Reinforcement learning for household finance: designing policy via responsiveness

Arka Prava Bandyopadhyay * Lilia Maliar †

January 17, 2024

Abstract

We use model-free reinforcement learning (RL) to investigate how a mortgage servicer can optimize her actions towards a borrower. Our methodology differs from the conventional heuristic approach, since we do not use subjective and qualitative judgments of industry and legal experts. We are the first to exploit the borrower's soft information post-securitization and her responsiveness to the servicer, to estimate an RL-policy rule. When maximizing her reward, the servicer learns the borrower's type dynamically. By doing so, the servicer can preempt the borrower's adversarial behavior, thereby increasing the borrower's cooperation.

Keywords: Reinforcement Learning, Household Finance, Soft Information, Moral Hazard, Experience Learning

JEL Codes: G2, G3, G4, G5, R2

*Bandyopadhyay: New York University, apb321@nyu.edu

†Maliar: Graduate Center, CUNY; Hoover Institution, Stanford University, lmaliar@gc.cuny.edu.
We thank Ben Keys, Michael Weber, Manisha Padi, Parag Pathak, Stefano Giglio, Nick Barberis, Nathaniel Pattison, David Zhang, Manuel Santos, Edmund Crawley, Brent Ambrose, Michael Seiler, Andrew Sanderford, Rick Peiser, Igor Halperin, Sean Flynn, Alberto Teguia, Gordon Ritter, Petter Kolm and Econometrics Society Summer 2022 conference participants for helpful comments.

1 Introduction

Asymmetric information between debt servicers and borrowers stands in the way of efficient contract modifications and hence it requires policy intervention. Debt servicers have incomplete information about individual borrowers, causing inefficiently low levels of renegotiation when borrowers experience payment shocks. Contract renegotiation happens more often during extreme economic circumstances like during the 2008 Great Financial Crisis (GFC) and the 2020 COVID pandemic. For example, after the 2008 GFC, the policymakers have attempted to give incentives for servicers to gather information dynamically from borrowers through Home Affordable Modification Program (HAMP).[1] However, the proportion of people who availed themselves of such policies was low, namely, one-third, as documented in Sumit Agarwal (2017). This demonstrates how challenging it is for servicers to optimally gather information and to provide targeted relief. Given a low rate of availing the above targeted policy, during the 2020 COVID pandemic, the policymakers implemented a blanket forbearance - a policy that allowed a borrower to skip mortgage payments without any justification. Blanket forbearance had unintended consequences however; see Matvos (2021) who documents the salience of forbearance relative to the default rate at the onset of the CARES Act.

In this paper, we propose a novel quantitative solution to the problem of asymmetric information between a borrower and servicer.[2] Specifically, we show how the servicer can use soft information about borrower's current circumstances and thus, can provide targeted relief vis-a-vis the most efficient contract modification.

We use the state-of-the-art reinforcement learning (RL) technique to design RL-optimal intervention policies. To the best of our knowledge, we are the first to do so. The RL-policy of the servicer maximizes the overall profit through the life of the loans. The servicer uses the quantified soft information about the borrowers from their communications (call transcripts) with the borrowers. Soft information in this context is the unstructured text of the call transcripts of the communication between the borrowers and the servicers. We are the first to use soft information post the securitization of the loans. With the updated soft information, the servicer can

[1]HAMP: https://home.treasury.gov/data/troubled-assets-relief-program/housing/mha/hamp

[2]The servicer handles the day-to-day tasks of managing the loan by processing loan payments, responding to the borrower's inquiries, keeping track of principal and interest payments and managing escrow accounts or initiating foreclosure.

alleviate the information asymmetry with the borrower. In addition, we have created a novel measure of the borrower's responsiveness. Using this measure, we learn the noisy borrower's type over time which enables the servicer to provide targeted and efficient contract modification.

An RL-policy is the best course of action, assuming the borrower is an adversary agent. Goodfellow (2014) provides a discussion on how to incorporate adversarial behavior in a machine learning framework. The borrower's interest may not always be aligned with that of the servicer or eventually the lender. RL incorporates adversarial borrower's strategic default choice and updates this choice in a feedback loop via the borrower's idiosyncratic attributes, a.k.a. the noisy *environment*. Consistent with the RL literature, we simulate the noisy environment using proprietary data about the borrower's spending habits, demography, income bracket, real-time unemployment status, ethnicity, marital status, short-term liabilities, and net worth.

Monitoring the borrower's behavior in real-time and making informed decisions are challenging and costly tasks for servicers. Mortgage servicers anecdotally rely on the carrot versus stick approach for monitoring borrowers, with the *carrot* being a reward for payment on time and the *stick* being a consequence for noncompliance. This is because servicers do not know how *responsive* the borrowers are and hence undertake either harsh of lenient actions based on their subjective expert judgement. Strategically, the servicers would prefer to know which borrowers to reach out to, by making outbound calls, and which borrowers to respond to, after the borrower initiates the communication via inbound calls, to increase the productivity of the curative process. Cooperation from the borrower's side enables the final resolution, e.g., servicers can negotiate with cooperative borrowers and offer a loan modification, thereby preempting bankruptcy chapter 13. Servicers can save valuable time by excluding those borrowers who are historically less cooperative and will surely go a bankruptcy/foreclosure route. Moreover, for bankruptcy proceedings, cooperation and responsiveness can determine the bankruptcy outcome, i.e., chapter 13 (restructuring) versus chapter 7 (liquidation); see Fay, Hurst, and White (2002) for a discussion of bankruptcy chapters. We optimize over the servicer's unrestricted (global) set of actions and find that the actions inferred from this model-free RL approach of learning from experience are better than when optimized over a restricted harsh ("sticks") or lenient ("carrots") set of actions.

In this paper, we resort to the "reward" terminology from the RL literature. It would be preferable to convert the rewards to loan yields. However, we do not have the data for a loan's entire life.[3] We have an access to soft information about the borrower-servicer communication and hard information only when the loans are in the servicer's books. For a large lender, using this soft information may violate the 14th amendment due to procedural reasons. For example, suppose the lender treats loans differently based on the borrower's responsiveness. In this case, the borrower could sue for discrimination if their ability to call back is hampered due to full-time or part-time employment status. Hence, our data are post-securitization, where the servicer has the flexibility to use the quantified soft information and make decisions based on the borrower's responsiveness from the borrower's communications.

The current methodology, used in practice, is based on the servicer's qualitative due diligence process. This sequential process is as follows: information related to a title, foreclosure, bankruptcy, and servicing comments are received and processed by a servicer. Compliance data are extracted from collateral files. Then, title, property and legal due diligence tasks are created, assigned, and completed. After that, combined legal grades are determined and exception reports are created. Then, the seller negotiations occur and the final loan/funding schedules are made and contracts are signed. As one can see, the above processes are heuristic and difficult to automate.

The legal grades formed by combining the above sequential information are qualitative. An overly conservative legal grade pushes the servicer out of every trade and an overly aggressive legal grade results in undersized returns. The grades reflect the likelihood of loss, as well as the time/cost and complexity involved in addressing the concerns. The five major grades and their impact on pricing are determined based on levels of risk and discounts applied. Grade A refers to a non-issue from a risk standpoint, and discount is not applied. Grade B has little to no material risk of loss; issues are of low complexity or covered by valid insurance and a small discount is applied. Grade C represents moderate risk of loss, hence a significant discount is used (10-25%). Grade D is likely to require litigation or significant time/cost expenditures to resolve and hence, comes with a substantial discount (50-90%). Grade E is nearly certain to result in a complete loss.

We train the model on monthly data from 09/2017 to 11/2019 and conduct cross-

[3] Yield is a well understood measure of return, analogous to reward in the RL paradigm. If we had the loan performance data for every individual mortgage, we could have computed the yields of each loan. Then, the RL algorithm would have used the yield as the reward instead of the cumulative overall dollar return.

validation. We test our optimal policy's accuracy (plausibility and impact) using out-of-sample testing on pre-COVID data on loans from 12/2019 to 03/2020. Finally, we compare this novel optimal policy to the current ad-hoc qualitative methodology used by the servicer, based on reward. The clear dollar difference in collections vis-a-vis our RL-policy over current servicer action, based on the above heuristic approach, provides direct evidence of the efficiency of our quantitative approach.

The rest of the paper is organized as follows: Section 2 reviews the related literature. Section 3 describes delinquency states and possible servicer actions and describes the proprietary household level variables used. In Section 4, we define a novel time-invariant measure of the borrower's responsiveness; we list a few variables that affect the borrower's responsiveness; finally, we document how the borrower's responsiveness can lead to higher cooperation. In Section 5, we specify the details of RL. In Section 6, we present the numerical results, as well as sensitivity experiments. In Section 7, we discuss possible extensions of our paper and put it into a perspective of broad implications. We finally conclude in Section 8.

2 Relation to the literature

Our RL-policy offers a path for the servicer to rengotiate optimally with the borrower, thereby increasing the dollar return through the life of the loan. The borrower most likely becomes more cooperative and less adversarial when she understands that the servicer is willing to work with her and there is a chance that her loan outcome will be better off. For example, a loan modification in terms of lower rate and longer duration may help a borrower in short term liquidity constraint. A foreclosure proceeding can be avoided if the servicer can corroborate the willingness of the borrower to make timely payments going forward. A loan can become cured from a severely delinquent state by recapitalizing the principal outstanding with the prior missed payments.

In this paper, we focus on household mortgage decisions and RL-optimal policy of the servicer. To the best of our knowledge, the only other paper that considers RL-policies in finance is Barberis and Jin (2022), namely, they compare an RL-policy with and without model assumptions and view them as joint drivers of investor behavior. Unlike Barberis and Jin (2022), we focus only on the model-free methodology from the servicer's perspective. We do not impose any assumptions in our framework and derive the RL-policy by maximizing the lifetime servicer's reward purely based on past servicer actions given a certain delinquency state of the borrower.

There has been a long strain of the literature, both theoretical and empirical, on renegotiation and optimality of contracts and their outcomes. First, there are seminal theoretical papers on renegotiation under different model assumptions. Our approach is data driven and devoid of the assumptions, e.g., made in Aghion, Dewatripont, and Rey (1994) and Hart and Moore (1998).[4]

Second, we are related to the previous literature on asymmetric information and moral hazard in terms of their role in renegotiation of contracts. In our approach, we include the aspects detailed in Fudenberg and Tirole (1990) on the impact of moral hazard before observing the consequences of actions. Similar to Roberts and Sufi (2009), we include new information about the credit quality of the borrower and the macroeconomy in a dynamic setting. Like Garleanu and Zwiebel (2008), we explore the impact of covenants in the renegotiation process but via RL.[5]

Third, our work documents the impact of commitment in renegotiation strategies. Similar to Maskin and Moore (1999), our RL-policy characterizes the choice rules that can be implemented by the servicer when borrowers are unable to commit themselves not to renegotiate the mechanism. Our work is also closely related to Hart and Tirole (1988) where they find a close relationship between the optimal long-term contract and the non-commitment outcome. Unlike Laffont and Tirole (1990), who fully characterize the equilibrium of a two-period procurement model with commitment and renegotiation, we provide a multi-period model-free RL-policy.

Finally, we treat the borrower as an adversary in the RL paradigm and document how frictions from covenant violations lead to renegotiation. We alleviate the inefficiencies that arise when negotiation between two parties takes place in the presence of transaction costs, documented in Anderlini and Felli (2001). We extend Piskorski, Seru, and Vig (2010) by exploring post securitization renegotiations in more granularity than just the action to foreclose a delinquent loan.

[4]The former paper analyzes a situation where renegotiation is always possible but contracts can influence the renegotiation process. The latter paper studies the optimal debt contract - specifically, the trade-off between the size of the loan and the repayment under the assumption that some debt contract is optimal.

[5]A covenant is a legally binding agreement which justifies harsh or lenient actions by promoting payment on time and punishing in the event of non-payment or delayed payment, respectively.

