On the Convergence of Credit Risk in Current Consumer Automobile Loans

Jackson P. Lautier, Vladimir Pozdnyakov, and Jun Yan*

March 22, 2023

ABSTRACT

Conditional credit risk of a current loan is understudied. Using large-sample statistics and asset-level consumer automobile asset-backed security data, we find default risk conditional on survival eventually converges for borrowers in disparate credit risk bands, a phenomenon we call credit risk convergence. We conservatively estimate that borrowers forwent $1,212 - $2,327 in savings through delayed prepayment and find the surprising result that current lower risk borrowers behave less efficiently than current higher risk borrowers. We also present visual evidence prepayments rose with used auto values and economic stimulus rather than financial acumen. Our results are robust to various sensitivity tests.

JEL Codes: C58, D11, D12, G32, G51, G53

Keywords— Adjustable premium loans, competing risks, consumer finance, Coronavirus, COVID-19, financial literacy, market inefficiency, Reg AB II

*All authors are at the University of Connecticut in the Department of Statistics. Contents of this presentation have benefited from seminar participants at Bentley University, the University of Connecticut (School of Business, Department of Finance), the University of Connecticut (Department of Economics), Université Concordia University (John Molson School of Business, Department of Finance), and personal correspondence with Brian Melzer, Jonathan A. Parker, and Jonathan Zinman. Any remaining errors are our own. Jackson P. Lautier was employed with Prudential Financial, Inc. from 2010-2019 but believes in good faith there are no conflicts of interest. The other authors have no potential conflicts to disclose. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DHE 1747453.

Correspondence: Jackson P. Lautier, University of Connecticut, Room 323, Philip E. Austin Building, 215 Glenbrook Road, Unit 4120, Storrs, Connecticut 06269-4120; e-mail: jackson.lautier@uconn.edu
In chronicling the loss curves for securitization pools of individual loans, there is a familiar pattern that even a junior credit analyst can sketch from memory: an initial rise in the early months of the securitization followed by a sustained flattening in the curve once the pool eventually settles into its long-term steady state. In higher risk or subprime pools of borrowers, the eventual loss percentage might be many multiples higher than lower risk or prime pools of borrowers, but the overall shape follows the familiar natural log-like pattern.\footnote{Junior analysts are trained to look for any sudden upward deviations in the pattern, which may indicate a rapid deterioration in the performance of the loans.}

It is striking that the loss curves all eventually flatten, despite the notable difference in cumulative totals. In other words, after a sustained period of performance, current loans appear to stay current, regardless of the loan’s initial risk classification. This observation, summarized in Figure 1 and formally studied herein, is the initial motivation of our efforts.

In attempting to price the default risk inherent within each pool of loans, it is well-accepted that riskier loans must pay higher interest rates. This is the common practice of risk-based pricing (e.g., Edelberg, 2006; Phillips, 2013). Traditionally, a borrower’s risk is assessed upon the initial loan application, and the borrowing cost comes through in the form of the annual percentage rate (APR) on the loan contract. In reflecting on the shape of the loss curves in Figure 1, it suggests that borrowers that remain current may eventually become better credits. Since the APR is set at the onset of the loan contract, it is possible that consumers paying a higher APR may eventually overpay in comparison to an updated risk assessment. This line of economic questioning is our second major item of study.\footnote{It may be tempting to point out the correct claim that it is the risky, high-interest rate loans that don’t default that offset losses from the risky, high-interest rate loans that do default. We briefly note here our data comes entirely from asset-backed securities, and so the lenders have already transferred default risk off of the balance sheet; they no longer have a direct interest in the performance of any loans we study. We return to this point in Section III.A.}

We thus present a paper with two intertwined contributions: a theoretical observation about the nature of consumer debt and empirical observations within consumer debt markets that suggest potential inefficiencies. At its heart, this is a work within the space of consumer finance, but it requires the use of statistical tools to empirically validate a current borrower’s declining credit risk over time. We now discuss each contribution in turn.

