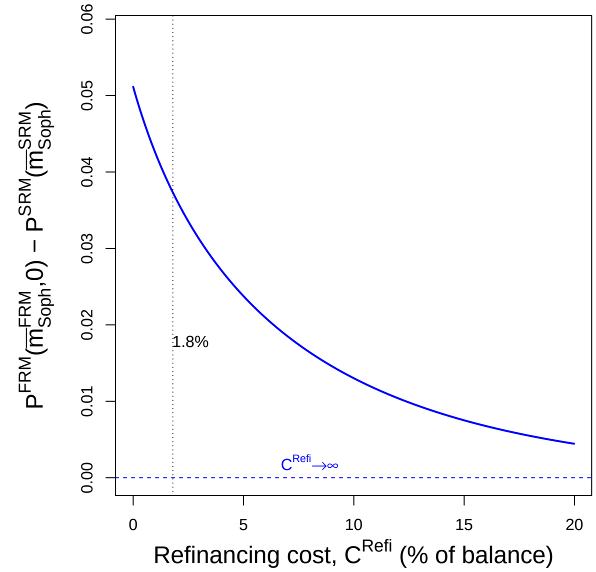
Figure 1d brings it all together, showing how lenders' rents depend on the refinancing cost. This combines two core effects: higher cost leads to a lower interest rate and therefore lower revenue from the Unsophisticated borrowers; the higher cost affects the deadeweight loss and therefore the amount 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 2 shows how profit depends on some of the other key parameters. As shown in Figure 2a, 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 2b shows

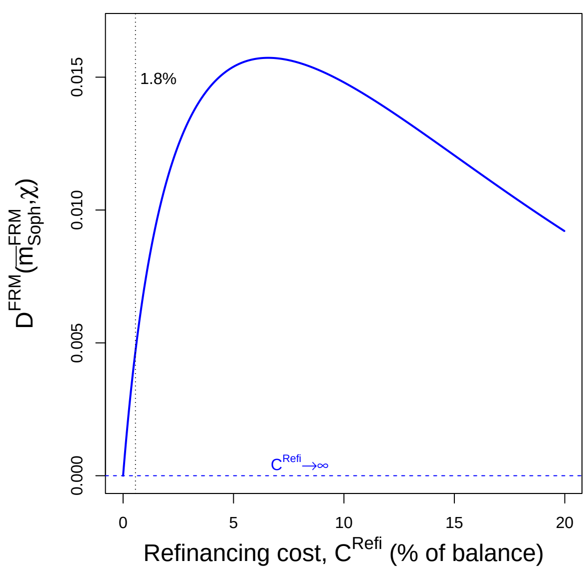
[44] Critically, Sophisticated borrowers are indifferent to where in the parameter space the model is, because their cost is constant – it is pinned down by their outside option of rejecting offers. This constant cost is in fact what pins down the equilibrium interest rate.



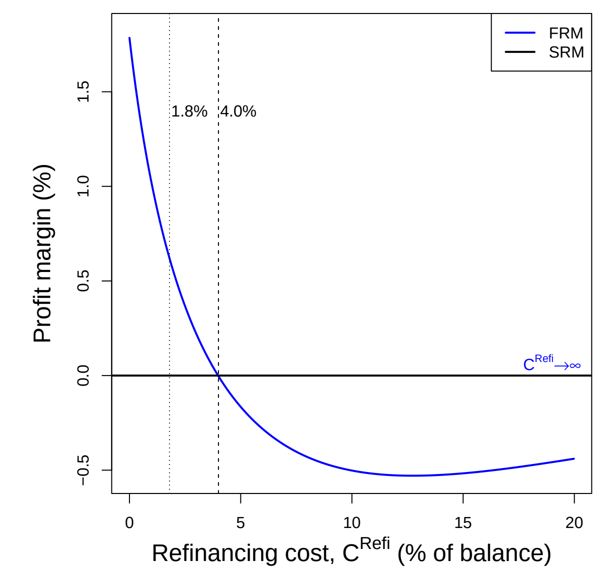
(a) Equilibrium FRM interest rate.



(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 1: Comparative statics with respect to refinancing cost (C^{Refi}) of key model outputs using the calibration from Section 3.1 (without tax deductibility or amortization). The baseline level of C^{Refi} (1.8%) is shown with a vertical line. Horizontal blue lines show the limiting values as $C^{Refi} \to \infty$. A second vertical line in Figure 1d shows the level of C^{Refi} at which the lender is indifferent between the FRM and SRM.

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

Figures 2c and 2d 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.

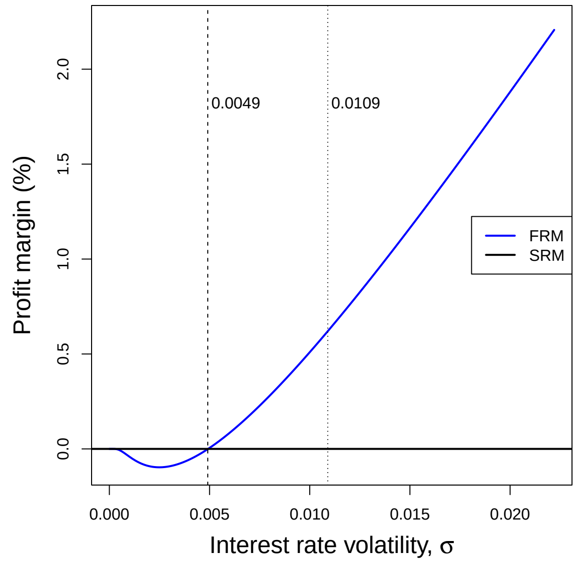
4 Expanding the Mortgage Space: Prepayment Penalties and Points

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.

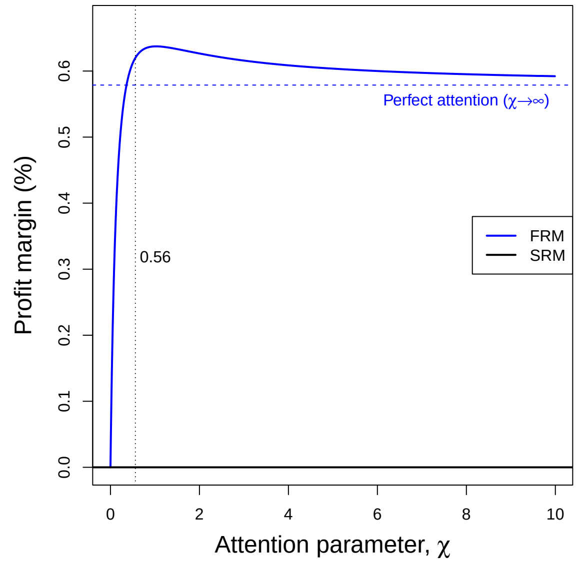
4.1 Prepayment Penalties

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.[45] This PPM-p presents the lenders a continuum of choices,