3 Data

We utilize a proprietary administrative data set for 23,693 loans from 09/2017 to 3/2020, containing detailed information on residential mortgage performance collected from daily mortgage servicing logs. This data set was provided by a single servicer who is in a joint venture with a private equity firm that focuses on real estate investments. This servicer has 15% of the national market share of Ginnie Mae Early Buyout loans in terms of deal flow. Although this data set is not representative of the entire United States residential mortgage market, there is a sizable portion of loans in this private equity firm's portfolio both in Government-backed and Non-Governmaent categories. Hence this servicer is one of the biggest players in the early buyout market for Ginnie loans and is representative of our analysis of servicer behavior in the United States. The restriction to a single servicer can also be seen as an advantage because it is free of any unobserved, servicer-specific effects. Accessing phone call transcripts and identifying which party initiated the call is a unique advantage of the data. We do not have data on the attributes of individual servicer call representatives or their compensation structure. This is by design so that the servicer is not legally liable for any possible discriminatory practices from their employees. Hence, the servicer call representatives are located in geographically distant locations, namely, TX, IN, CA and the calls get routed to them randomly. Also, to avoid lawsuits from plausible racial bias, each of the above locations have a mix of servicer call representatives of different ethnicities, namely white, african american, hispanic, asian, etc.

3.1 Delinquency States

Given our data, we enumerate possible delinquency states. In particular, we use a *granular* set of delinquency states. Figure 1 shows possible transitions from one state to another.

Borrowers with L30D loans can be aggressively negotiated with to improve future payments. W30-60D loans have reasonable chances of getting cured. W60-90D are loans which have missed two payments consecutively or after a while. There is still some chance that these borrowers reperform from this stage. Several of the strategic defaulters in the W90-120D bucket were observed post the great financial crisis in 2008. B120D loans are in limbo, neither resolved nor cured. Bankruptcy could be filed in chapter 13 (BKCh13) while the borrowers are trying to negotiate favorable

Table 1: Definition of the Delinquency States: We define 9 delinquency states.

State	Definition of different loans
L30D	current or less than 30 days delinquent.
W30-60D	within 30 to 60 days of delinquency.
W60-90D	within 60 and 90 days of delinquency.
W90-120D	in default after 90 days of delinquency with ongoing payments after missing 3 months of payment.
B120D	already beyond 120 days of delinquency.
BK	the borrower has filed for bankruptcy.
Frclsr	have entered the foreclosure proceedings.
PIF	already paid in full.
REO	repossessed by the original lender/servicer representing the lender.
ShrtSal	auctioned in public market for short sale.

terms and trying to obtain temporary relief for missed payments. The bankruptcy can also be filed in chapter 7 (BKCh7) for a complete liquidation of assets. The differential recourse laws across states make this challenging. PIF loans have no remaining cashflows in terms of debt obligations.

Frclsr, BKCh7, PIF, REO, ShortSal are terminal states. But how a specific loan ends up in one of these terminal states is subjective and the optimal strategy of the servicer, given the current delinquency status of the loan is not clear. This is exactly what necessitates the use of the dynamic setting of RL.

3.2 Action space

The disposition strategy (action space) which the servicer chooses to take could be any of the transitions in the figure. The action space consists of but is not limited to the set of actions presented in Table 2.

When there is a *pending claim*, the servicer is in the process of recovering the principal amount of payments missed by the borrower. During *Modification in review*, the servicer is patiently hearing out the issues faced by the borrower and the reasons for missed payment. This phase also leads to the plausible resolution of the missed payments when both parties agree to a reasonable change in the contractual terms of the loans that makes it feasible for the borrower to again start making payments on time in adequate amounts. There are other actions (not mentioned in Table 2)

Figure 1: Possible transitions during the life of a mortgage loan:
This diagram provides a finite state automaton for the life of a mortgage loan. Various possible transitions between the states are considered in this diagram. PIF, BKCh7, REO, ShortSal are absorbing terminal states, i.e., if a loan enters that state, it stays there. Hence, they are marked with two circles.

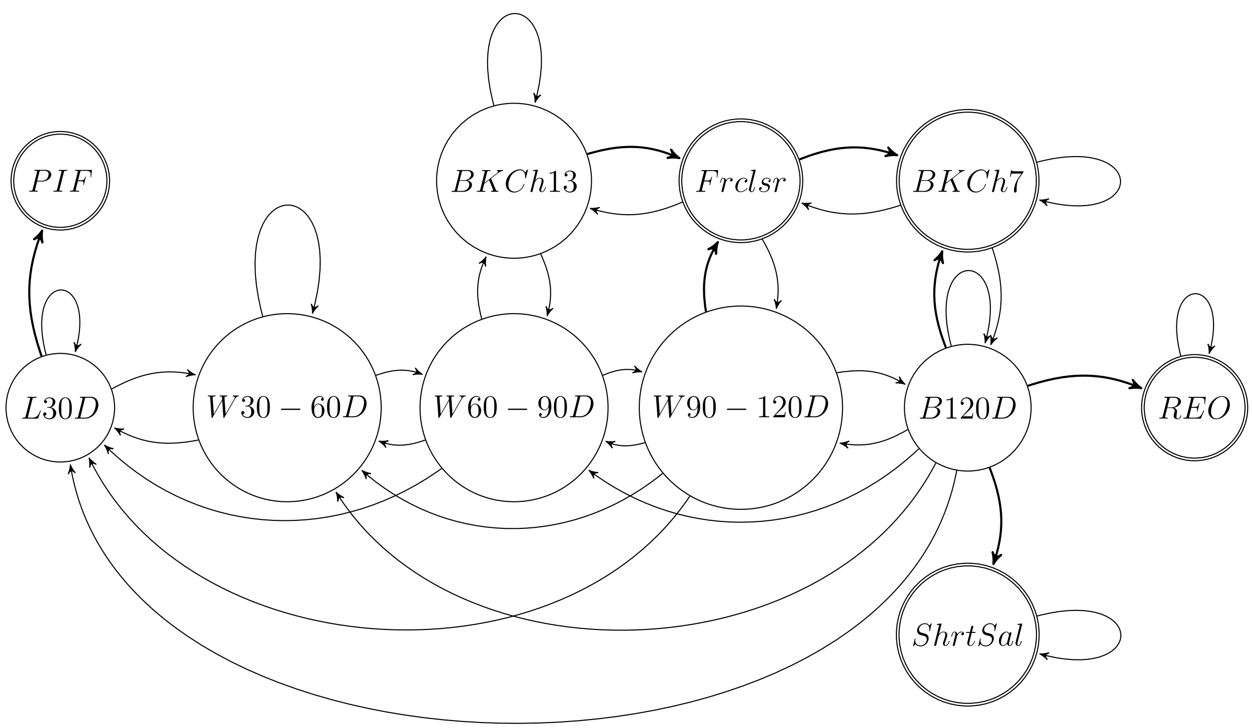


Table 2: Definition of the action space We define the set of possible actions in our data that a servicer can take

Action	Definition
Pending claim	the servicer has filed for a HUD claim.
Modification in review	the ongoing phase of active negotiation between borrowers and servicers.
No Action	the servicer has taken no action.
Pending foreclosure completion	a foreclosure process about to close in the near future.
Real Estate Owned	the process is which the lender or the servicer has gained back possession of the property after offering deed-in-lieu.
Bankruptcy	the ongoing bankruptcy filed the borrower, chapter 11 for a renegotiation or chapter 7 for a complete liquidations of assets.
Not referred for short refinance	not offering a loan modification to the borrower based on servicer's discretion.

like *Notice of intent filed: not in foreclosure.* This implies a notice of intent is filed with regards to foreclosure proceedings. But foreclosure process has not started. *Pending deed-in-lieu* implies an offer has been made to the homeowner to vacate the property as is, without having to settle the missed payments. This is done so that the property could be auctioned in short sale, real-estate owned, etc. *Pending short sale* are loans that are in the process of short sale. This is one of the methods in which the property can be liquidated and any outstanding debt obligations can be mitigated. *Pending repurchase* These are loans where the original buyer expresses interest to repurchase the property from the lender after having sold it. A *consent judgment* is a settlement agreement approved by the court where the borrower acknowledges what they owe to the lender and/or the servicer. *Modification completed* implies that the changes in the contractual terms of the loan has actually taken place. This could be a change in the interest rate, term remaining for the loan or a recapitalization of the remaining balance of the loan. *Performing* loans are those where the borrower has missed a payment or is close to missing a payment. *Rolling delinquency* implies that the severity of delinquency does not increase in the eyes of the servicer.

3.3 Household and demographic data

To capture the spending patterns, demography, relocation, and several other aspects of these borrowers, we use proprietary data from Epsilon at the household level.

Epsilon provides data for roughly 100 million US households.[6] We have mapped the borrowers in our data to Epsilon's database and have extracted several attributes. The variables from this data are ethnic group code, language code, household marital status, number of adults, length of residence, household age, presence of children, household size, household education, income bracket, net worth, liquid resources, investment resources, wealth resources, short term liability, household political party, banking access, move residence date, year the home was built, trigger for buying a house, trigger for move in residence, trigger for home loan. This set of variables helps us create a *noisy environment* via which the servicer actions can lead to optimal outcomes in the presence of the adversarial borrower.

4 Cross-sectional results: motivation for RL

In this section, we propose a novel measure of the borrower's *responsiveness* on the basis of our cross-sectional analysis. This helps us identify the borrower's type and provides a motivation for using RL, which is dynamic by construction.

4.1 Responsiveness measure

We claim that more responsive borrowers *cooperate* more with the servicer than borrowers who are less responsive. The responsiveness measure helps the servicer shortlist borrowers for real-time monitoring and making informed decisions. Servicers can actively engage with the borrowers but can do so more effectively based on how responsive the borrower can be. From a strategic viewpoint, the servicer can decide which borrowers to reach out to and which borrowers are responsive after the servicers initiate communication. Hence, monitoring the borrowers via their responsiveness can enormously streamline the curative process.

The borrower's responsiveness helps servicers identify who can be negotiated with for better terms on the one hand and not waste valuable time and resources on those who are surely going to lead to bankruptcy or foreclosure. With the help of a responsiveness score, a costly Chapter 7 (liquidation) bankruptcy process can be avoided by renegotiating with the more responsive borrowers when they file for the chapter 13 (restructuring) bankruptcy. Loan modifications can be offered to these

[6] See https://www.epsilon.com/us

more responsive borrowers and they may be reperforming after an income shock or a life event. The outreach methodology is better done via calls, and not via text, mail, or email. For severely adverse loans, the servicer can gauge from the borrower's responsiveness whether a preemptive deed-in-lieu (DIL) is possible.[7] The servicer may also choose to retain legal counsel, which may be costly in terms of billable hours. The other possibility would be to settle quickly. To conduct these cost-benefit tradeoff analyses related to bankruptcy or foreclosure timeline, the servicer needs to know the borrower's responsiveness.

Our *time-invariant* measure of the borrower's responsiveness is based on the empirical cumulative distribution function (ECDF) (Langrene and Warin (2020)) of the five variables described below. The ECDF is a robust measure of responsiveness since it weighs all five variables equally while combining the individual marginal distributions. To ensure that all the five inputs significantly contribute towards responsiveness, we conduct a principal component analysis (PCA) in Table 3. None of the four variables can explain more than 95% of the variation and hence we keep the following five variables to define the responsiveness.[8]

The variables used for creating a novel measure of responsiveness are defined below in details:

1. *Months of delinquency:* We assign numerical values to months of delinquency in reverse order, which is in line with increasing responsiveness, *Paid Ahead* := 4, *L30D* := 3, *1 month behind* := 2, *2 months behind* := 1 and *3+ months behind* := 0. This definition enforces a positive correlation between months of delinquency and responsiveness.

2. *Loan delinquency status:* We assign numerical values to Loan Delinquency Status. This is a more accurate layer of the delinquency status of the loans. We assign less scores to more adverse status in the following manner: *L30D* := 6, *30 days delinquent* := 5, *60 days delinquent* := 4, *90 days delinquent* := 3, *Bankruptcy* := 2, *Foreclosure* := 1 and *120+ days delinquent* := 0.

[7]Deed-in-lieu entails the borrower willingly vacating the property and surrendering the deed of the property because of non-payment or delayed payment.

[8]Technically, one can use a two-stage measure of responsiveness based on the latent variable approach. The frequency of conversation, time lag between calls, action post-call, and time for action implementation are valid approaches one can take to measure the level of engagement and cooperation. But the quality of the call transcript and the way they are recorded does not allow us to explore a structural model.

3. *Inbound (IB):* We calculate the sum of all known IB calls from inception till the date for each borrower.

4. *Inbound per Outbound (IB per OB):* We create a relative measure of the number of return IB calls by the borrower per OB call of the servicer.