The first major contribution is the concept of credit risk convergence, in which borrowers of secured consumer automobile loans in different credit risk bands that remain current...
Figure 1: **Securitization Loss Curves.** A plot of three securitization loss curves, as a cumulative percentage of total defaults by the age of the securitization. The highest loss curve corresponds to the riskiest pool of loans in terms of traditional credit metrics. It is striking that all curves eventually flatten (i.e., after the dashed vertical line, age 40), despite the differences in underlying borrower credit quality. This hints at an observable phenomenon that we call *credit risk convergence*, and it is studied formally in Section II.

eventually converge in default risk.\(^3\) To empirical validate credit risk convergence, we rely on large sample asymptotic statistics and a survival (or time-to-event) analysis tool known as the hazard rate. The hazard rate measures the probability of an event conditional on survival, and it is thus the ideal quantity of interest for a current loan analysis. Due to some incomplete data challenges in working with loan data sampled from securitization pools, we take some effort to arrive at an estimator that is theoretical suitable for our application. In particular, we utilize a competing risks framework to estimate a cause-specific hazard

\(^3\)It should be stated that two consumers who completely repay their loans will trivially converge in credit risk to a zero probability of default. The major analysis, therefore, is determining if such convergence occurs prior to loan termination and, if so, its financial implications.
(CSH) rate, which allows us to model both conditional default and conditional prepayment probabilities. The estimator we utilize also has convenient large sample properties, which we state formally in Appendix A and prove in the Online Appendix A (a contribution to the statistical literature in its own right). Specifically, it yields a large sample hypothesis test to determine if the CSH rate for default between two different risk bands is different at a statistically significant level. If not, then we cannot claim the CSH rates are different between the two risk bands (i.e., conditional default risk has converged).

To perform an empirical assessment of credit risk convergence, we utilize consumer automobile loan performance data from publicly issued asset-backed securities (ABS) containing borrowers over a wide spectrum of different credit risk profiles. The asset-level performance data is available to the public through the Electronic Data Gathering, Analysis, and Retrieval (EDGAR) system operated by the Securities and Exchange Commission (SEC). We thus compile a unique data set of borrower attributes and monthly loan performance from four completed ABS bonds (i.e., CarMax, 2017; Ally, 2017; Santander, 2017b,a). We select loans to be as comparable as possible, and we assign risk bands based on the loan’s APR. In total, we consider 58,118 individual consumer automobile loans. For more data details, see Section I. We find evidence of convergence between disparate risk bands of 72-73 month auto loans between 10-52 months for current loans, depending on the two risk bands being compared. For details, see Section II.B and Table 2. We find these results are robust to sensitivity analysis considering the economic impact of COVID-19 (see Section II.C), collateral type, and the business model of the loan originator’s parent company (for the latter two, see Section II.D). The entirety of the credit risk convergence analysis may be found in Section II.

The second major contribution is the financial implications of our credit risk convergence results in light of the differences in APR due to risk-based pricing. In a study of hypothetical lender profitability in Section III.A, we demonstrate that the CSH rates for default may be

---

4All our analysis adjusts for prepayment behavior. In other words, while we consider both prepayments and repayments as “non-defaults”, the distribution of loan repayments and therefore all subsequent analysis is adjusted for the timing of observed prepayments.

5Recent regulatory changes have made such data available to the public for the first time (Securities and Exchange Commission, 2014, 2016), though it remains surprisingly underutilized in the literature.

6We also perform robustness analysis using a second set of 65,892 loans sampled from the bonds same set of four bonds but issued in 2019 (i.e., CarMax, 2019; Ally, 2019; Santander, 2019b,a) and a third sample of 16,412 loans from the set of 2017 ABS bonds. See Sections II.C and II.D for additional details.
used within an actuarial analysis to solve for a risk-adjusted rate of return conditional on loan survival. We demonstrate that lender profits are back-loaded, which is consistent with the insurance-like arrangement within pools of risky loans. We then transition to consider the consumer perspective in Section III.B. We first find meaningful outstanding balances even for mature loans, which suggests that refinance savings may be substantial, even late into a loan’s lifetime. We then estimate the potential savings assuming the average borrower in one risk band refinanced at the average rate in a higher risk band, once eligible based on our loan risk convergence point estimates. We find that the riskiest borrowers (deep subprime, subprime) can save between $11-62 dollars in monthly payments or $136-1,616 in total by refinancing. Our estimates suggest deep subprime and prime borrowers should refinance after about 42-50 months, when they become prime borrowers. We find evidence that these borrowers generally wait too long to refinance. In a surprise, we find that less risky loans (near-prime, prime) leave even more money on the table, with total savings ranging from $148-2,327 (or $12-39 in monthly payments). Our estimates suggest that near-prime and prime borrowers should refinance quickly, after about only one year, but they also generally wait too long. Hence, in a result counter to expectations about borrower sophistication, it is the near-prime and prime loans that behave less efficiently.