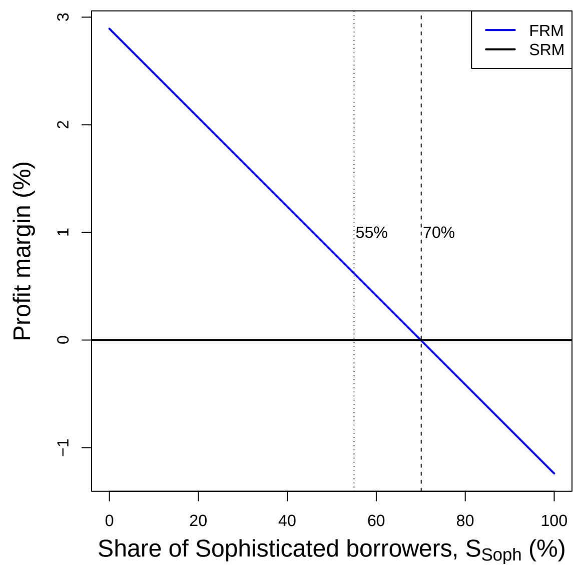
[45]Whether the penalty is also imposed when the borrower moves has no bearing on any outcomes of consequence. If the mortgage had only a "refinancing penalty," which were imposed upon a refinance but



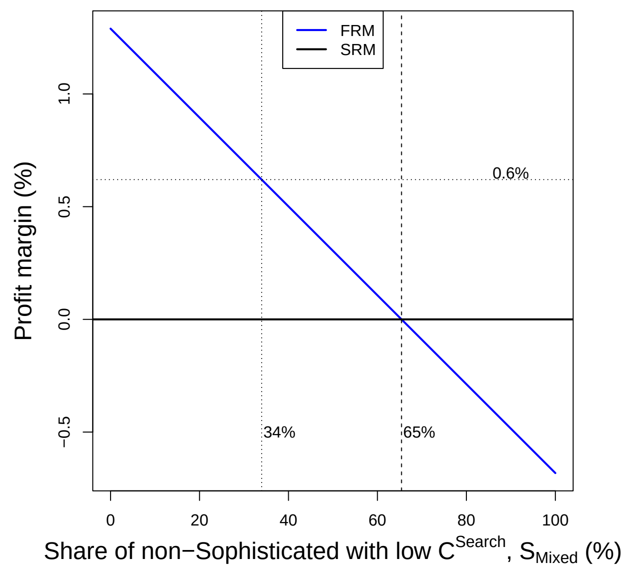
(a) Interest rate volatility.



(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 2: Comparative static of lender profit with respect to key model inputs using the calibration from Section 3.1 (without tax deductibility or amortization). For each parameter, the baseline level from Table 1 is shown with a vertical line. The horizontal blue line in Figure 2b shows the limit values as $\chi \to \infty$. With the exception of Figure 2b, 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 2d shows how the calibrated value of $S_{Mixed} = 34\%$ was chosen, by generating a profit margin of 0.62%.

as they get to choose 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 $\lim_{p\to\infty}$ PPM-p in this model. The two mortgages seem to be polar opposites, in that the SRM refinances without any effort, while the $\lim_{p\to\infty}$ 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.[46]

To see why, note that the interest rate on the $\lim_{p\to\infty}$ PPM-p will be $\lim_{p\to\infty} m_0^{\text{PPM-}p} = i_0 + v_{Soph}$, 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.[47] This works out because the $\lim_{p\to\infty}$ 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.

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

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).

[46] 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 $\lim_{p\to\infty}$ 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 between the two mortgages. But in the risk-neutral, unconstrained setting of this paper, this convenient equivalence holds.

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

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.

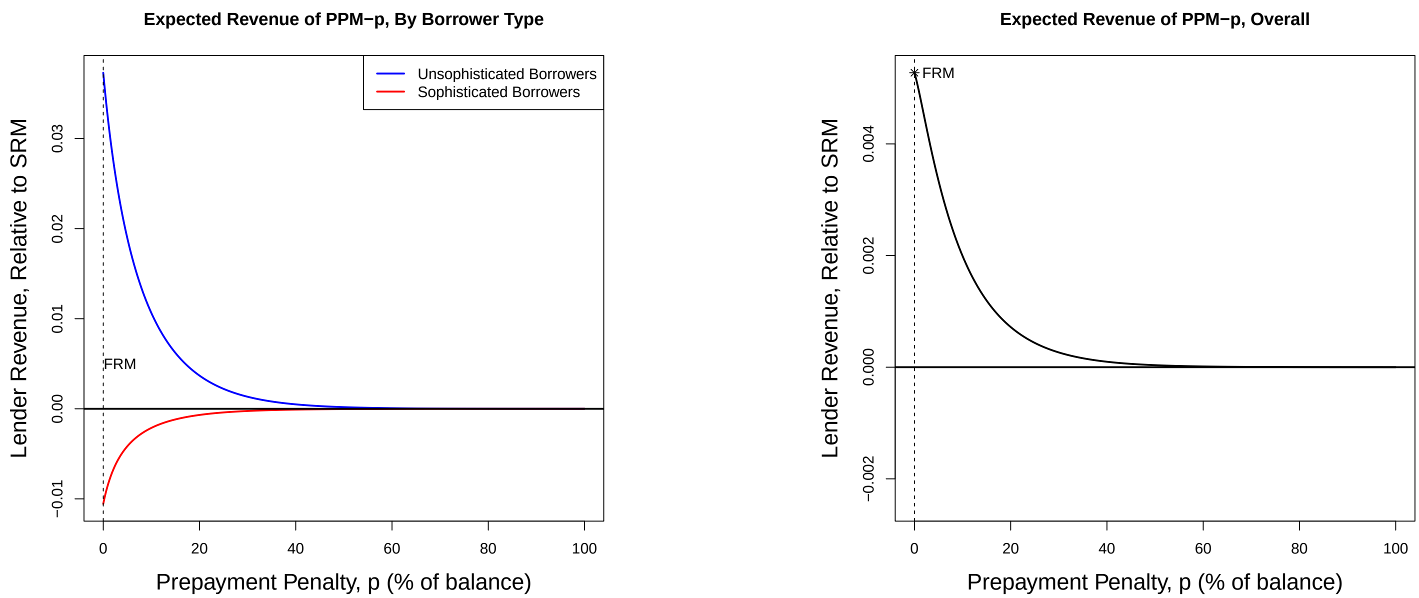
4.1.3 Baseline Calibration

Figure 3 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 3a shows the tradeoff discussed above: as the 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. Figure 3b shows that given this tradeoff, the profit-maximizing PPM-p is the PPM-0 – 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.[48] This is consistent with the very small share of mortgages with prepayment penalties observed in recent decades, a point expanded on below.[49]

[48] One interesting implication of these results is that lenders should offer prepayment "bonuses" rather than penalties (i.e. $p < 0$). By increasing refinancing among Sophisticated borrowers, this would allow lenders to ratchet up the interest rate and extract even more from Unsophisticated borrowers. This idea is expanded upon in Section 4.2, which shows how allowing borrowers to "buy negative points" can serve this purpose.