5. *Information content* captures the reasons for the calls, including OB calls, IB return calls, communications regarding forbearance, foreclosure moratorium, loan modification, borrower's reported unemployment and curtailment of income.

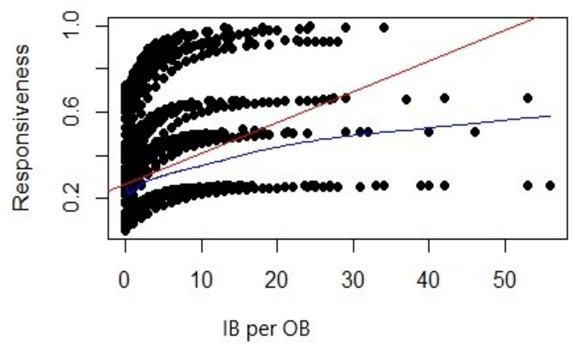
Table 3: Checking with PCA: The PCA yields 5 principal components and even 4 of them cannot explain more than 95% variation in the five inputs of Responsiveness score. Hence all of the five inputs have significant contribution towards Responsiveness

	PC1	PC2	PC3	PC4	PC5
Standard deviation	1.4230	1.0320	0.9859	0.7775	0.5775
Proportion of Variance	0.4050	0.2130	0.1944	0.1209	0.0667
Cumulative Proportion	0.4050	0.6180	0.8124	**0.9333**	1.0000

Months of delinquency and *loan delinquency status* are variables related to the borrower's quality and are a good proxy for past loan performance. Total *IB* calls scale the responsiveness measure by capturing the total number of communications returned by the borrower in response to the servicer reaching out to them. *IB per OB*, as in Figure 2, is the average number of Inbound calls normalized by the number of Outbound calls and is a direct measure of borrower responsiveness. *Information content* is a broader variable capturing the high-dimensional spectrum of borrower's behavior and interaction with the servicer. First, for borrowers who did not call the servicer in the recent past (IB is zero), say the last six months, the success of responsiveness is highly uncertain. Hence, the servicer should contact the borrower (OB is non-zero) if the loan-performance variables are derogatory since there is no attempt to communicate by the borrower. If there is more than one IB call in a month, the borrower is more responsive.

In Figure 2, we plot IB per OB for 19,481 borrowers from the training set. We can distinguish four borrowers' types in terms of their responsiveness with respect to their proactive inbound calls for each servicer outbound call. This is very crucial for a time-constrained servicer for addressing borrower concerns or to reach out to borrowers who have missed payments. This is the first indication that a servicer can

Figure 2: Responsiveness and IB per OB calls
This diagram indicates a novel finding that on an average, there are four different types of borrowers in terms of responsiveness. The straight line is the linear fit of all four types and the curved line is estimated by the non-parametric fitting technique LOWESS.



focus her time and resources on certain borrowers who may be more likely to negotiate which may lead to a better outcome for the borrower, as well as the servicer. For this diagram, we observe and quantify these four borrowers' types in terms of approximate responsiveness buckets: [0,0.2), [0.2,0.5), [0.5,0.6) and [0.6,1]. The buckets are approximate since there is some overlap between them when the number of inbound calls per outbound calls is less than or equal to 5.

We plot responsiveness with several household characteristics to conduct sanity check and find monotonic relationships. We document these relationships in Figure A1 for short-term liability, in Figure A2 for liquidity, in Figure A3 for net worth, in Figure A4 for investments, in Figure A5 for age.

Table 4 demonstrates that there is a clear difference between these buckets in terms of length of residence, household age, income, net worth, liquidity, investments, wealth and short term liabilities. Although the first three buckets show monotonic trends, the most responsive bucket is slightly lower than the penultimate. This is because of the plethora of issues that are the talking points for these most responsive borrowers, not all of which have a high information content.

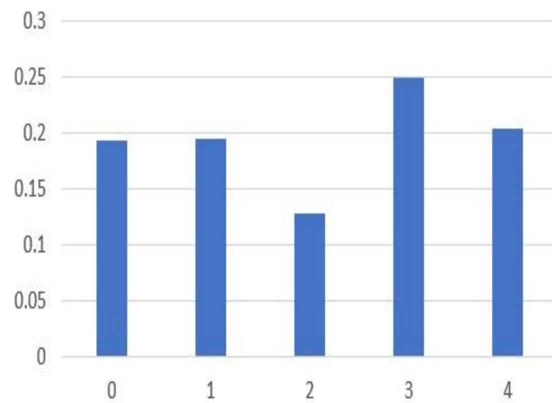
Table 4: Responsiveness and household characteristics:
We quantify the four borrowers' types in terms of four approximate responsiveness (R) buckets. This table demonstrates the clear difference from the p-value of the relative mean of household attributes in the first 3 buckets with the most responsive bucket.

Bucket	**[0,0.2)**	**[0.2,0.5)**	**[0.5,0.6)**	**[0.6,1]**
Number	8009	9406	1007	1059
Length res	12.24 (< 0.05)	13.56 (0.15)	14.50 (< 0.05)	13.78
Age	47.98 (< 0.05)	50.25 (< 0.05)	53.24 (0.22)	52.84
Income	77048.94 (< 0.05)	76183.82 (< 0.05)	76832.17 (0.06)	73142.12
Net worth	154053.25 (< 0.05)	170372.63 (0.06)	196288.48 (0.10)	182448.06
Liquidity	9125.89 (< 0.05)	10007.84 (0.3)	12841.36 (< 0.05)	10339.00
Investment	85800.97 (0.18)	92742.13 (0.22)	108430.98 (< 0.05)	89596.32
Wealth	105219.13 (0.07)	112436.21 (0.38)	131330.69 (< 0.05)	114235.13
Short liab	26661.72 (< 0.05)	24784.71 (< 0.05)	22663.85 (0.25)	22089.24

Figure 3 shows a non-monotonic, non-linear relationship between the borrower's responsiveness and months of delinquency. The average responsiveness of a borrower increases until 60 days of delinquency. The borrower reduces communication dramatically around 60 days of delinquency and then again starts communicating in more adverse delinquency states. This is the first indication that if there is a decent negotiation with a responsive borrower before 60 days of delinquency, there is a high chance that the loan may not even cross the 60-day delinquency threshold. This is a significant finding since this result helps servicers preemptively offer loan modification to borrowers who may otherwise be considered *performing* and within 60 days of delinquency.

We also conduct robustness check on the above four buckets of responsiveness

Figure 3: Responsiveness and months of delinquency
The borrower reduces the communication dramatically around 60 days of delinquency and then again starts communicating in more delinquency states. A robust negotiation at this juncture can lead to a resolution and a better outcome.



on both sides of 60 days of delinquency. We find in FigureA6 and FigureA7 in the appendix that the 4 buckets remain consistent whether the communications take place before 60 days of delinquency or after, respectively. This provides conclusive evidence of the existence of different types of borrowers identified on the basis of their responsiveness.

The entire exposition and analysis of the soft information using natural language processing (NLP) is beyond the scope of this paper. Hence, we present some of the results in the appendix for the readers to appreciate how insightful these call transcripts really are. To achieve this objective, we first plot the 5 main adverse delinquency states, namely, delinquent, bankruptcy, foreclosure, reo, short sale, using T-SNE, a.k.a, stochastic nearest neighbor embedding (see van der Maaten and Hinton (2008) for methodology), which is just a two-dimensional visualization of the clusters of similar words in Figure A8. The T-SNE plots helps the readers appreciate the high-dimensionality and complexity of the information content in these communications, illustrating their intuitive underlying structure as recurrent co-location of topics. The axes do not have any units or physical significance and is chosen automatically by the T-SNE algorithm to appropriately fit the important similar words in one diagram. The impact of legal- and title-related conversations and how they overlap with the basic 5 delinquency states can be visualized in Figure A9 and Figure A10, respectively.

4.2 Features, variable importance & tree for responsiveness

To capture different aspects of responsiveness score, we use a regression tree technique to fit the score on a set of loan and borrower's specific variables. Different measures of variable importance in Table 5 enable us to understand the marginal contributions of aspects related to modification date, stage of foreclosure, length of residence and several others. The high importance of modification date, current rate, current fico score, foreclosure stage, year the home was built (Year Home Built) and delay in the bankruptcy process underscores the temporal aspect of the borrower's responsiveness beyond time-invariant factors like original fico score, original ltv, original rate, etc. Hence, we explore RL as a methodology to capture the optimal action of the servicer given the servicer now knows borrower responsiveness and has learnt the borrower's type using soft information captured during these monitoring communications.

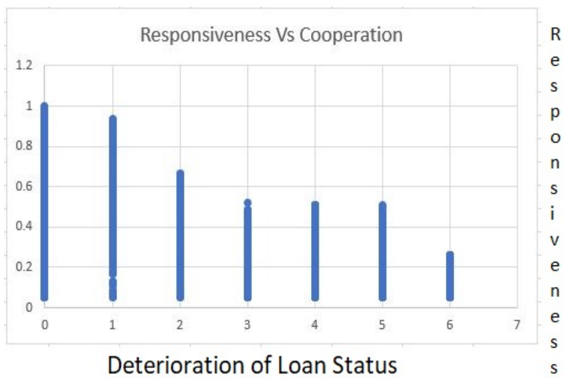
Table 5: Variable importance for tree-based regression:
We use regression tree technique to fit the responsiveness score on a set of loan-specific, borrower-specific variables. The variable importance table enables us to understand the marginal contributions of household variables.

	variable	scaled_importance	percentage
1	modification_date	1.00	0.17
2	Original_fico	0.41	0.07
3	Current_rate	0.41	0.07
4	Current_fico	0.38	0.07
5	Original_ltv	0.27	0.05
6	Original_rate	0.27	0.05
7	Foreclosure_stage	0.25	0.04
8	Year Home Built	0.25	0.04
9	Buy a House Rank	0.24	0.04
10	Bankruptcy_delay	0.23	0.04
11	Home Loan Rank	0.23	0.04
12	Move Residence Date	0.20	0.03
13	Length of Residence	0.14	0.03
14	Investment Resources	0.14	0.02
15	Short Term Liability	0.13	0.02
16	Household Size	0.13	0.02
17	Household Age	0.12	0.02
18	Liquid Resources	0.12	0.02
19	Net Worth	0.12	0.02
20	Wealth Resources	0.09	0.02
21	Household Education	0.09	0.02
22	...	...	...

4.3 Cooperation

We call a borrower cooperative if the borrower's delinquency status does not deteriorate over time, i.e., the loans remain in the same delinquency bucket or improve. In Figure 4, we plot responsiveness on the vertical axis with increasing deterioration of loan status on the horizontal axis. The deterioration of loan status is a numeric variables calculated by the maximum length of movement for a loan from any present loan status to a worse loan status. From this figure, it is evident that the most responsive borrowers can maintain the same delinquency status or improve. This happens due to borrower's cooperation resulting from responsiveness and renegotiation. In this context, we allude to cooperation when there is no further deterioration of the delinquency status of a loan.

Figure 4: Responsiveness versus cooperation
We plot responsiveness with increasing deterioration of loan status on the X-axis. The responsiveness score decreases with time variation in delinquency status.



However, responsiveness provides cross-sectional results. This necessitates the use of the dynamic RL framework to capture the time-varying nature of these borrower-servicer negotiations.

5 Methodology of reinforcement learning

We implement Q-learning(QL), an RL algorithm with experience replay, on our data.[9] The RL algorithm learns an optimal policy based on state-transition tuple $(s_i, a_i, r_{i+1}, s_{i+1})$, where s_i is the current delinquency state, a_i is the selected servicer action in the current state, r_{i+1} is the immediate reward received after transitioning from the current state to the next state, and s_{i+1} is the next delinquency state.

Our model-free Q learning approach estimates $Q(s, a)$, the value of undertaking an action a in a state s. It does so by learning from experience and not by making assumptions about the probabilities of future states and rewards. One caveat of this model-free approach is that it is operational over a limited time range when the agent

[9]We use the software package:
https://cran.r-project.org/web/packages/ReinforcementLearning/vignettes/ReinforcementLearning.html

actively interacts with the environment. But the limited time range of the data works to our advantage since no weight is given to the behavior in the distant past. Our methodology differs from Malmendier and Nagel (2011) in that a servicer's action solely depends on her experience and not on past prejudice. Malmendier and Nagel (2011) asks a completely different research question about the impact of depression on the long terms habits of people born during the depression. Here, we are simply comparing the methodology. Our study is more in line with the methodology used in Greenwood and Shleifer (2014) while comparing expectation of returns and expected returns. But our conclusions are different from Greenwood and Shleifer (2014) because we do find evidence of converge and optimality of servicer's actions.