We then attempt to assess motivations for borrower behavior. A visual examination finds that borrower prepayment behavior is likely attributable to unusual increases in the value of used automobiles (and thus trade-in values) and economic stimulus payments rather than any self-assessment of an updated risk profile (see Figure 10 for details). Section III.B closes by suggesting market frictions may exist that prevent both borrowers and lenders alike from reducing these market inefficiencies. In hopes of encouraging related research, we also proffer that lenders may consider offering new loan products that reward borrowers for good performance or potential regulator interventions, among other potential solutions.

In attempting to contextualize our contributions, we feel it is best to compare related work within consumer finance. A perhaps related field is payday lending and so-called high-cost alternative financial service (AFS) lending. Melzer (2011) finds not only no evidence that payday loans reduce economic distress but that access to such loans may exacerbate financial difficulties for low income individuals. Bertrand and Morse (2011) use a field experiment to show that a disclosure designed to make the cumulative costs of payday lending more apparent led to an 11% reduction in the utilization of such loans in the subsequent four
months. In times of unexpected financial distress, Morse (2011) finds some evidence payday lenders offer a positive service. Financial literacy also plays a role in consumers electing to use high-cost methods of borrowing, as Lusardi and de Bassa Scheresberg (2013) find that consumers with higher financial literacy are much less likely to have utilized high-cost borrowing methods. Lim et al. (2014) stress the need to better understand the payday loan industry in the context of social work advocacy. Robb et al. (2015) reach a similar financial literacy conclusion as Lusardi and de Bassa Scheresberg (2013), but they also note that borrower overconfidence of self-assessed financial acumen increases the chance of using high-cost AFS. Dobbie et al. (2021) find significant bias against immigrants and older applicants when reviewing a high-cost lender’s long-run profit measure for a U.K. high-cost lender. The model of Allcott et al. (2021) suggests that banning payday lending would reduce economic welfare, though they note that limits on repeat borrowing might increase welfare. Payday lending and AFS are generally more short-term products, however, and so research has not considered the dynamics of default risk of current borrowers over time.

In a closer relative, credit card lending has also attracted a similar level of study. Both Ausubel (1991) and Calem and Mester (1995) report that credit card interest rates remain sticky relative to the cost of funds, which does not conform to behavioral assumptions of a perfect competition model. Gross and Souleles (2002) note that conventional models cannot explain why borrowers carry a high-interest balance on their credit cards and simultaneously hold low yielding assets (the so-called “credit card puzzle”; see a proposed model based on credit volatility in Fulford (2015)). Alan and Loranth (2013) find that only low-risk borrowers that fully utilize their credit cards reduced credit demand when presented with an increase in interest rates. They also find that a rate increase of 5% would significantly increase lender profitability without inducing more delinquencies over a short time horizon. Agarwal et al. (2014) find that regulation in the credit card market can benefit consumers by analyzing the effectiveness of the Credit Card Accountability Responsibility and Disclosure (CARD) Act. Of note, they estimate the CARD Act saved consumers $11.9 billion a year. Heidhues and Köszegi (2016) find that credit card lenders use information about a borrower’s naïveté to increase interest revenue. We consider consumer automobile loans, however, and we again note these studies do not consider the borrower risk profile over time.

Our work naturally aligns with previous studies on the consumer automobile lending space, which have found that consumers are subject to various forms of troubling economic
behavior. For example, there is evidence of racial discrimination found in studies that span decades (e.g., Ayres and Siegelman, 1995; Edelberg, 2007; Butler et al., 2022). For an overview of the used car industry and the challenges presented to poor consumers in purchasing and keeping transportation, see Karger (2003). Adams et al. (2009) look at the effect of borrower liquidity on short-term purchase behavior within the subprime auto market. Namely, they observe sharp increases in demand during tax rebate season and high sensitivity to minimum down payment requirements. Grunewald et al. (2020) find that arrangements between auto dealers and lenders lead to incentives that increase loan prices. They also find consumers are less responsive to finance charges than vehicle charges and that consumers benefit when dealers do not have discretion to price loans. While consumer auto loans and subprime borrowers have attracted significant attention, we again do not find consideration of the borrower risk profile over the lifespan of the loan.