[49] One notable exception was the subprime market in the early 2000s, in which prepayment penalties were very common. Refinancing was extremely common among this group, but Mayer *et al.* (2013) point out that the incentive arose for a different reason than traditional refinancing. In the subprime market, refinancing often resulted not because of falling market interest rates but rather because the borrower's creditworthiness improved, allowing them to achieve a lower rate for that reason. Motivated by this, Mayer *et al.* (2013) present an interesting explanation of prepayment penalties as a form of risk-sharing: they lower the equilibrium interest rate, and then borrowers who experience a positive innovation in their creditworthiness will pay the prepayment penalty to refinance into a non-subprime mortgage. In their model, then, the prepayment penalty essentially serves to spread the benefits of improved creditworthiness to all borrowers. In that model, borrowers are *ex-ante* identical but experience idiosyncratic credit shocks *ex-post*. The model in the present paper, which has *ex-ante* differences among borrowers, provides a different potential rationale for the existence of prepayment penalties: prepayment penalties affect the tradeoff between deadweight loss caused by Sophisticated borrowers and extracting surplus from Unsophisticated borrowers.



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

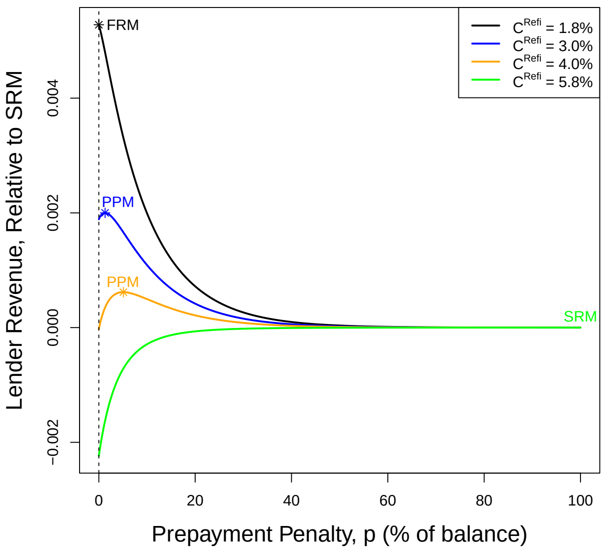
(b) Expected overall FRM revenue relative to SRM, as a function of prepayment penalty, p.

Figure 3: Analysis of prepayment penalty using the calibration from Section 3. A borrower must pay p every time she refinances or moves.

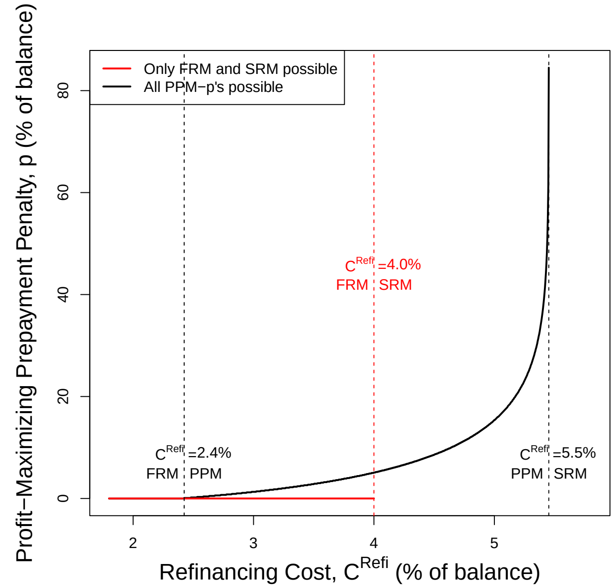
4.1.4 Comparative Statics

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 4 shows how this changes when the PPM is possible. Figure 4a 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.



(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 4: Analysis of prepayment penalty assuming different refinancing costs, C^{Refi}, using the calibration from Section 3. A borrower must pay p every time she refinances (or moves). The first black vertical line in Figure 4b 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 $C^{Refi} = 1.8\%$.

Figure 4b 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 4b present some nuance to that narrative by allowing prepayment penalties to be something that is endogenously produced by the market rather than exogenously 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 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 4b, as decreasing C^{Refi} leads to lower prepayment penalties and then, eventually, their disappearance.

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 will be 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.[50] 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,[51] 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 5 shows how this plays out in the model calibrated above. The mechanics of the mortgage offers are shown in Figures 5a-5b. When $v_{Uns} = \omega^{FRM}_{Soph}$, the Unsophisticated bor-

[50] 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.

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

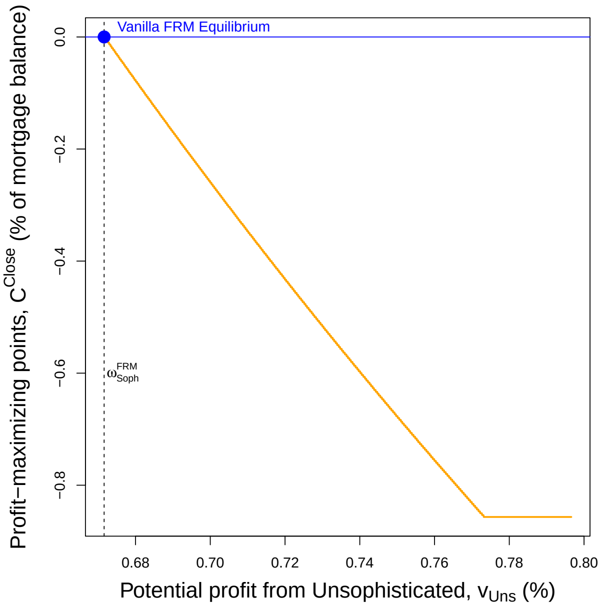
rower is paying her reservation interest rate and points cannot be used effectively. However, as v_{Uns} increases, the lender can extract more surplus by offering two mortgages with higher interest rates: one targeted to the Unsophisticated borrower with no points; and another targeted to the Sophisticated borrower with negative points and an even higher interest rate. As shown in Figure 5c, this allows the lender to modestly increase its profit margin above the 0.62% calibrated in 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, once $C^{Close} = -0.85\%$, further extraction is not practical as it becomes too costly due to the refinancing costs it induces. Notably, 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 5, is consistent with the real "rate sheet" shown in Figure 1 of Zhang (2023).[52]

Figure 5d shows the sensitivity of this analysis to the refinancing resource cost, C^{Refi}. In the neighborhood of the baseline level of 1.8%, the benefit of points is a decreasing function of C^{Refi}. As C^{Refi} falls, the additional deadweight refinancing cost associated with introducing points weakens and so the lender can benefit by aggressively decreasing C^{Close} to drive up the interest rates on the mortgages and extract more from the Unsophisticated borrower. When $C^{Refi} = 2.4\%$ – an increase of 0.6pp from the baseline – lenders cannot improve upon the vanilla FRM outcome by using points. At that level, $\omega^{FRM}_{Soph} = v_{Uns}$ and so the lender is extracting the maximum possible from the Unsophisticated borrower.[53] This point will be revisited in Section 5.2.