The servicer maximizes a Q-value defined as

$$Q^*(s,a) = \max_{a'} E_0 \left[\sum_{t=1}^{T} \gamma^t r_t \right] \tag{1}$$

where r_t, E_0 and $Q^*(s,a)$ satisfy the Bellman equation:

$$Q^*(s,a) = E_t \left[r_{t+1} + \gamma \max_{a'} Q^*(s_{t+1}, a) | s_t = s, a_t = a \right] \tag{2}$$

where the expectation is taken over future possible rewards r_{t+1} and states s_{t+1} by way of the probability distribution $p(r_{t+1}, s_{t+1} | s_t, a_t)$.

QL can formulate the optimal action a of the servicer at time t and in state s and the resulting reward r_{t+1} at time $t+1$ that brings the loan to state s_{t+1}. Suppose also that, at time t, the algorithm's initial estimate of $Q^*(s,a)$ is $Q^{old}(s,a)$. At time $t+1$, after observing the reward r_{t+1}, its estimate of $Q^*(s,a)$ is updated as follows:

$$Q^{new}(s,a) = Q^{old}(s,a) + \alpha_t \left[r_{t+1} + \gamma \max_{a'} Q^{old}(s', a') - Q^{old}(s,a) \right] \tag{3}$$

where α_t is the learning rate.[10] Eq 3 can be written as a convex combination of Eq 1 and Eq 2.

[10]The term in square brackets in Eq 3 is the reward prediction error (RPE). RPE is the realized value of taking the action a relative to its previously anticipated value.

$$Q^{new}(s,a) = (1-\alpha_t)Q^{old}(s,a) + \alpha_t \left[r_{t+1} + \gamma \max_{a'} Q^{old}(s',a') \right] \tag{4}$$

The Q-learning algorithm takes an estimate of the right-hand side of (2) and then updates $Q^{old}(s,a)$ in the direction of this estimate to an extent determined by the learning rate α_t. Specifically, it proxies for the expected reward $E_t(r_{t+1})$ in (2) by the realized reward r_{t+1} and for $E_t[\max_{a'} Q^*(s_{t+1},a')]$ by $\max_{a'} Q^{old}(s_{t+1},a')$.

The algorithm does not simply choose the action with the highest estimated value of $Q^*(s,a)$, i.e., with the highest value of $Q^{old}(s,a)$. Rather, it chooses an action probabilistically, where the probability of choosing a given action is an increasing function of its Q value,

$$p(a_t = a | s_t = s) = \frac{e^{\beta Q^{old}(s,a)}}{\sum_{\{a'\}} e^{\beta Q^{old}(s,a')}} \tag{5}$$

where β is called a *inverse temperature* parameter, but we refer to it more simply as the exploration parameter. In the limit as $\beta \to \infty$, the algorithm chooses the action with the highest Q value; in the limit as $\beta \to 0$, it chooses an action randomly.

This probabilistic choice, often known as a *softmax* approach, encourages the algorithm to *explore* an action other than the one that currently has the highest Q value in order to see whether this other action has an even higher Q value.

6 Numerical results

We compare our RL-optimal policy with other policies by constraining the set of actions the servicer can take based on their expert judgement. We first conduct this comparison with respect to the Q-values of the policies. Then, we convert the Q-values across delinquencies to the probability distribution by normalizing the Q-values so that the sum is 1 for a given policy. Then we compare the frequencies of actions based on our RL-policy with the counterfactual of the actual actions currently taken by the servicers. We explore how our RL-policy changes with the interplay of the learning rate and discounting.

Using the frequency distribution of the reward (the ratio of Collections to the

Original Balance of the loans) in Figure A11 and severe clustering of rewards near zero in Figure A12, we define standardized set of rewards. This is because our Q learning cannot handle continuous rewards and the rewards have to be a predefined finite set of discrete values. Then, we shortlist the set of (state, action) pair which has material impact on the rewards and eliminate redundant possibilities to improve the convergence of our Q learning algorithm. To do this, we first perform OLS on rewards with states, actions and the Cartesian product of these states and actions in Table A1 to find the combinations which are statistically significant. We further narrow down the possible combinations by performing fixed effect regressions on rewards in Table A2.

6.1 Reinforcement learning optimal, harsh and lenient policies

The RL-optimal policy is computed via several learning iterations for 7 delinquency states and 18 possible servicer actions, which are described in Section 3. It led to a total reward of 14.4. The current approach is based on legal and industry expertise, and hence the policies used in practice are either systematically harsh or lenient. The harsh servicer actions exclude the actions: *No Action*, *Performing*, *Rolling Delinquency*, *Not Referred for Short Refinance*, explained in Section 3. For a policy based on harsh servicer actions, the total reward is 1.8, which is significantly lower than 14.4, the reward from our optimal action.

The following servicer actions are excluded in a lenient disposition towards the borrowers: *Notice of Intent filed:Not in FC*, *Pending Claim*, *Pending Foreclosure Completion*, *Pending Deed-in-Lieu*, *REO*, *Modification Completed*, explained in Section 3. For the policy based on lenient servicer actions, the total reward is 2.7, again much lower than 14.4, the reward from our RL-policy.

The Q-matrices for optimal, harsh and lenient servicer actions are presented in Table 6. The corresponding Q values for each state-action pair for the optimal, harsh and lenient servicer actions are presented in Table 7.

The first thing to notice from the first column of Table 7 is that the Q-value is zero for the states "L30D" and "B120D". This is because loans less than 30 days of delinquency cannot be negotiated any better for the borrower or the servicer. Also, when the loans have crossed 120 days of delinquency, there is not much that can

Table 6: Optimal action for each state for RL-optimal, harsh and lenient strategies: The first column refers to RL-optimal actions, the second and third columns refer to the cases where the servicer restricts herself to a subset of actions, either harsh or lenient.

		Disposition strategies	
States	**RL-optimal**	**Harsh**	**Lenient**
L30D	Mod Review	Mod Review	Mod Review
W30-60D	No Action	Pend FC Complet	No Action
W60-90D	No Action	Mod Review	Not refer Short Refi
W90-120D	Pend FC Complet	Mod Review	Mod Review
B120D	Mod Review	Mod Review	Mod Review
BK	Pending Claim	Pend FC Complet	No Action
FC	REO	Pend FC Complet	Mod Review

Table 7: Optimal Q-value across states for RL-optimal, harsh and lenient strategies: For RL-optimal policy, the Q value drops when the loan is 90 days delinquent, because loan modification is offered by the servicer. During bankruptcy, the Q value jumps back to 0.9.

		Disposition strategies	
States	**RL-optimal**	**Harsh**	**Lenient**
L30D	0	0	0
W30-60D	0.9	0.9	0.89991
W60-90D	0.9	0	0.89991
W90-120D	0.44991	0	0
B120D	0	0	0
BK	0.9	0.9	0.89991
FC	0.89991	0	0

be done except offering reasonable modification terms to the borrower. For the RL-optimal policy, the Q value drops when the loan is 90 days delinquent. This is because the RL-optimal policy indicates an aggressive foreclosure procedure immediately after the loan crosses 90 days of delinquency. The Q value jumps back to 0.9 when the borrower files for bankruptcy, since the servicer can recoup lost principal and interest by filing a claim.

For a harsh set of actions, the servicer tries to preempt further downgrade of the loan beyond 30 days of delinquency or bankruptcy. But, invariably a loan that is 30 days delinquent, worsens more often than getting cured. In the event the borrower files for bankruptcy, the Q value remains very close to column 1 (0.9). Most bankruptcy filings by the borrower result in foreclosure proceeding by the servicer which brings back the Q value to zero, when the servicer restricts herself to a set of harsh actions.

The lenient set of actions have similar Q-values as the harsh set of actions. Because of leniency during the first 90 days of delinquency, the servicer has slightly better opportunity, i.e., Q-values of 0.89 to recoup some of the delayed payments by giving better terms to the borrower.

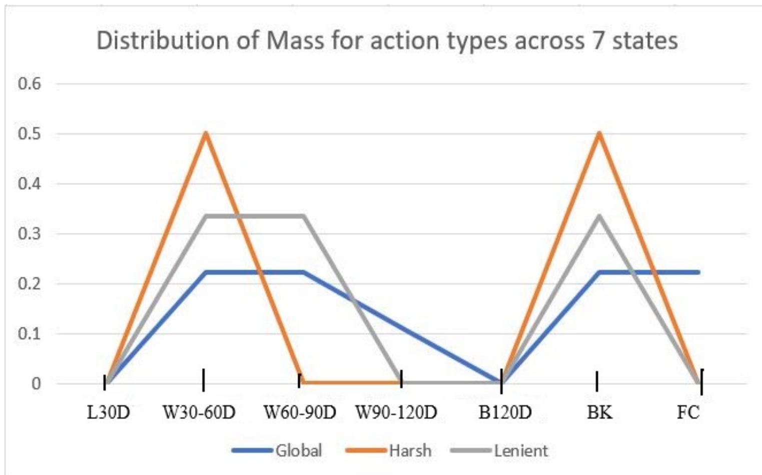
6.2 Comparison of policies in terms of flexibility

Figure 5 is the visualization based on the normalized Q-values in Table 7. Hence, we can compare the distributions of the Q-values across delinquency states for the optimal RL policy versus the harsh and lenient policies. The blue, orange and gray colors, respectively, represent the probability distributions of the RL-policies based on optimal, harsh and lenient servicer actions.

Based on the Q-values, we compare how the masses are shifted across the 7 delinquency classes for the RL-optimal action, as well as the harsh and lenient actions. In Figure 5, for the harsh action, the mass is shifted to extreme cases, either less than 60 days delinquent or bankruptcy/foreclosure. The lenient action, on the other hand, allows the borrower to file bankruptcy. There is less communication regarding negotiation in the critical 90 days and 120 days of delinquency. The RL-optimal action spreads the mass of possible actions on almost all delinquency classes and hence extracts the maximum reward vis-a-vis maximizing the Q-value dynamically.

For all the actions, the Q-values for L30D and W90-120D loans are zero. For L30D loans, there is not much requirement for servicer's actions as the borrower is

Figure 5: Distribution of optimal, harsh and lenient actions across states: The harsh and lenient actions shift the mass to lesser number of states. The RL-optimal action spreads the mass of possible actions on almost all delinquency classes and hence increases the possibility of overall reward.



in good standing and hence there is no reward to be extracted from the borrower resulting in a zero Q-value. However, we show in Subsection 6.3 that loans that are 1-29 days days of delinquency can be offered decent loan modification and although this not local optimal, it will turn out to be global optimal as it will preempt lot of the negative ripple effects for more adverse states of the loan which may turn up later. This is true for each of the above three optimal servicer actions. Loans that are 120 days delinquent are in limbo. These borrowers are technically at default but communication and negotiation is the best strategy for the servicer to encourage these borrowers to re-perform; otherwise, these loans almost surely would end up in bankruptcy and/or foreclosure.

6.3 Transition matrix for RL-optimal and actual actions

In Table 8, we present the transition matrix from data: **bold font** specifies RL-optimal actions and ***bold and italics*** **font** documents actions that are currently taken by the servicers. Clearly from the current servicer actions, we can argue that these actions are adhoc and far from optimal based on our RL methodology. Broadly, there are differences in the percentage of times the servicer takes the RL-optimal action versus a different action. There is some opportunity of reward maximization that is lost by not always choosing the RL-optimal actions. But these differences are stark for severe delinquency classes. This makes the current servicer actions inefficient in terms of both resolution and profitability.

Table 8: Transition matrix from data: Percentages in **bold font** are RL-optimal actions and percentages in ***bold and italics*** font are those actually taken by the servicer.