So how does one classify this paper? As this previous research indicates, the consumer already faces an uphill battle in most financial transactions. Given the complexity of modern financial markets, pervasive financial ignorance is unsurprising and contributes meaningfully to wealth inequality (Campbell, 2016). In thinking about economic inequality, Pressman and Scott (2009) argue that consumer debt should be a component of measuring poverty and economic inequality. Perhaps partially motivated by this backdrop, Zingales (2015) asks us to “blow the whistle” on financial practices that do not work. In this spirit, we have identified an ubiquitous financial practice in risk-based pricing that can be improved, updated, and modernized to better serve a consumer in a core lending space necessary to purchase an essential economic asset: the automobile.

The paper proceeds as follows. The data and related details are introduced first in Section I. Section II formalizes the concept of credit risk convergence. Section III uses the results of Section II to perform financial analysis, and Section IV concludes. Supporting details may be found the Appendix or Online Appendix and are referenced where appropriate.

---

7For an additional example, lenders knowingly hide interest rates to select for borrowers who make decisions based on payment size (Stango and Zinman, 2011).
I. Data

On September 24, 2014, the SEC adopted significant revisions to Regulation AB and other rules governing the offering, disclosure, and reporting for ABS (Securities and Exchange Commission, 2014). One component of these large scale revisions, which took effect November 23, 2016, has required public issuers of ABS to make freely available pertinent loan-level information and payment performance on a monthly basis (Securities and Exchange Commission, 2016). We have utilized the EDGAR system operated by the SEC to obtain complete loan-level performance data for the consumer automobile loan ABS bonds CarMax Auto Owner Trust 2017-2 (CarMax, 2017), Ally Auto Receivables Trust 2017-3 (Ally, 2017), Santander Drive Auto Receivables Trust 2017-2 (Santander, 2017b), and Drive Auto Receivables Trust 2017-1 (Santander, 2017a). Henceforth, we will use the standard industry shorthand of CARMX, AART, SDART, and DRIVE to refer to each of these four bonds, respectively.

By count, the total number of loans for CARMX, AART, SDART, and DRIVE were 55,000, 67,797, 80,636, and 72,515, respectively.

The bonds were selected because of the credit profile of the underlying loans, the lack of a direct connection to a specific auto manufacturer, and the observation window of each bond’s performance spanning approximately the same macroeconomic environment. To elaborate, the credit profile of a DRIVE borrower is generally deep subprime to subprime, SDART is subprime to near-prime, CARMX is near-prime to prime, and AART is prime to super-prime.\(^8\) Thus, the collection of all four bonds taken together span the full credit spectrum of individual borrowers. Figure 2 in Section I.B provides additional details. Further, it is common that an auto manufacturer will originate loans using its financial subsidiary (e.g., Ford Credit Auto Owner Trust). The bonds selected do not have a direct connection to a specific auto manufacturer, however, and so we may allay concerns our results may be influenced by oversampling loans secured by a specific brand of automobile.\(^9\) Furthermore, the bonds were selected to span approximately the same months to ensure all underlying loans were

---

\(^8\)The standard definitions of the terms deep subprime to super prime stem from the borrower’s credit score. Specifically, credit scores below 580 are considered “deep subprime”, credit scores between 580-619 are “subprime”, 620-659 is “near-prime”, 660-719 is “prime”, and credit scores of 720 and above are “super-prime” Consumer Financial Protection Bureau (2019).

\(^9\)It is true that the business objectives of CarMax, a used auto sales company, will differ from those of the traditional banks of Santander and Ally. We sensitivity test this point in the robustness checks of Section II.D.
subject to the same macroeconomic environment. Specifically, CARMX, AART, SDART, and DRIVE began actively paying in March, April, May, and April of 2017, respectively, and each trust was active for 50, 44, 52, and 52 months, respectively.

We elected to use consumer automobile loans for two reasons. The first is the practicality of subject matter expertise of the lead author to the nuances of the auto loan ABS market (e.g., which bonds to select, obtainment of data, etc.). Second, consumer auto loans typically represent a high priority of payment for a borrower amid many potential monthly credit obligations: housing, auto loans, credit cards, student loans, consumer loans, etc. In other words, if we do not observe a declining conditional default risk over time for a secured, high-priority of payment loan like a consumer auto loan, then it likely is not an observable phenomenon in other types of consumer loans with a lower priority of payment.