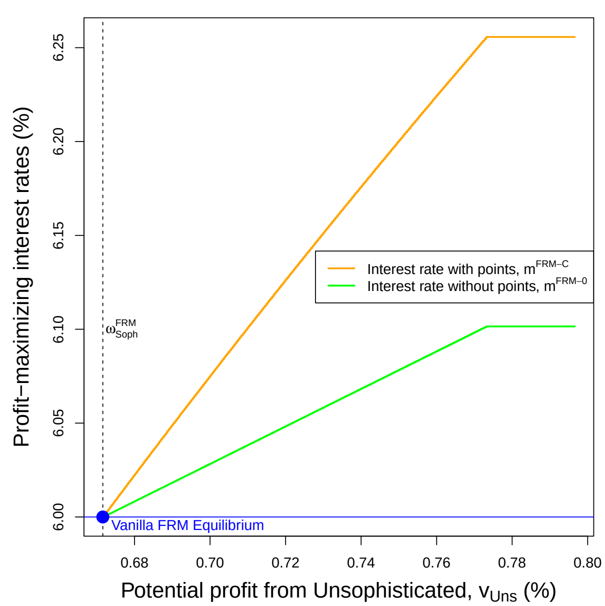
As Zhang (2023) has demonstrated, points exacerbate the efficiency and distributional impacts of the FRM, leading to higher cost for unsophisticated borrowers and more refinancing by sophisticated ones. The core difference in this paper is that this redistribution from unsophisticated borrowers goes to lenders rather than sophisticated borrowers, as in his work, since lender profits in my model are not pinned down by Perfect Competition. More importantly for the purposes of this paper, this analysis shows that the introduction of points

[52]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 (2023), 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.

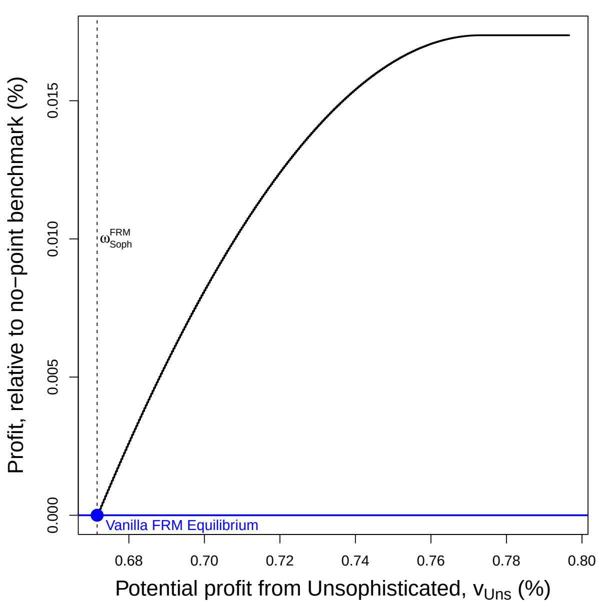
[53]For $C^{Refi} > 2.4\%$, the lender increases profit by charging positive points $C^{Close} > 0$ to the Sophisticated borrower as a way to reduce the deadweight loss of their refinancing. As discussed above, evidence in Woodward and Hall (2010), Woodward and Hall (2012), Agarwal *et al.* (2017), and Zhang (2023) find that borrowers who refinance will have lower (in this model, negative) points, which supports the validity of the condition that $v_{Uns} > \omega^{FRM}_{Soph}$, discussed in depth in Section 3.2.



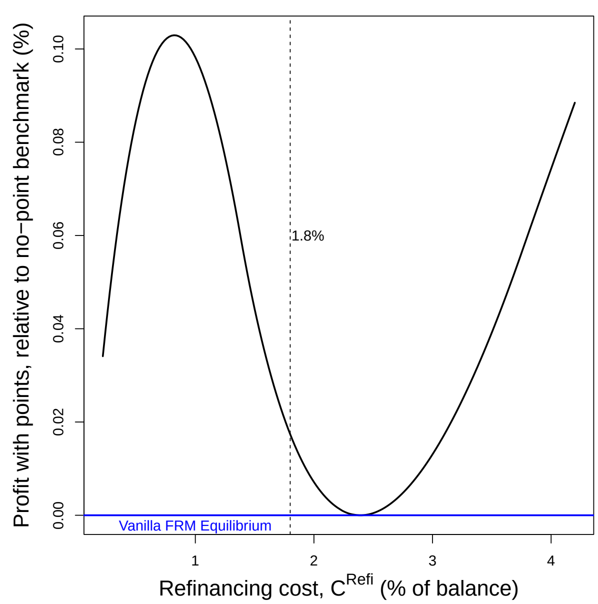
(a) Equilibrium points as v_{Uns} varies.



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



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



(d) Profit margin as refinancing resource cost varies, holding $v_{Uns} = 1\%$.

Figure 5: Analysis of points using the calibration from Section 3.1 (without tax deductibility or amortization). Figures 5a-5b show the points and interest rates on a menu of two mortgages for different values of Unsophisticated borrowers' search costs, as captured be 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 normalized to be zero. Figure 5c shows the profit level associated with this approach. Figures 5a-5c show values of v_{Uns} greater than ω^{FRM}_{Soph} so that the conditions of Proposition 1 are met. Panel 5d hold v_{Uns} fixed at 1% and shows how the excess profit from introducing points varies with the resource cost of refinancing, C^{Refi}.

can further bolster the FRM as the mortgage design of choice for lenders over the SRM. The seed of price discrimination within the vanilla FRM equilibrium can be amplified by points, as demonstrated above. 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: A Refinancing Tax

Previous papers studying the failure of unsophisticated borrowers to refinance FRMs have 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.

To aid unsophisticated borrowers, the model of this paper recommends a policy that may be somewhat surprising: tax borrowers who actively refinance their mortgages. (To be clear, if the interest rate falls automatically, as with an SRM, that would not be taxed.)[54] 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 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

[54]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.

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.