				States			
Actions	**L30D**	**W30-60D**	**W60-90D**	**W90-120D**	**B120D**	**BK**	**FC**
Pend FC	2.32%	2.40%	2.59%	**2.61%**	8.87%	2.55%	***77.57%***
Pend Claim	0.00%	0.00%	0.00%	0.00%	6.41%	**0.02%**	0.99%
Bankruptcy	0.97%	0.67%	0.11%	0.83%	2.57%	***87.65%***	5.72%
No Action	8.18%	9.70%	11.54%	11.62%	7.10%	7.29%	4.55%
Performing	***68.54%***	***50.39%***	***44.03%***	***42.81%***	**17.00%**	0.90%	0.81%
No Refi	3.48%	1.86%	0.90%	0.44%	0.17%	0.00%	0.33%
REO	0.00%	0.00%	0.00%	0.00%	0.02%	0.00%	**0.18%**
Mod In Rev	**0.17%**	0.30%	0.64%	1.28%	3.01%	0.23%	1.14%
Pend ShrtSale	0.00%	0.00%	0.00%	0.01%	0.56%	0.00%	0.62%
Pend D-In-L	0.00%	0.00%	0.00%	0.00%	0.20%	0.00%	0.17%
Pend Repurch	0.00%	0.00%	0.00%	0.00%	0.01%	0.00%	0.11%
Cnsnt Judgm	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%
Mod Compltd	1.50%	0.89%	0.85%	0.55%	0.31%	0.06%	0.91%
Roll Delinq	0.00%	0.05%	0.11%	0.04%	0.02%	0.00%	0.00%
NOI: Not FC	14.83%	33.73%	39.24%	39.81%	***53.74%***	1.30%	6.88%

When a loan is L30D or less than 30 days of delinquency, the servicer assumes that loans are performing and hence the servicer does not communicate with the servicer. However, this is precisely where the servicer can proactively offer and negotiate, since the borrower is technically not delinquent yet. Communicating with the borrower and talking through the reasons and financial health of the borrower can help the servicer gauge the terms of a loan modification which can prevent the number of days of delinquency from being less than 30 days. In reality, the servicer is not taking the RL-action and assuming the loan is performing 68.54% of the time. The RL-optimal servicer action of *Mod In Rev* is conducted only 0.17% of the time. This has an enormous ripple effect on the future rewards of the servicer as well the loan outcome from the borrower's perspective.

When a loan is within 30 and 60 days delinquent (W30-60D), the servicer is taking the right decision of assuming the borrower is performing 50.39% and 44.03%, respectively. This is borne out from a single number that is in bold and italic in these columns. In other words, there is no bold number as actual and RL-policy coincide. However, they are still losing out on rewards around 33.73% and 39.24% of the times, respectively, by filing a notice of intent for several borrowers for these delinquency states.

When a loan approaches 90 days of delinquency (W60-90D), the servicer assumes the borrower is performing 42.81% of the time, filing a notice of intent, and not proceeding with foreclosure 39.81% of the time. The RL-optimal action for the servicer should be to file for foreclosure, but it has currently been done only 2.61% of the time.

When a loan is more than 90 days but less than 120 days delinquent (W90-120D), technically, the loans are in limbo, and proactive communication with the borrower should be the best way to avoid costly bankruptcy and foreclosure proceedings. But active communication takes place only 17% of the time. 53.74% of the time, the servicer is just filing a notice of intent and not actual foreclosure.

Since our data set mostly has Ginnie Mae loans (FHA, VA and USDA), the best way to recoup the principal and some of the interest is to file for HUD (Housing and Urban Development) claims and not let the borrower file for bankruptcy. But 87.65% of the time, the borrower is filing for bankruptcy which is costly for the servicer. Claims are filed only 0.02%, which should be the RL-optimal action of the servicer in this situation. When a loan is in foreclosure proceedings and pending foreclosure completion, the RL-optimal action is to conduct REO (Real Estate owned) and recover the property. This is done only 0.18% of the time.

The above results demonstrate the massive inefficiencies in the servicer's actions which are currently ad-hoc based on qualitative judgements. Our RL-optimal policy opens the door for finding a more profitable strategy for the servicer; at the same time, the borrower is not worse off.

6.4 Sensitivity analysis

In this section, we study the results of our sensitivity experiments with respect to the parameters α and γ. The learning rate, denoted as α in Equation 3, is what allows to iteratively update an old Q-value to a new Q-value. Our model-free RL approach works because we only use deterministic reward and transition functions.

The parameter γ, in Equation 3, can be interpreted as the time value of the future reward or the discount factor and it affects the learning rate. If γ is equal to one, i.e., if the discount rate is zero, the agent values future reward *just as much* as the current reward. That is, if the servicer proactively negotiates and resolves the delinquency in the loan, this is *just as valuable* as paying on time without missing any payments. As a result, learning does not work well at high γ values. Conversely, a zero value of γ will cause the agent to only value immediate rewards, which is a very myopic way to look at the situation.

In Table 9, we keep a very low discount rate, in other words, $\gamma = 0.99$. Then

Table 9: Optimal action for each state for iterations = 10,000 and $\gamma = 0.99$
Here, we keep a very low discount rate ($\gamma = 0.99$) and choose a range of values for the learning parameter α to compute the optimal set of actions for each delinquency state and their overall cumulative rewards change.

	α=**0.99**	α=**0.95**	α=**0.9**	α=**0.8**	α=**0.7**	α=**0.6**
L30D	Mod	Mod	Mod	Mod	Mod	Mod
W30-60D	NA	NA	Mod	NRSR	NA	NA
W60-90D	NA	NA	Mod	NA	NA	NA
W90-120D	FC	FC	Mod	Mod	FC	Mod
B120D	Mod	Mod	Mod	Mod	Mod	Mod
BK	FC	PC	PC	PC	PC	NA
FC	DIL	DIL	REO	DIL	DIL	DIL

we choose a range of the learning parameter α from $\{0.99, 0.95, 0.9, 0.8, 0.7, 0.6\}$ and observe how the RL-optimal set of actions for each delinquency and their overall cumulative rewards change. Also, we separately track the first five delinquency states and the worst two states to see the differential impact of the decreasing learning rate on the RL-optimal policies and their corresponding rewards. Here, 'NA' means "No Action", 'DIL' means "Deed-in-Lieu", 'Mod' means "Modification in Review", 'FC' means "Pending Foreclosure Completion", 'PC' means "Pending HUD Claim", 'NRSR' means "Not referred for Short Refinance".[11] For any learning rate higher than 0.9, the RL algorithm can identify the most optimal set of actions. The rest of the actions for learning rates lower than 0.9 are suboptimal, as those actions are derived from their decreasing cumulative optimal Q-values. For example, the RL algorithm learns to preempt borrower bankruptcy filing with a foreclosure proceeding, only with a high learning rate of 0.99.

In Table 10, we fix α at 0.99 and vary both the discount factor γ and the number of iterations (N). In other words, when the servicer learns from experience, the discount factor does not play an important role. There is a minor difference in the RL-optimal policy when the borrower files for bankruptcy with increasing number of iterations.

Specifically in the delinquency state of bankruptcy, when there is less room for communication and negotiation between the borrower and servicer (N= 10,000), there

[11] HUD refers to the Housing and Urban Development wing of the US Govt. which has designed the Ginnie Mae loans and guarantees the proceeds to the investor is the event of borrower's default. This eliminates the credit risk almost completely, since the servicer can recoup the principal and most of the interest by filing a HUD claim.

Table 10: Impact of discount factor with high learning α=0.99:
We find that the optimal actions, in a high learning rate, are the same independent of the discount factor. This result is important as it indicates that in a high learning environment, past experiences do not bias the optimal servicer actions.

	N=10,000	**N=10,000**	**N=100,000**	**N=100,000**
	γ=0.99	**γ=0.01**	**γ=0.99**	**γ=0.01**
L30D	Mod	Mod	Mod	Mod
W30-60D	NA	NA	NA	NA
W60-90D	NA	NA	NA	NA
W90-120D	FC	FC	FC	FC
B120D	Mod	Mod	Mod	Mod
BK	*FC*	*PC*	NA	NA
FC	DIL	DIL	DIL	DIL

is a residual information asymmetry. This prompts the servicer to start the foreclosure proceedings when the borrower initiates bankruptcy. This is more so when the discount rate is low, i.e., $\gamma = 0.99$. When the discount rate becomes materially higher, then the servicer does not value older rewards and is focused on the present reward going forward. Hence, filing a HUD claim is the optimal strategy for the servicer to reclaim the principal when the borrower files for bankruptcy. The tradeoff is the loss in the missed interest payments which cannot be recovered from HUD claims.

When the servicer has enough back and forth communications (N= 100,000), we find that the RL-optimal actions for a high learning rate $\alpha = 0.99$ are exactly same for both $\gamma = 0.99$ and $\gamma = 0.01$ with 10,000 iterations of learning. Even for the severe state of bankruptcy, the optimal actions are the same, i.e., no action. This is because most bankruptcies are chapter 11 which leaves room for renegotiation. After several rounds of dialogue, one can argue that the information asymmetry is mitigated and the servicer has almost perfect information about the borrower. Hence, taking no preemtive action may be the best strategy to avoid a costly chapter 7 bankruptcy filing. In other words, the loan can still cure. This result is important since it indicates that in a high learning environment, past experiences do not bias the RL-optimal servicer actions. In other words, when the servicer learns from experience, the discount factor does not play a role. This result proves that, unlike in Malmendier and Nagel (2011), past prejudices about the borrower's performance are not deciding factors for the RL-optimal action of a servicer.

As an agent begins the learning process, we want this process to take random actions to explore more paths. But as the agent gets better, the Q-function converges to more consistent Q-values. Now we would like our agent to exploit paths with highest Q-value, i.e., take greedy actions. This is where ϵ comes in during exploration. So the agent takes random actions for probability ϵ and greedy action for probability 1-ϵ. However, we do not find any first or second order impact on our RL-optimal policy based on greediness of actions. Hence, the results of our model-free approach are as good as model based approaches in this context.

7 Discussion and extension

First, the servicers do not estimate the probability distribution of returns in practice, although such information is available. It just may be that servicers do not possess the quantitative know-how to appreciate the structure of asset returns. It can very well be that this model-free RL is fundamental to human decision-making. Second, our paper points to learning rates as a primary driver of risk management and profit maximization from the servicer's standpoint. Third, in the extensive literature on the Bellman equation, investors' beliefs and preferences are priors, but in model-free RL the value function itself is a prior. In our analysis, the learning rate is assumed to be constant from 2017-2019. However, the learning rates may have temporal variance during unprecedented times like COVID-19 and may lead to new results.

8 Conclusion

Because of information asymmetry at the loan-level, the financial intermediary in the mortgage market, namely the servicers, have anecdotally used the sticks and carrots approach. We show that it is possible to measure the borrower's responsiveness based on our unique administrative data set of call transcripts between the borrowers and servicers. We provide evidence that more responsive borrowers cooperate upon communication with them. This requires a setting where one can evaluate an RL-optimal strategy for the servicer, which is most aligned with the lender and investor. But sometimes, the lenders do not prefer uncertainty and are willing to take a hit if they can quantify their loss. This situation is not aligned with the servicer, who makes money on fees collected from interest payments by the borrower. So, the servicer would instead work with the borrower and work out a loan disposition strategy that benefits both the borrower and servicer. Hence, we frame the research question in terms of the servicer, and our results detail the servicer's RL-optimal actions given the borrower's delinquency status. Our paper provides a quantitative framework for servicers to target specific responsive borrowers with a higher propensity to cooperate. Hence, an RL-optimal action by the servicer enables us to document the most efficient transition among delinquency states during the life of a loan. The borrower's outcome is also not worse off vis-a-vis this servicer's RL-optimal policy.

We find that this RL approach is more efficient than the qualitative approach of sticks and carrots currently used by the mortgage servicing industry. We find that learning rate has the most important effect on the set of RL-optimal actions for the servicers. We demonstrate the divergence in learning the RL-optimal policy based on decreasing learning rates. For high learning rates, the discount factor does not matter. This provides a new perspective to mitigate the differences in the existing viewpoints on beliefs. Past experiences do not dominate the RL-optimal actions of an agent in a high learning environment. At the same time, more experience replay alleviates the errors one may estimate from narrow framing.

References

Aghion, Philippe, Mathias Dewatripont, and Patrick Rey, 1994, Renegotiation Design with Unverifiable Information, *Econometrica* 62, 257–82.

Anderlini, Luca, and Leonardo Felli, 2001, Costly Bargaining and Renegotiation, *Econometrica* 69, 377–411.