I.A. Loan Selection and Defining Risk Bands

To ensure the underlying loans in our analysis are as comparable as possible, we employ a number of filtering mechanisms. First, we remove any loan contracts that include a co-borrower. Second, we require each loan to have been underwritten to the level of “stated not verified” (obligorIncomeVerificationLevelCode), which is a prescribed description of the amount of verification done to a borrower’s stated income level on an initial loan application (Securities and Exchange Commission, 2016). Third, we remove all loans originated with any form of subvention (i.e., additional financial incentives, such as added trade-in compensation or price reductions on the final sale price). We then require all loans to correspond to the sale of a used vehicle. We further drop any loan with a current status of “repossessed” as of the first available reporting month of the corresponding ABS. Further, to minimize the chance of inadvertently including a loan that has been previously refinanced or modified, we only consider loans younger than 18 months as of the first available ABS reporting month. For loan term, we only include loans with an original term of 72 or 73 months. As a final data integrity check, we remove any loans that did not pay enough total principle to pay-off

---

10 On this point, the unwritten code of traders in the consumer ABS space is “you can live in your car, but you can’t drive your house to work”.

11 This was mainly to keep the loans from CARMX, of which used cars predominate. We sensitivity test this requirement in the robustness checks of Section II.D.

12 Pragmatically, the most common loan term in the data was 72/73 months, and so our loan term choice allows us to maximize the sample size.
the outstanding balance as of the first month the trust was active and paying but had a missing value (NA) for the outstanding balance in the final month the trust was active and paying. In other words, the loan outcome was not clear from the data; the loan did not pay enough principal to pay off the outstanding balance nor default but stopped reporting monthly payment data. In total, this final data integrity check impacts only 2,630 or 4.3% of the filtered loan population. We are left with 58,118 individual consumer auto loan contracts in total, summary details of which may be found in Section I.B. Complete replication code and other data details may be found in the online supplementary material.

Next, we assign each loan into a credit risk category or risk band depending on the original interest rate \( \text{originalInterestRatePercentage} \) assigned to the contracted loan. The interest rate is the ideal measure of perceived borrower risk within a risk-based pricing framework (Edelberg, 2006; Phillips, 2013) because a borrower’s risk profile is a multidimensional function of factors like credit score, loan amount, down payment percentage (% down), vehicle or collateral value, income, payment-to-income (PTI), etc., in addition to many of the factors of which we have already filtered. In other words, given we have already controlled for prevailing market rates by selecting loans originated within a close temporal proximity, the interest rate serves as the market’s best estimate of a loan’s risk profile.

We now formalize this discussion slightly. Working from Phillips (2013), a borrower’s interest rate in risk band \( a \), \( r_a \), is

\[
r_a = r_c + m + l_a,
\]

where \( r_c \) is the cost of capital, \( m \) is the added profit margin, and \( l_a \) is a factor that varies by risk band. The components \( r_c \) and \( m \) will be shared by all risk bands, and so there exists some functional relationship

\[
l_a = f(\text{PTI}, \% \text{ down}, \text{Loan Amt}, \text{Vehicle Val}, \ldots).
\]

Rather than attempt to recover this unknown \( f \), therefore, we are in effect treating the lender’s credit scoring model as an accurate reflection of the borrower’s risk.\(^{13}\) Specifically, we assign borrower’s with an APR of 0-5% to the super-prime risk band, 5-10% to the

\(^{13}\text{Indeed, these models are often quite sophisticated (Einav et al., 2012).}\)
prime risk band, 10-15% to the near-prime risk band, 15-20% to the subprime risk band, and 20%+ to the deep subprime risk band. In a review of Figure 2 in Section I.B, we can see that the risk bands assigned by interest rate compare favorably to the traditional credit score borrower risk band definition (Consumer Financial Protection Bureau, 2019).