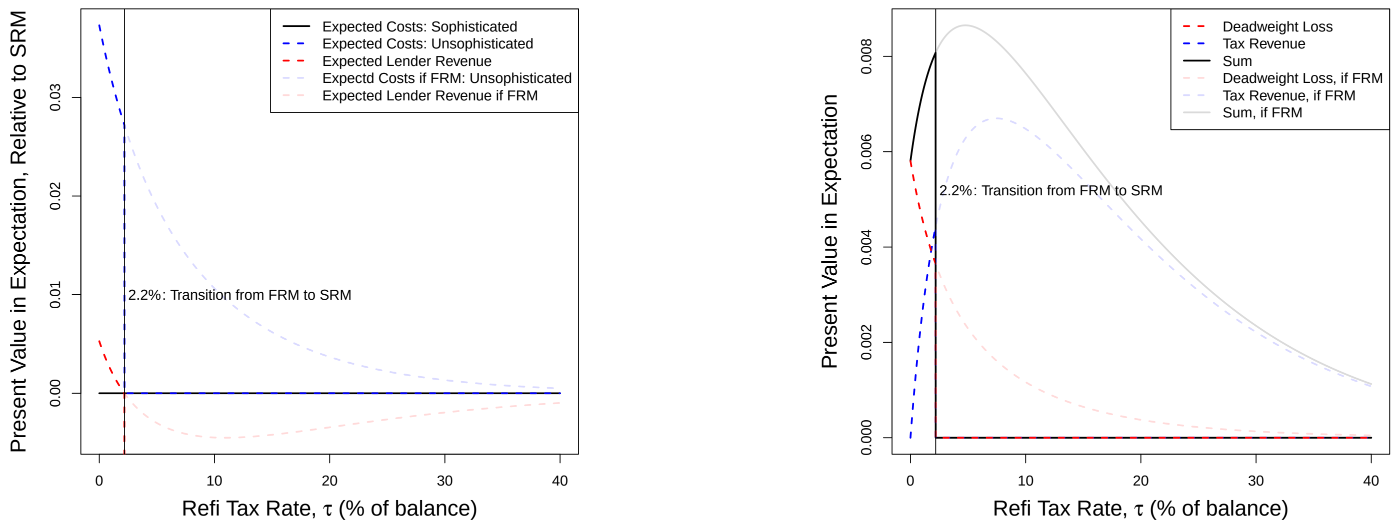
5.1 Refinancing Tax when PPM-p Is Not Possible

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 6 demonstrates the impact of a refinancing tax using the calibration presented in Section 3. The tax is implemented by requiring the borrower to pay τ to the government each time she refinances. Figure 6a 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.[55] Figure 6b shows that the tax does lower deadweight loss, but this is initially swamped by the rise in government 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.

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.

[55] 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.

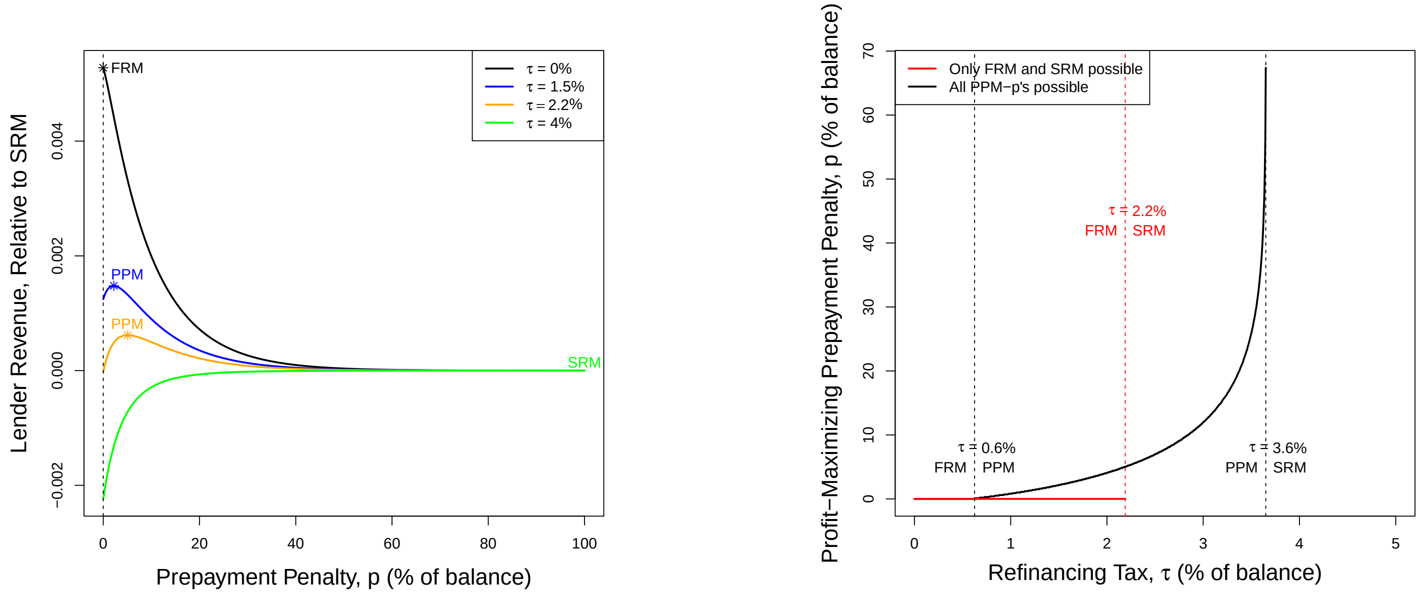
Figure 6: Analysis of a refi tax using the calibration discussed in Section 3. A borrower must pay τ every time she refinances.

5.2 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 7a shows lender revenue as a function of the prepayment penalty for different levels of the refinancing tax. Critically, 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 to SRM could instead lead to the adoption of a PPM. Figure 7b 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%,[56] 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.

Suppose the policymaker did set the tax to, say, 2.2% and so the market transitioned to a PPM rather than a SRM; would this be a bad policy? It depends on the policymaker's

[56] Recall from Section 4.2 that an increase of 0.6pp in the refinancing cost is needed in order to cause the lender to switch from setting negative points to positive points (i.e. a prepayment penalty).



(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 7: Analysis of prepayment penalty assuming different refinancing taxes, using the calibration from Section 3. 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 7b 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.

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. Both the tax itself and the lenders' resulting prepayment penalties would lower refinancing, which 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 the PPM would offer less payment relief when interest rates decline than did the FRM, since there will be less refinancing. So while all of these goals are promoted by a SRM, whether the PPM is preferable to the FRM depends on whether one is primarily concerned with macroprudential issues or equity and efficiency issues in steady state.

6 Discussion

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 Sections 2.3 and 3.2, in the limiting case of Perfect Competition, all borrowers will end up with the desirable SRM. The 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. For these reasons, the answer from the person on the street may not be compelling to an economist.

This paper has presented a model that confronts these two critiques and is able to generate a no-SRM equilibrium and restore the explanation of the person on the street. There are two key elements that drive the result. The first is some ability of lenders to extract rents – in this case, because of borrower search costs – that gives the lender the ability to earn a profit and withhold a product that the borrower wants, up to an extent. Borrower heterogeneity is the second key element. Together, they lead to an equilibrium in which lenders benefit from Unsophisticated borrowers' failure to refinance their FRMs. Unsophisticated borrowers accept the FRMs because it is too difficult to find an alternative, and lenders cannot do this heavy extraction from Unsophisticated borrowers with an expensive SRM because Sophisticated borrowers would not accept that. Ultimately, then, the model basically agrees with the person on the street: the appeal of the FRM to the lender is that Unsophisticated borrowers do not refinance them, which creates a windfall that the SRM would eliminate. 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 the observed mortgage design.