Barberis, Nicholas, and Lawrence Jin, 2022, Model-free and Model-based Learning as Joint Drivers of Investor Behavior, .

Fay, Scott, Erik Hurst, and Michelle J. White, 2002, The Household Bankruptcy Decision, *American Economic Review* 92, 706–718.

Fudenberg, Drew, and Jean Tirole, 1990, Moral Hazard and Renegotiation in Agency Contracts, *Econometrica* 58, 1279–1319.

Garleanu, Nicolae, and Jeffrey Zwiebel, 2008, Design and Renegotiation of Debt Covenants, *The Review of Financial Studies* 22, 749–781.

Goodfellow, et. al., 2014, Generative Adversarial Networks, .

Greenwood, Robin, and Andrei Shleifer, 2014, Expectations of Returns and Expected Returns, *The Review of Financial Studies* 27, 714–746.

Hart, Oliver, and John Moore, 1998, Default and Renegotiation: A Dynamic Model of Debt*, *The Quarterly Journal of Economics* 113, 1–41.

Hart, Oliver D., and Jean Tirole, 1988, Contract Renegotiation and Coasian Dynamics, *The Review of Economic Studies* 55, 509–540.

Laffont, Jean-Jacques, and Jean Tirole, 1990, Adverse Selection and Renegotiation in Procurement, *The Review of Economic Studies* 57, 597–625.

Langrene, Nicolas, and Xavier Warin, 2020, Fast multivariate empirical cumulative distribution function with connection to kernel density estimation, .

Malmendier, Ulrike, and Stefan Nagel, 2011, Depression Babies: Do Macroeconomic Experiences Affect Risk Taking, *The Quarterly Journal of Economics* 126, 373–416.

Maskin, Eric, and John Moore, 1999, Implementation and Renegotiation, *Review of Economic Studies* 66, 39–56.

Matvos, et. al., 2021, Government and Private Household Debt Relief during COVID-19, Working Paper 28357 National Bureau of Economic Research.

Piskorski, Tomasz, Amit Seru, and Vikrant Vig, 2010, Securitization and distressed loan renegotiation: Evidence from the subprime mortgage crisis, *Journal of Financial Economics* 97, 369–397.

Roberts, Michael R., and Amir Sufi, 2009, Renegotiation of financial contracts: Evidence from private credit agreements, *Journal of Financial Economics* 93, 159–184.

Sumit Agarwal, et. al., 2017, Policy Intervention in Debt Renegotiation: Evidence from the HAMP, *Journal of Political Economy* 125, 654–712.

van der Maaten, L.J.P., and G.E. Hinton, 2008, Visualizing High-Dimensional Data Using t-SNE, *Journal of Machine Learning Research* 9, 2579–2605 Pagination: 27.

A Appendix

In this section, we document some of the esoteric details of this unique data. Also, we report more results that corroborate our findings, but which can be viewed as supplementary to the main paper.

A.1 Household and demographic determinants of Responsiveness

Identifying the exact borrower type is very complicated, if not impossible. Hence, we provide several attributes that contribute to this borrower segmentation. We observe short-term liability as a significant determinant of responsiveness in Figure A1. The less short-term liability a borrower has, the less financially constrained she is and the more responsive she is willing to be. The liquidity also tells the same story in Figure A2 in terms of responsiveness directly related to the available cash the borrowers have at a given time. We do not find any relationship of responsiveness with household income brackets as expected since the income for these households is not smooth over time. Borrowers become more responsive when they have much to lose, either in terms of net worth in Figure A3 or overall investments they may have in Figure A4. Responsiveness also has a positive relationship with age as evidenced by Figure A5.

Figure A1: Responsiveness versus short term liability

Figure A2: Responsiveness versus liquidity

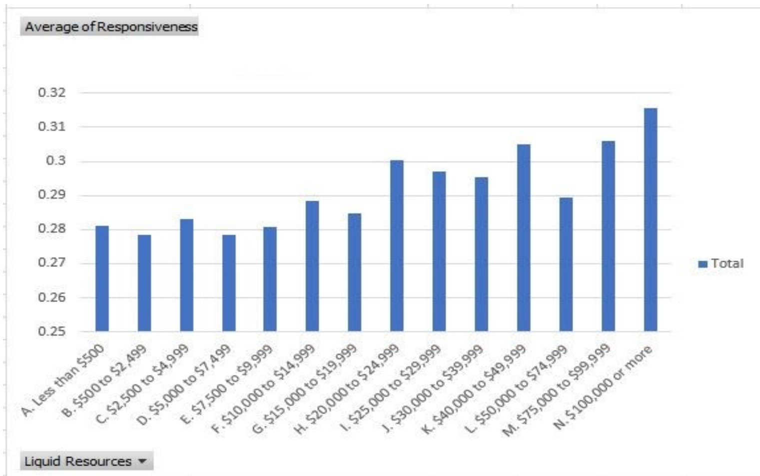


Figure A3: Responsiveness versus net worth

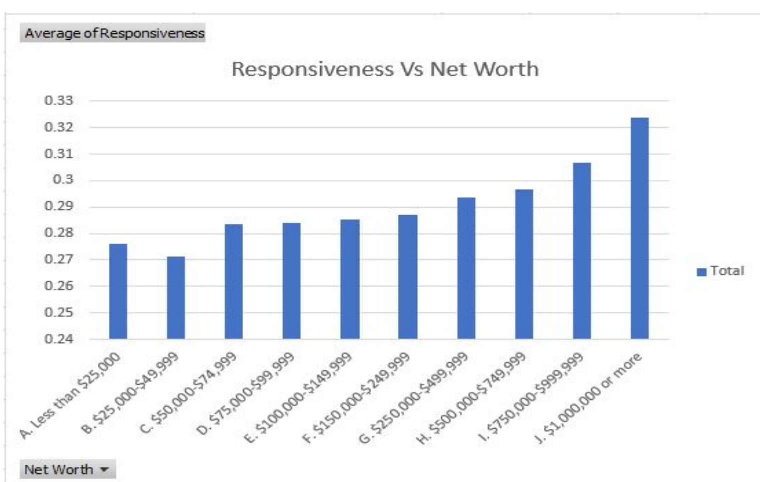


Figure A4: Responsiveness versus investments

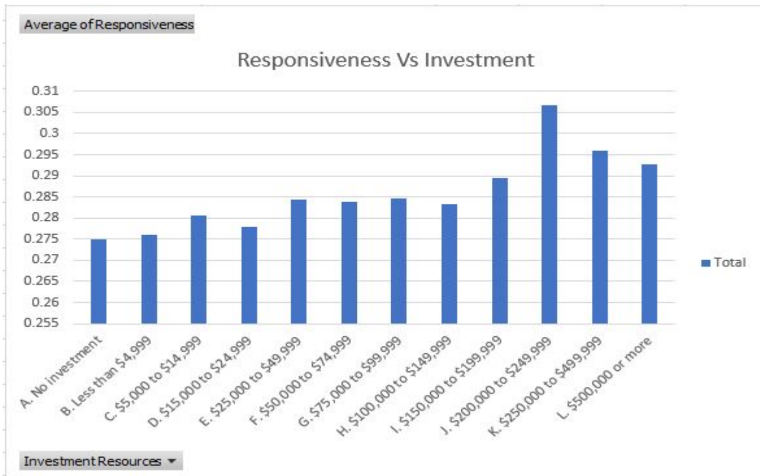
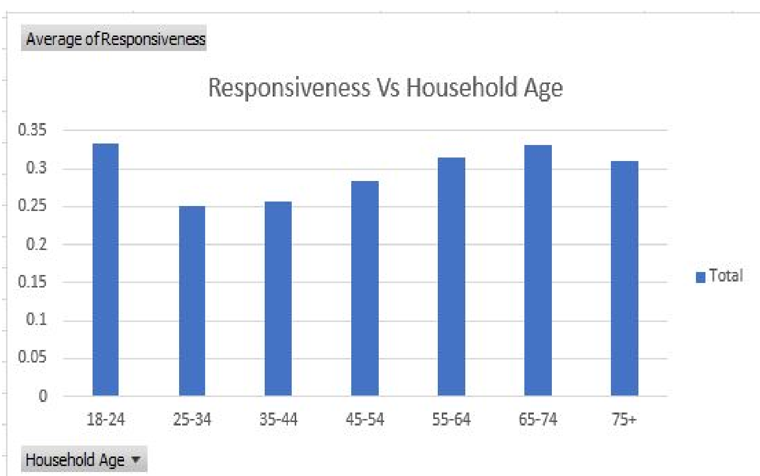


Figure A5: Responsiveness versus household average age



A.2 Robustness of borrower types around 60 days of delinquency

Since the responsiveness score goes down during 60 days of delinquency, arguably the behavior of borrowers across different tiers of responsiveness may vary differentially around the the threshold of 60 days of delinquency. We conduct a robustness check in Figure A6 and Figure A7 to ensure there still are four distinct segments in terms of levels of responsiveness.

Figure A6: Responsiveness versus inbound calls per outbound call for households with less than 60 days of delinquency

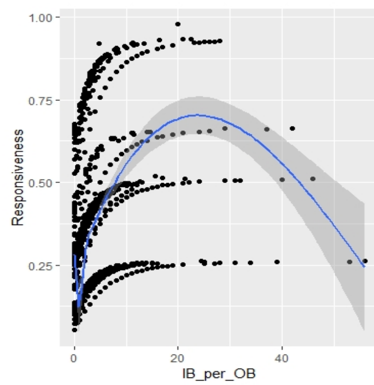
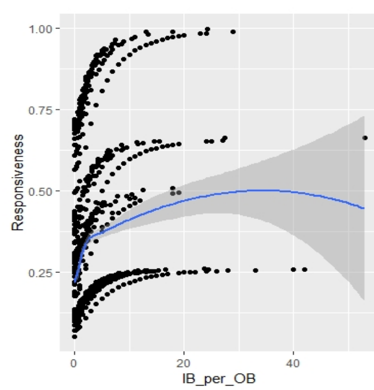


Figure A7: Responsiveness versus inbound calls per outbound call for households with more than 60 days of delinquency



A.3 Soft information and curative processes

This section examines the degree to which soft information collected during borrower-servicer communication can be used by issuers in their early buyout decisions. The data includes a text-based summary of the call log as well as several other types of soft information that inform the curative process. The data provides insight not only into the likelihood of the borrower reperforming, but also potential property, title, legal, document, and program compliance related issues that must be resolved before the loan can be considered cured. We exploit this soft information to examine whether it can be used to identify loans that are more likely to self-cure, require a loan modification, or default.

We find clusters of similar words to our pre-specified keywords (chosen from qualitative expert judgement and historical perspective). These similar words can be converted to vectors and can be used to quantify higher-dimensional *soft* information (beyond positive or negative sentiment). To achieve this objective, we first plot the main adverse delinquency states, namely, delinquent, bankruptcy, foreclosure, reo, short sale, using T-SNE, a.k.a, stochastic nearest neighbor embedding (see van der Maaten and Hinton (2008) for methodology), which is just a two-dimensional visualization of the clusters of similar words in Figure A8. The T-SNE plots help digest the high-dimensionality and complexity of the information content in these communications, illustrating their intuitive underlying structure as recurrent co-location of topics. The axes do not have any units or physical significance and is chosen automatically by the T-SNE algorithm to appropriately fit the important similar words in one diagram.

Next, we investigate and report the T-SNE results for several key events that happen during the lifecycle of a loan and extract intricate relationships and interactions with the main 5 delinquency states described above. I add each category on top of the main 5 delinquency states to visualize how each category results in one of these 5 delinquency states. As is evident in Figure A9, the legal cluster is mostly intertwined with the foreclosure cluster. There are two types of legal proceedings, one related to the title-issue from improper or missing promissory note for the residential mortgage. The other legal sub-cluster refers to the court proceedings related to the foreclosure process.