I.B. Summary of Selected Loans

After the data cleaning and filtering of Section I.A, we have payment performance for 58,118 consumer auto loans that span a wide range of borrower credit quality based on the traditional credit score metric. Figure 2 presents a summary of each bond by obligor credit score and interest rate as of loan origination. Judging by credit score, we can see that generally DRIVE is a deep subprime to subprime pool of borrowers, SDART is a subprime to near-prime pool, CARMX is a near-prime to prime pool, and AART is a prime to super-prime pool of borrowers (Consumer Financial Protection Bureau, 2019). As expected in a risk-based pricing framework, the density plot of each borrower’s interest rate has an inverse relationship to the density plot of each borrower’s credit score: lower credit scores correspond to higher interest rates (compare the first two rows of Figure 2). As such, we can see the annual percentage rates (APRs) are higher for the DRIVE and SDART bonds, generally sitting within a range around 20% and then declining to under 15% for CARMX and finally under 10% for AART. The bottom two rows of Figure 2 demonstrate that defining risk bands by interest rate corresponds closely to the traditional credit score risk band definitions (Consumer Financial Protection Bureau, 2019), as the expected inverse relationship holds.

The loans are well dispersed geographically among all 50 states and Washington, D.C., with the top five concentrations of Texas (13%), Florida (12%), California (9%), Georgia (7%), and North Carolina (4%). Similarly, the loans are well diversified among auto manufacturers, with the top five concentrations of Nissan (13%), Chevrolet (10%), Ford (7%), Toyota (7%), and Hyundai (7%). Thus, our sample is not overly representative to one state-level economic locale or auto manufacturer. For additional details on the makeup of the loans, see the associated prospectuses (Ally, 2017; CarMax, 2017; Santander, 2017a,b).

Table 1 provides a summary of borrower counts by bond and performance. The total pool of 58,118 loans is weighted towards deep subprime and subprime borrowers, which are each 37% of the total and together 74%. Similarly, DRIVE and SDART supply around 85%
Figure 2: **Borrower Credit Profile and APR by Bond, Risk Band Classification.** A summary of the borrower credit profiles (1st row) and charged APR (2nd row) of the 58,118 filtered consumer automobile loans used in the analysis of Sections II and III by ABS bonds CarMax Auto Owner Trust 2017-2 (CarMax, 2017) (CARMX, 6,835), Ally Auto Receivables Trust 2017-3 (Ally, 2017) (AART, 2,171), Santander Drive Auto Receivables Trust 2017-2 (Santander, 2017b) (SDART, 20,192), and Drive Auto Receivables Trust 2017-1 (Santander, 2017a) (DRIVE, 28,920). The bottom two rows show the distribution of credit scores and interest rates, respectively, by our APR-based risk band classification: super-prime (0-5%), prime (5-10%), near-prime (10-15%), subprime (15-20%), and deep subprime (20%+) for the same set of 58,118 loans.

of the total loans in our sample. The smallest risk band is super-prime, which totals 2,179 loans for 4% of the total of 58,118.\footnote{Generally speaking, it is difficult to find a securitization pool of super-prime borrowers that has not been directly issued by an auto manufacturer. This explains why our sample is more heavily-weighted to deep subprime and subprime borrowers. Conversely, it is difficult to find deep subprime and subprime borrowers within securitization pools that have been issued by auto manufacturer financing subsidiaries; there is a give-and-take. That said, our asymptotic results scale by sample size, so the confidence interval width adjusts appropriately. Furthermore, even the smallest sample is quite robust (2,179) for the statistical methods we apply throughout.}
In terms of loan performance, we can observe some clear trends in Table 1. First, more than half of all deep subprime risk band loans defaulted, and this percentage declines by risk band until super-prime, in which only 4% of loans defaulted during the observation window. We also see that performance is fairly consistent by risk band, even among different bonds. For example, super-prime default percentages are within a tight range (3-6%) across each bond. The same may be said for deep subprime defaults. We see some wider ranges in the default percentages of the subprime (33-40%), prime (8-19%), and near-prime (17-24%) risk bands by bond, but they remain close enough to suggest there is not a worrisome difference between the credit scoring models employed by each different issuer. Overall, the percentage of defaulted loans declines as the credit quality of the risk band increases. This is further evidence that our APR-based risk band definition has yielded appropriate classification results.