Working through the model, however, provides additional insights that the person on the street was unlikely to think of. The key one, stressed throughout the paper, is that while the FRM does extract more from Unsophisticated borrowers, it does so through a mechanism that requires deadweight costs from Sophisticated borrowers. This lowers the profit extracted from them and therefore creates a tradeoff between the FRM and SRM, the latter of which is efficient. More subtly, as explored throughout Section 4, lenders can use prepayment penalties and points to more deftly navigate this tradeoff. The calibrations of Sections 3 and 4 show that, taking all of these issues into consideration, the FRM generates more revenue for lenders than a SRM or any mortgage with positive prepayment penalties. Offering a menu of of FRMs with varying levels of points can further amplify this profitable mechanism.

Zooming out from the details of the model and its results, the highest-level goal of this paper was to write a model that has an FRM mortgage design as an output, rather than an input. As discussed in Section 1, it has been standard in the literature to assess the distributional consequences of heterogeneous refinancing using models that exogenously restrict the mortgage space to only an FRM. This restriction is crucial because it has also been standard to assume assume a zero-profit condition, which, as shown in Section 2.3, would lead to the unraveling of the FRM market if mortgage design were endogneous. It is instructive to consider how the conclusions of the present paper, which includes search costs and endogenous mortgage design, agree with and differ from a zero-profit, exogenous-FRM model.

In a model with zero profits and an exogenous FRM mortgage design, Unsophisticated borrowers would be better off if borrower types were revealed, because then lenders would charge them a lower interest rate to reflect their low likelihood of refinancing. Similarly, Sophisticated borrowers would be worse off. In the model with search costs presented in this paper, Unsophisticated borrowers would actually be worse off if their types were revealed, because lenders would then extract a large profit margin from them, knowing they would not search for a better alternative. More broadly, lenders are essentially a "veil" in the zero-profit model, and so all redistribution happens between borrowers; in the present paper's model with the potential for rents, on the other hand, lenders play a more active role and the key distributional margin is between Unsophisticated borrowers and lenders. This is true not only of the impact of the abstract possibility of revealing borrower types, but also of prepayment penalties, points, and the refinancing tax explored in Section 5, which all impact how much the lender can capture and how much is extracted from Unsophisticated borrowers.

This segues into a discussion of policy, where the two models have some similarities and some differences. In both models, a refinancing tax would increase efficiency and lower the mortgage interest rate, to the benefit of Unsophisticated borrowers. The models differ in that the windfall to the Unsophisticated borrowers comes at the expense of Sophisticated borrowers in the zero-profit model, as their shrinking likelihood of refinancing brings their expected costs closer to the opportunity cost of capital. In contrast, as discussed in Section 5, the Unsophisticated borrowers' gains come at the expense of the lenders in the model with search costs, as the lower interest rate limits how much revenue can be extracted from them. Nonetheless, both models show that a refinancing tax would enhance efficiency and benefit Unsophisticated borrowers. In contrast to that similarity, the biggest difference is that, by definition, the model with exogenous mortgage design does not speak to how policy can change the structure of mortgages, whereas the analysis in Section 5 shows that when mortgage design is endogenous, the refinancing tax can reach a level at which lenders introduce the SRM, further benefiting Unsophisticated borrowers and enhancing efficiency.

To conclude, this paper has presented a model in which a SRM is possible and yet is not offered, despite the fact that some borrowers are hurt by their failure to refinance their FRMs and the remaining borrowers must pay a cost to do so. Whether the reader finds this particular model convincing or thinks some other explanation is at work, explicitly modeling the design of mortgages (and likely other products, financial and otherwise) is an important avenue for future research. While models with exogenous mortgage design can potentially demonstrate relative advantages of various products, they cannot directly answer the policy question of how to move from one to another. A model that cannot explain why the SRM is absent in the first place can offer limited guidance in how to encourage its adoption. Hopefully, this paper has made some progress on that front.

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.

—, 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.

Beltratti, A., Benetton, M. and Gavazza, A. (2017). The Role of Prepayment Penalties in Mortgage Loans. *Journal of Banking and Finance*, **82**, 165–179.

Beraja, M., Fuster, A., Hurst, E. and Vavra, J. (2019). Regional Heterogeneity and the Refinancing Channel of Monetary Policy. *Quarterly Journal of Economics*, **134** (1), 109–183.

Berger, D., Milbradt, K., Tourre, F. and Vavra, J. (2023a). *Optimal Mortgage Refinancing with Inattention*. Technical Report.

—, —, — and — (2023b). *Refinancing Frictions, Mortgage Pricing and Redistribution*. Working Paper.

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. (2024). Paying Too Much? Borrower Sophistication and Overpayment in the US Mortgage Market. *Federal Reserve Bank of Philadelphia Working Paper*.

— and Keys, B. J. (2016). Interest rates and equity extraction during the housing boom. *American Economic Review*, **106** (7), 1742–1774.

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

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

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

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

— and Cocco, J. F. (2003). Household Risk Management and Optimal Mortgage Choice. *The Quarterly Journal of Economics*, **118** (4), 1449–1494.

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

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

Chen, H., Michaux, M. and Roussanov, N. (2019). Houses as ATMs? Mortgage Refinancing and Macroeconomic Uncertainty. *Journal of Finance*, **75** (1), 323–375.

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

Di Maggio, M., Kermani, A., Keys, B. J., Piskorski, T., Ramcharan, R., Seru, A. and Yao, V. (2017). Interest Rate Pass-Through: Mortgage Rates, Household Consumption, and Voluntary Deleveraging. *American Economic Review*, **107** (11), 3550–3588.

—, — and Palmer, C. (2020). How quantitative easing works: Evidence on the refinancing channel. *The Review of Economic Studies*, **87** (3), 1498–1528.

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

Ehrlich, G. and Perry, J. (2015). *Do Large-Scale Refinancing Programs Reduce Mortgage Defaults? Evidence From a Regression Discontinuity Design*. Working Paper Series 2015-06, CBO.

Elliehausen, G. E., Staten, M. E. and Steinbuks, J. (2008). The effect of prepayment penalties on the pricing of subprime mortgages. *Journal of Economics and Business*, **60**, 33–46.

Enkhbold, A. (2023). *Monetary Policy Transmission, Bank Market Power, and Wholesale Funding Reliance*. Bank of Canada Staff Working Paper.

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

Flesaker, B. and Ronn, E. I. (1993). The Pricing of FIREARMs ("Falling Interest Rate Adjustable-Rate Mortgages"). *Journal of Real Estate Finance and Economics*, **6**, 251–275.

Fuster, A., Lo, S. and Willen, P. (forthcoming). The Time-Varying Price of Financial Intermediation in the Mortgage Market. *Journal of Finance.*

— and Willen, P. S. (2017). Payment Size, Negative Equity, and Mortgage Default. *American Economic Journal: Economic Policy*, **9** (4), 167–191.