Figure A10 lists keywords contest, lien strip, title issue, repurchase related to the Title of the promissory note related to the mortgage and ownership of the res-

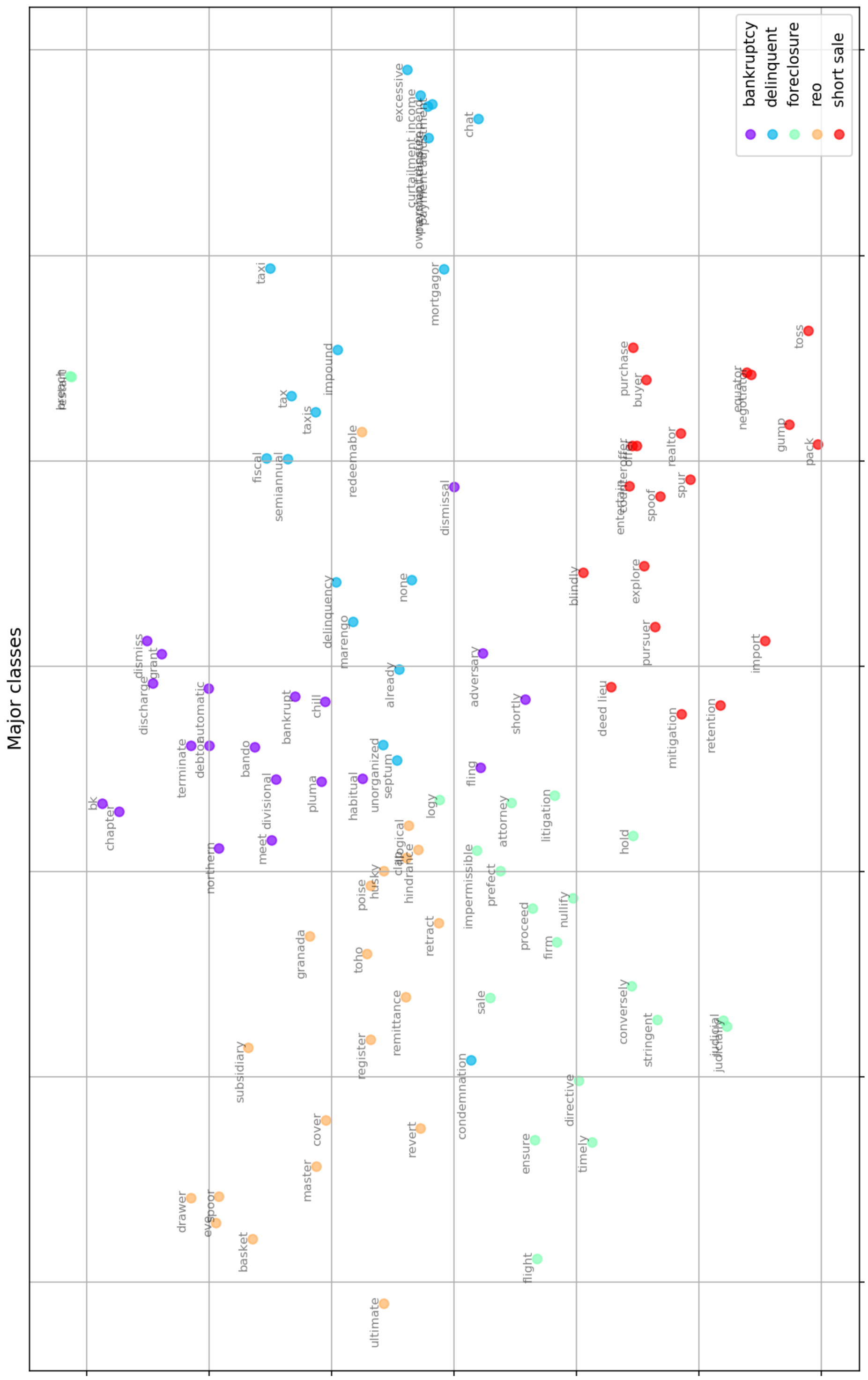


Figure A8: Key delinquency states
The t-SNE plots help appreciate the high-dimensionality and complexity of servicer comments, illustrating their intuitive underlying structure as recurrent co-location of topics.

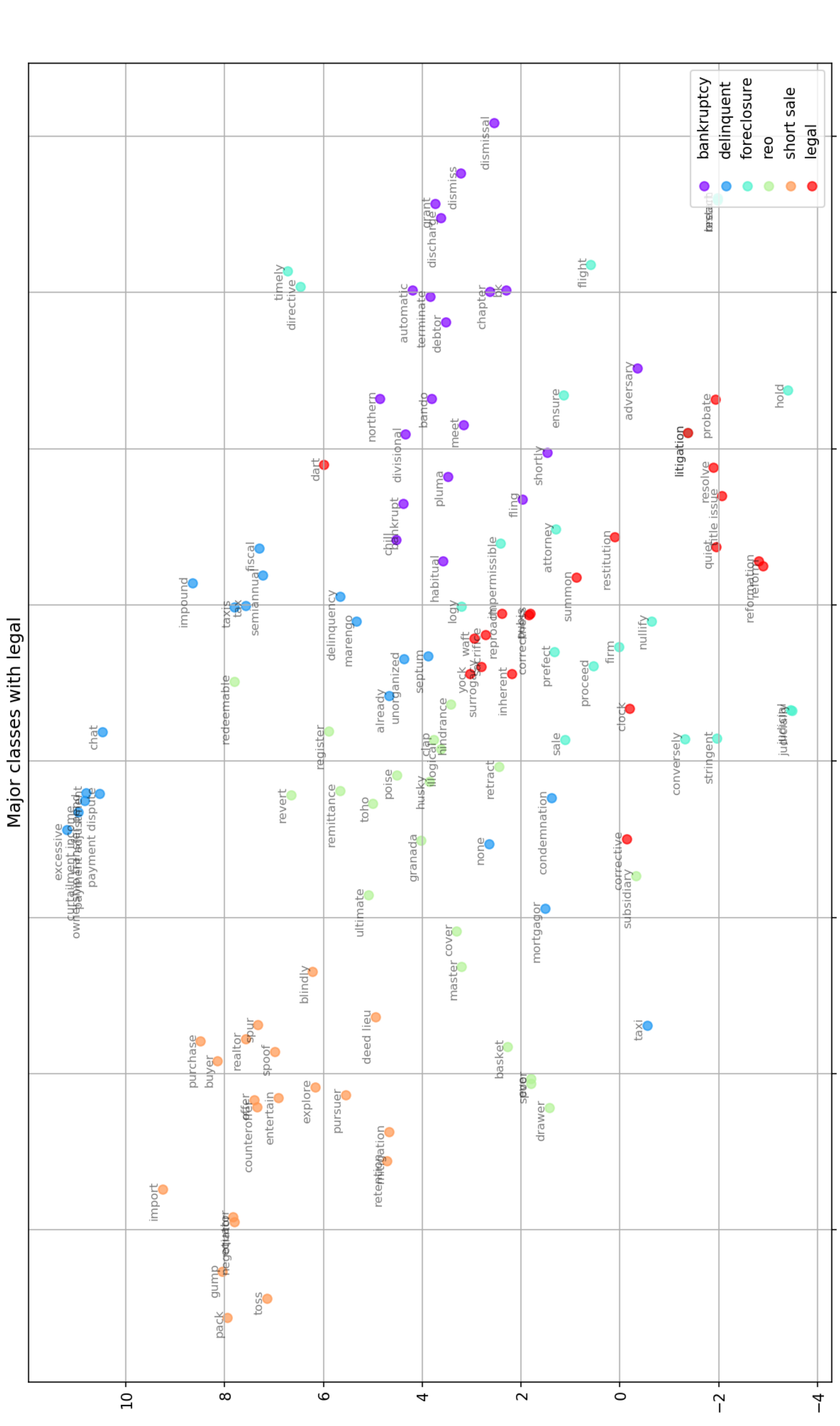


Figure A9: Key delinquency states and legal keywords
There are two types of legal proceedings, one related to the title-issue from improper or missing promissory note for the residential mortgage. The other "legal" sub-cluster refers to the court proceedings related to the foreclosure process.

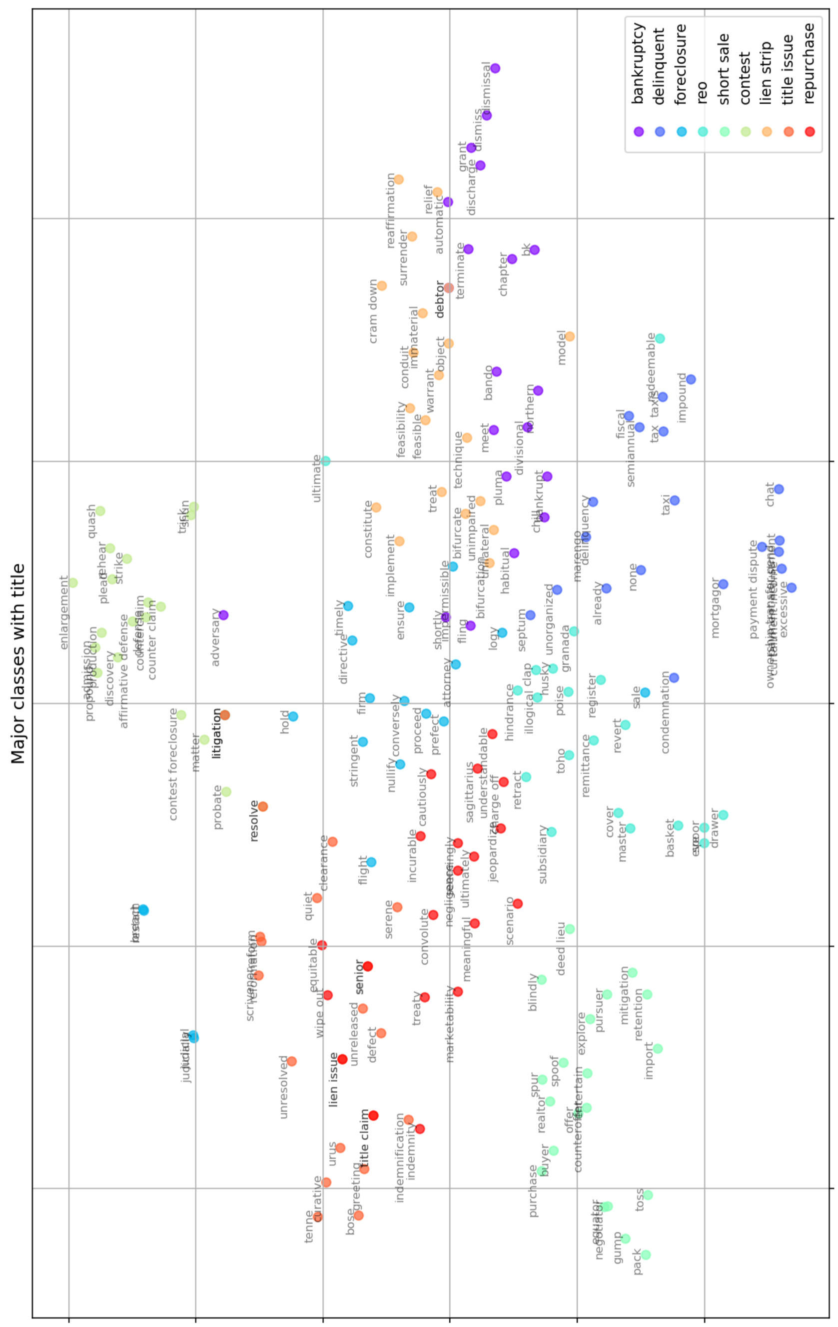


Figure A10: Key delinquency states and title-related keyword

idential property. The peach color lien strip has heavy intersection with the violet bankruptcy cluster, which could be an outcome of the involuntary bankruptcy proceedings following foreclosure petition by the lender. One light blue dot also defines foreclosure as the channel for lien strip leading to the bankruptcy process. The title issue and contest are intertwined due to obvious reasons. The contest cluster is closer to foreclosure and hence contest can be used as a leading indicator for foreclosure proceedings. The red color repurchase cluster has a broad intersection with several clusters. Repurchase can first be can a natural outcome of lien strip. A repurchase is also associated with an REO proceeding by the lender. One red dot in the light green short sale cluster indicates repurchase from short-sale.

A.4 Computation of reward for formation of buckets

To test our results in a dynamic setting, we first resort to simple linear models. First, we define a reward function as the ratio of *Collections* to the *Original Balance* of the loans. This is to ensure we account for the heterogeneity of different loans and create a measure which can be used for all loans across all times. The reward variable gives us an indication of the relative advantage of obtaining the *Collections* compared to the overall size of the loan at origination.