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.

Greenspan, A. (2004). Testimony before the committee on financial services, u.s. house of representatives, february 11, 2004.

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

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

Ho, G. and Pennington-Cross, A. (2008). Predatory lending laws and the cost of credit. *Real Estate Economics*, **36** (2), 175–211.

Hurst, E. and Stafford, F. (2004). Home Is Where the Equity Is: Mortgage Refinancing and Household Consumption. *Journal of Money, Credit, and Banking*, **36** (6), 985–1014.

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.

Karamon, K., McManus, D. and Zhu, J. (2016). Refinance and Mortgage Default: A Regression Discontinuity Analysis. *The Journal of Real Estate Finance and Ecnonomics*, pp. 1–19.

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. (2023). *Regional Variation in Transaction Costs, Mortgage Rate Heterogeneity, and Mortgage Refinancing Behavior*. Working Paper, FDIC.

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. (2023). *Private Information and Price Regulation in the US Credit Card Market*. Working Paper.

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. (2016). *Market Power in Mortgage Lending and the Transmission of Monetary Policy*. Working Paper, Harvard University.

Tracy, J. and Wright, J. (2016). Payment Changes and Default Risk: The Impact of Refinancing on Expected Credit Losses. *Journal of Urban Economics*, **93**, 60–70.

Wang, Y., Whited, T. M., Wu, Y. and Xiao, K. (2022). Bank Market Power and Monetary Policy Transmission: Evidence from a Structural Estimation. *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. (2023). *Closing Costs, Refinancing, and Inefficients in the Mortgage Market.* Working Paper.

Zhu, J., Janowiak, J., Ji, L., Karamon, K. and McManus, D. (2015). The effect of mortgage payment reduction on default: Evidence from the home affordable refinance program. *Real Estate Economics*, **43** (4), 1035–1054.

Appendix

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:

$$\begin{aligned} K^{FRM}(m_0^{FRM}, 0) &= \int_0^{\infty} m_0^{FRM} \cdot exp(-(\rho+\mu+\pi)\cdot t)\cdot dt + \int_0^{\infty} \mu \cdot exp(-\mu \cdot t)\cdot exp(-(\rho+\pi)\cdot t)\cdot dt \\ &= \frac{m_0^{FRM}}{\rho+\mu+\pi} + \frac{\mu}{\rho+\mu+\pi} \\ &= \frac{m_0^{FRM}+\mu}{\rho+\mu+\pi}. \end{aligned} \tag{19}$$

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:

$$K^{FRM}(m_0^{FRM}, \chi) = K^{FRM}(m_0^{FRM}, 0) + \Omega^{FRM}(0, \chi). \tag{20}$$

To solve for the option value, note:

$$\begin{aligned} \Omega^{FRM}(y, \chi) = \min_{y^*(\chi)} E\Big[\int_0^{\infty} exp(-(\rho+\mu+\pi)\cdot t)\cdot \big(C^{Refi} + \frac{y_{t-}}{\rho+\mu+\pi}\big)\cdot A_t \cdot dt\Big] \\ \text{s.t. } dy_t = \sigma \cdot dz_t - y_{t-} \cdot A_t, \end{aligned} \tag{21}$$

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:

$$\begin{aligned}\rho\cdot\Omega^{FRM}(y,\chi) &= -(\mu+\pi)\cdot\Omega^{FRM}(y,\chi)+\frac{E[d\Omega^{FRM}(y,\chi)]}{dt}\\ (\rho+\mu+\pi)\cdot\Omega^{FRM}(y,\chi) &= \frac{\sigma^2}{2}\cdot\frac{\partial^2\Omega^{FRM}(y,\chi)}{\partial y^2}+\chi\cdot\min\big(0,\Omega^{FRM}(0,\chi)+C^{Refi}+\frac{y}{\rho+\mu+\pi}-\Omega^{FRM}(y,\chi)\big)\\ (\rho+\mu+\pi+\chi)\cdot\Omega^{FRM}(y,\chi) &= \frac{\sigma^2}{2}\cdot\frac{\partial^2\Omega^{FRM}(y,\chi)}{\partial y^2}+\chi\cdot\min\big(\Omega^{FRM}(y,\chi),\Omega^{FRM}(0,\chi)+C^{Refi}+\frac{y}{\rho+\mu+\pi}\big).\end{aligned} \tag{22}$$

This differential equation is solved by:

$$\Omega^{FRM}(y,\chi)=\begin{cases}J_\psi\cdot exp\big(-\psi\cdot(y-y^*(\chi))\big) & \text{if } y\geq y^*(\chi)\\ J_\eta\cdot exp\big(\eta\cdot(y-y^*(\chi))\big)+\frac{\chi}{\rho+\mu+\pi+\chi}\cdot\big(J_\psi\cdot exp(\psi\cdot y^*(\chi))+C^{Refi}+\frac{y}{\rho+\mu+\pi}\big) & \text{if } y\leq y^*(\chi)\end{cases} \tag{23}$$

Plugging this into Equation 22 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:

$$J_\psi=J_\eta+\frac{\chi}{\rho+\mu+\pi+\chi}\cdot\big(J_\psi\cdot exp(\psi\cdot y^*(\chi))+C^{Refi}+\frac{y^*(\chi)}{\rho+\mu+\pi}\big) \tag{24}$$

and

$$-\psi\cdot J_\psi=\eta\cdot J_\eta+\frac{\chi}{(\rho+\mu+\pi)\cdot(\rho+\mu+\pi+\chi)}. \tag{25}$$

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:

$$\begin{aligned}\Omega^{FRM}(y^*(\chi),\chi) &= \Omega^{FRM}(0,\chi)+C^{Refi}+\frac{y^*(\chi)}{\rho+\mu+\pi}\\ J_\psi &= J_\psi\cdot exp(\psi\cdot y^*(\chi))+C^{Refi}+\frac{y^*(\chi)}{\rho+\mu+\pi}.\end{aligned} \tag{26}$$

Equations 24-26 are three equations in three unknowns, and as shown by Berger *et al.* (2023a), there is a unique solution with $y^*(\chi) < 0$.

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

$$\begin{aligned}\Omega^{FRM}(0,\chi) &= J_\psi \cdot exp(\psi \cdot y^*(\chi)) \\ &= \frac{\chi \cdot \left(\eta \cdot \left(\frac{y^*(\chi)}{\rho+\mu+\pi} + C^{Refi}\right) - \frac{1}{\rho+\mu+\pi}\right)}{(\rho+\mu+\pi+\chi)\cdot(\eta+\psi)\cdot exp(-\psi\cdot y^*(\chi)) - \chi\cdot\eta}.\end{aligned} \tag{27}$$

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:

$$K^{FRM}(m_0^{FRM},\chi) = P^{FRM}(m_0^{FRM},\chi) + D^{FRM}(\chi), \tag{28}$$

where, based on Equations 19, 20, and 27:

$$P^{FRM}(m_0^{FRM},\chi) = \frac{m_0^{FRM}+\mu}{\rho+\mu+\pi} + \frac{\chi\cdot\left(\eta\cdot\frac{y^*(\chi)}{\rho+\mu+\pi} - \frac{1}{\rho+\mu+\pi}\right)}{(\rho+\mu+\pi+\chi)\cdot(\eta+\psi)\cdot exp(-\psi\cdot y^*(\chi)) - \chi\cdot\eta} \tag{29}$$

and

$$D^{FRM}(\chi) = \frac{\chi\cdot\eta\cdot C^{Refi}}{(\rho+\mu+\pi+\chi)\cdot(\eta+\psi)\cdot exp(-\psi\cdot y^*(\chi)) - \chi\cdot\eta}. \tag{30}$$

As $y^*(\chi)$ is produced as the solution to Equations 24-26, Equations 29 and 30 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 $C^{Refi} = 0$.

As further justification that $D^{FRM}(\chi)$ represents the expected PDV of deadweight loss, note that we can solve for $D^{FRM}(\chi)$ directly in the limiting case of perfect attention, i.e. when $\chi \to \infty$. In that case:

$$\lim_{\chi \to \infty} D^{FRM}(\chi) = \int_0^\infty \big(C^{Refi} + \lim_{\chi \to \infty} D^{FRM}(\chi)\big) \cdot exp(-(\rho + \mu + \pi) \cdot t^*) \cdot f(t^*) dt^*, \tag{31}$$

where t^* is the first time that $y_t = \lim_{\chi \to \infty} y^*(\chi) \equiv y^*_{lim}$, and $f(t^*)$ is its PDF. Therefore:

$$\lim_{\chi \to \infty} D^{FRM}(\chi) = C^{Refi} \cdot \frac{J}{1 - J}, \tag{32}$$

where $J \equiv \int_0^\infty exp(-(\rho + \mu + \pi) \cdot t^*) \cdot f(t^*) dt^*$.

To find J, note that because y_t is a Brownian Motion with $y_t \sim N(0, \sigma^2 \cdot t)$, it follows that $Pr(t^* < t) = 2 \cdot \Phi\Big(\frac{y^*_{lim}}{\sigma \cdot \sqrt{t}}\Big)$, where $\Phi()$ is the CDF of a Standard Normal random variable. As a result:

$$f(t) = -\frac{y^*_{lim}}{\sigma} \cdot t^{-3/2} \cdot \phi\Big(\frac{y^*_{lim}}{\sigma \cdot \sqrt{t}}\Big), \tag{33}$$

where $\phi()$ is the PDF of a Standard Normal random variable. Therefore:

$$\lim_{\chi \to \infty} D^{FRM}(\chi) = -\frac{y^*_{lim} \cdot C^{Refi}}{\sigma} \cdot \frac{\int_0^\infty exp(-(\rho + \mu + \pi) \cdot t) \cdot t^{-3/2} \cdot \phi\Big(\frac{y^*_{lim}}{\sigma \cdot \sqrt{t}}\Big) dt}{1 + \frac{y^*_{lim}}{\sigma} \cdot \int_0^\infty exp(-(\rho + \mu + \pi) \cdot t) \cdot t^{-3/2} \cdot \phi\Big(\frac{y^*_{lim}}{\sigma \cdot \sqrt{t}}\Big) dt}. \tag{34}$$

This is the direct approach. The "indirect approach" derived from the solution to the optimization problem, shown in Equation 30, simplifies to the following in the case of perfect attention:

$$\lim_{\chi \to \infty} D^{FRM}(\chi) = \frac{C^{Refi}}{exp(-\psi \cdot y^*_{lim}) - 1}. \tag{35}$$

Finally, note that in this limiting case, $y^*_{lim} = -\frac{1}{\psi} \cdot \Big[1 + \psi \cdot (\rho + \mu + \pi) \cdot C^{Refi} + W(-exp(-(1 + \psi \cdot (\rho + \mu + \pi) \cdot C^{Refi})))\Big]$, where $W()$ is the principal branch of Lambert's W-function, as shown by Agarwal *et al.* (2013).

While an equivalence between Equations 34 and 35 is not obvious, numerical checks confirm that they align. This is possible because of the connection between Lambert's W and the Normal distribution: Lambert's W is part of the approximate characteristic function of the log-normal distribution.

Proof of Proposition 1

Assume, as stipulated in Proposition 1, 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:

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

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:

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

Therefore:

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

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:

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

which is Expression 13 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 13 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 13 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 13 does not hold.

In symmetric equilibria in which all lenders offer only a FRM at $\bar{m}^{FRM}_{Soph}$ when Expression 13 holds and only a SRM at $\bar{m}^{SRM}_{Soph}$ when Expression 13 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 13 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_t^{FRM} will follow a Brownian motion, as assumed throughout the paper.

Similarly, if Expression 13 does not hold, the SRM will be observed, with interest rate $m_t^{SRM} = i_t + \Delta^{SRM}$, where $\Delta^{SRM} = \upsilon_{Soph} + \omega^{SRM}$ is a constant wedge. Therefore, since i_t follows a Brownian motion, m_t^{SRM} will follow a Brownian motion, as assumed throughout the paper.

Proof of Proposition 2

Assume, as stipulated in Proposition 2, 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 $(i_t + 1/\psi)$ will cover the lender's opportunity cost of capital:

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

Because $\bar{m}_{Uns}^{SRM} > \bar{m}_{Soph}^{SRM} = i_t + \upsilon_{Soph} + 1/\psi$ 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 $\upsilon_{Soph} \geq 0$. Such a choice will lead to expected profit of:

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

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

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

Therefore:

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

where the final step follows from the fact that $\bar{m}_{Soph}^{FRM}$ and $\bar{m}_{Soph}^{SRM}$ 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:

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

which is Expression 14 shown in Proposition 2.

If Expression 14 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:

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

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 14 holds, an all-FRM equilibrium is not possible.

If Expression 14 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:

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

where the first inequality comes from definition of reservation interest rates and the fact that $\bar{m}^{FRM}_{Uns} < \bar{m}^{FRM}_{Soph}$; 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 14 does not hold. Therefore, if Expression 14 holds, an all-FRM equilibrium is not possible.

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

Tax Deductibility

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 T be the household's marginal tax rate, let N 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 $\theta \approx 0.21$,[57] though this is an approximation because the true hazard varies over time. Then, define:

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

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 $\frac{1}{1-T}$ term, which increases the effective cost of refinancing) as well as the fact that the ability to spread a deduction of closing costs ($T \cdot C^{Refi}$) over the remaining course of the mortgage (N 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} in all instances, and the results of the model go through.

Following Agarwal *et al.* (2013), $T = 0.28$ and $N = 30$. All other parameters take on the same values as used throughout the paper, shown in Table 1.

[57]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 $\mu = 0.10$, $\theta \approx 0.21$.