We use OLS regressions on continuous reward variable on loan delinquency states, servicer actions and their all possible interactions. In Table A1 Model 1, we document which delinquency states and servicer actions are salient in generating the reward for the servicer. After controlling for delinquency states and servicer action, we also document the statistically significant interactions of $(delinquencystate, serviceraction)$ so that we can shortlist them for RL later. For example, we see that given the loan is L30D, a loan modification offered by the servicer (L30D:Loan Mod) offers upfront fee before refinancing. Hence, the impact is positive. Similarly, when the loan status is L30D, the servicer action of filing a notice of intent without foreclosure proceedings (L30D: NOI Filed: Not in FC), has a negative impact on the servicer reward. Also, given that the loan is in foreclosure proceedings, a servicer action of REO is costly and geneerates negative reward. We further robustly test for the same results in Model 2, including borrower financial constraints emanating from unemployment or curtailment of income, access to insurance, financial sophistication, religion, language, family composition, length of residence, etc. and find they have marginal impact on the reward of the servicer.

In Table A2 we further test the reward on shortlisted delinquency states, servicer actions and $(state, action)$ pairs to find the direction of impact on the reward. This gives us a handle on which states and which actions the servicer should focus on while optimizing their reward based on borrower responsiveness and communication with the borrower. We then test for robustness of these results with zipcode fixed effect in panel 2, zipcode and month fixed effects in panel 3, all possible interactions of zipcode and months in panel 4 and finally with loan fixed effect which takes out any loan specific idiosyncrasies.

The frequency distribution of the reward informs us about severe clustering of rewards near zero, as expected. The mean is 0.04012 and the third quartile is 0.00693. Now we bucket the reward variables into five groups of equal width. Once, we round the rewards of these loans to zero, we have a frequency distribution of the remaining one-fourth of the observations in Figure A11. Now, we resort to $(delinquencystate, serviceraction)$ pair and use OLS regression to bucket these rewards in Figure A12 for later use in RL.

Figure A11: Histogram of reward beyond third quartile

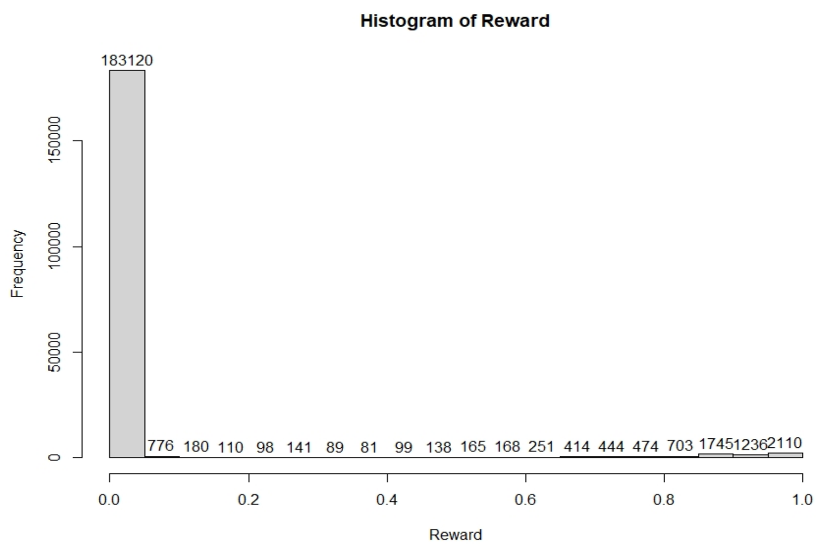


Figure A12: Reward buckets based on actual collections data regression results

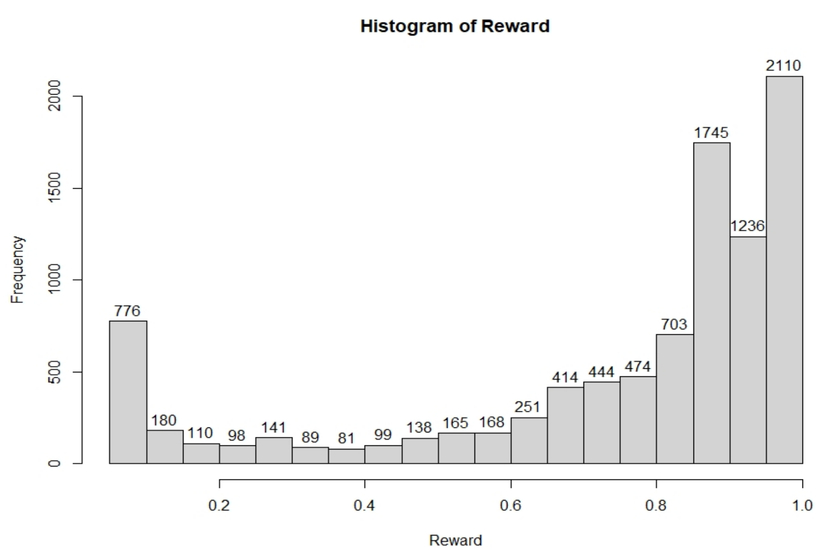


Table A1: Determinants of reward: We document which delinquency states and servicer actions are salient in generating the reward for the servicer.

	Model 1	Model 2
(Intercept)	0.03^{***}	0.10^{*}
30Days	-0.02^{*}	-0.03^{*}
L30D	0.12^{***}	0.12^{***}
ModificationinReview	-0.02^{*}	-0.03^{**}
NOIFiled:NotinFC	-0.02^{***}	-0.03^{***}
NULL	-0.02^{**}	-0.03^{***}
PendingClaim	0.02^{**}	0.01
PendDeedinLieu	0.05^{*}	0.08^{**}
PndFC	0.04^{***}	0.03^{***}
PendingShortSale	0.04^{**}	0.04^{**}
Performing	-0.02^{***}	-0.03^{***}
REO	0.27^{***}	0.25^{***}
BK:ModCompleted	0.13^{***}	0.13^{***}
L30D:ModCompleted	0.24^{***}	0.25^{***}
30Days:NOIFiled:NotinFC	0.02	0.03^{*}
L30D:NOIFiled:NotinFC	-0.03^{***}	-0.03^{**}
L30D:NShortRefi	-0.06^{**}	-0.07^{**}
L30D:NULL	-0.09^{***}	-0.10^{***}
FC:NULL	0.01	0.02^{*}
FC:PendingClaim	0.02	0.06^{***}
30Days:PndFC	-0.03^{*}	-0.03
BK:PndFC	-0.05^{***}	-0.03^{***}
L30D:PndFC	-0.02^{**}	-0.01
FC:PndFC	-0.04^{***}	-0.03^{***}
30Days:Performing	0.03^{*}	0.03^{*}
BK:Performing	0.04^{***}	0.05^{***}
L30D:Performing	-0.04^{***}	-0.04^{***}
FC:REO	-0.29^{***}	-0.30^{***}
UninsuredforHealth		0.00^{**}
PaychecktoPaycheckConsumers		0.00^{**}
FinancialHealthNewsletterSubscribers		0.00^{**}
ConservativeInvestmentStyleConsumers		0.00^{***}
ACAHealthInsurancePurchasers		0.00^{**}
BudgetMealPlanners		0.00^{**}
Person1LanguageCodeL1		0.28^{**}
Person1LanguageCodeN2		-0.22^{**}
Person1ReligionJ		0.23^{*}
Person1ReligionL		0.19^{**}
Curtailment_of_Income_Flag_Count		0.00^{***}
Unemployed_Flag_Count		0.00^{***}
LengthofResidence		-0.01^{***}
NumberofAdults		-0.00^{**}
PresenceofChildren		0.01^{***}
R^2	0.06	0.07
Adj. R^2	0.06	0.07
Num. obs.	308200	278969

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table A2: Determinants of Reward using 5 models of various fixed effects: We further test the reward on shortlisted delinquency states, servicer actions and $(state, action)$ pairs to find the direction of impact on the reward. We then test for robustness of these results with zipcode fixed effect in panel 2, zipcode and month fixed effects in panel 3, all possible interactions of zipcode and months in panel 4 and finally with loan fixed effect.

Reward	OLS	Zipcode Fixed Effect	Zip+Month Fixed Effect	Zip*Month Fixed Effect	+Loan Fixed Effect
30Days	-0.02	-0.02^{*}	-0.02^{*}	-0.04^{*}	-0.00
60Days	-0.02	-0.02^{*}	-0.02	0.01	0.04^{*}
L30D	0.11^{***}	0.13^{***}	0.13^{***}	0.08^{***}	0.10^{**}
ModinReview	-0.01	-0.02	-0.02^{**}	-0.01	0.02
NOIFiled:NotinFC	-0.02^{*}	-0.03^{***}	-0.03^{***}	-0.02^{*}	0.00
NShortRefi	-0.02	-0.03^{**}	-0.04^{***}	-0.22	-0.08
NULL	-0.02^{*}	-0.03^{***}	-0.03^{***}	-0.02	0.02
PendingClaim	0.02^{**}	0.03^{**}	0.03^{*}	0.02	-0.05
PendDeedinLieu	0.05^{*}	0.10^{**}	0.09^{**}	0.05	0.01
PndFC	0.04^{***}	0.05^{***}	0.05^{***}	0.02	-0.02
PendingPayoff	-0.03	-0.02^{*}	-0.03^{***}	-0.02	0.04^{*}
Pend3rdPartySale	0.06	0.08^{***}	0.09^{***}	0.08^{***}	0.03^{**}
Performing	-0.02^{**}	-0.02^{**}	-0.03^{***}	-0.03^{*}	0.02
RollingDlq	-0.03	-0.02^{**}	-0.02^{*}	-0.04^{**}	0.06
L30D:ModComplt	0.21^{***}	0.26^{***}	0.26^{***}	0.25^{***}	0.24^{***}
30Days:NOI:NotFC	0.02	0.02^{*}	0.03^{**}	0.05^{*}	0.03
60Days:NOI:NotFC	0.02	0.02^{*}	0.02^{*}	-0.00	-0.02
BK:NOI:NotFC	0.01	0.02	0.03^{*}	0.01	-0.01
30Days:NShortRefi	0.03	0.05^{***}	0.04^{**}	0.24	0.16
60Days:NShortRefi	0.04	0.05^{**}	0.05^{**}	0.24	0.10
BK:NULL	0.01	0.02^{*}	0.02^{*}	0.02	0.01
L30D:NULL	-0.09^{***}	-0.11^{***}	-0.11^{***}	-0.05^{*}	-0.06^{*}
FC:NULL	0.01	0.02^{*}	0.02^{*}	0.04^{*}	0.02
BK:PendingClaim	-0.04	-0.05^{***}	-0.04^{**}	-0.01	0.12^{**}
FC:PendingClaim	-0.00	0.05	0.05	0.09^{*}	0.22^{***}
FC:PendDeedinLieu	-0.06^{*}	-0.08	-0.08	-0.10^{*}	-0.06
30Days:PndFC	-0.04^{*}	-0.04^{**}	-0.04^{**}	0.00	0.05
60Days:PndFC	-0.04	-0.04^{***}	-0.04^{***}	-0.04^{*}	0.00
90Days:PndFC	-0.04	-0.05^{***}	-0.04^{***}	-0.02	0.02
BK:PndFC	-0.06^{***}	-0.06^{***}	-0.05^{***}	-0.03^{*}	0.02
L30D:PndFC	-0.03^{**}	-0.04	-0.04	0.01	0.06
FC:PndFC	-0.05^{***}	-0.06^{***}	-0.06^{***}	-0.02	0.03
PndSvcTrans:PndFC	-0.09^{*}	-0.10^{*}	-0.09^{*}	-0.15	-0.12
30Days:Performing	0.02	0.03^{**}	0.03^{**}	0.05^{**}	0.03
60Days:Performing	0.02	0.02^{*}	0.02^{*}	0.01	-0.01
BK:Performing	0.03^{*}	0.04^{*}	0.05^{**}	0.04^{*}	0.03
FC:REO	-0.21^{*}	-0.22	-0.22	-0.05^{***}	0.18^{**}
30Days:RollingDlq	0.02	0.03^{*}	0.03^{*}	0.10^{***}	0.03
60Days:RollingDlq	0.03	0.03^{**}	0.02	0.04	0.00
Num. obs.	187213	159169	159169	159169	159169
R^2	0.06	0.14	0.15	0.59	0.70

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$