II. Credit Risk Convergence

This section is the main theoretical contribution of this work to the field of finance: credit risk convergence. We begin with a review of the relevant statistical results in Section II.A. We will introduce the field of survival analysis, its subfield of competing risks, and its history in default modeling. We then present the estimator we employ and the associated large sample statistical hypothesis test relied on throughout. The formal statements are available for reference in Appendix A, and we provide complete proofs in the Online Appendix A. We then move to using these statistical results for an empirical study on the ABS data of Section I. That is, the empirical evidence for credit risk convergence may be found in Section II.B. Because our data spans the economic events of the Coronavirus pandemic, we make the obligatory remarks and robustness analysis in Section II.C. This section closes with Section II.D, which is additional sensitivity analysis related the loan filtering of Section I.A and potential generalizations of credit risk convergence to other types of debt.

II.A. Relevant Statistical Results

From an economic perspective, not all defaults are equivalent. For example, there is an obvious profitability difference between a loan that defaults shortly after it is originated...
Table 1: **Borrower Counts by Risk Band, Bond, and Loan Outcome.** Summary statistics and loan outcomes of the 58,118 filtered consumer automobile loans summarized in Figure 2. Percentages may not total to 100% due to rounding.

| Deep Prime | Subprime | Near-Prime | Prime | Super-Prime | Total |
|------------|----------|------------|-------|-------------|-------|
| **Total**  | 21,630 (37%) | 21,332 (37%) | 6,677 (11%) | 6,300 (11%) | 2,179 (4%) | 58,118 (100%) |
| **DRIVE**  | 14,079 (65%) | 12,884 (60%) | 1,443 (22%) | 220 (3%) | 294 (13%) | 28,920 (50%) |
| **SDART**  | 7,551 (35%) | 8,327 (39%) | 2,782 (42%) | 861 (14%) | 671 (31%) | 20,192 (35%) |
| **CARMX**  | 0 (0%) | 120 (1%) | 2,128 (32%) | 3,752 (60%) | 835 (38%) | 6,835 (12%) |
| **AART**   | 0 (0%) | 1 (0%) | 324 (5%) | 1,467 (23%) | 379 (17%) | 2,171 (4%) |
| **Total**  | 21,630 (100%) | 21,332 (100%) | 6,677 (100%) | 6,300 (100%) | 2,179 (100%) | 58,118 (100%) |

| Defaulted | 11,210 (52%) | 7,900 (37%) | 1,422 (21%) | 624 (10%) | 92 (4%) | 21,248 (37%) |
| Censored  | 3,547 (16%) | 4,599 (22%) | 1,997 (30%) | 2,556 (41%) | 948 (44%) | 13,647 (23%) |
| Repaid    | 6,873 (32%) | 8,833 (41%) | 3,258 (49%) | 3,120 (50%) | 1,139 (52%) | 23,223 (40%) |
| **Total**  | 21,630 (100%) | 21,332 (100%) | 6,677 (100%) | 6,300 (100%) | 2,179 (100%) | 58,118 (100%) |

| Defaulted | 7,518 (53%) | 5,115 (40%) | 351 (24%) | 42 (19%) | 14 (5%) | 13,040 (45%) |
| Censored  | 2,214 (16%) | 2,641 (20%) | 324 (22%) | 60 (27%) | 119 (40%) | 5,358 (19%) |
| Repaid    | 4,347 (31%) | 5,128 (40%) | 768 (53%) | 118 (54%) | 161 (55%) | 10,522 (36%) |
| **Total**  | 14,079 (100%) | 12,884 (100%) | 1,443 (100%) | 220 (100%) | 294 (100%) | 28,920 (100%) |

| Defaulted | 3,692 (49%) | 2,740 (33%) | 590 (21%) | 105 (12%) | 29 (4%) | 7,156 (35%) |
| Censored  | 1,333 (18%) | 1,915 (23%) | 715 (26%) | 255 (30%) | 299 (45%) | 4,517 (22%) |
| Repaid    | 2,526 (33%) | 3,672 (44%) | 1,477 (33%) | 501 (58%) | 343 (51%) | 8,519 (42%) |
| **Total**  | 7,551 (100%) | 8,327 (100%) | 2,782 (100%) | 861 (100%) | 671 (100%) | 20,192 (100%) |

| Defaulted | 0 | 45 (38%) | 427 (20%) | 296 (8%) | 25 (3%) | 793 (12%) |
| Censored  | 0 | 43 (36%) | 854 (40%) | 1,736 (46%) | 392 (47%) | 3,025 (44%) |
| Repaid    | 0 | 32 (27%) | 847 (40%) | 1,720 (46%) | 418 (50%) | 3,017 (44%) |
| **Total**  | 0 | 120 (100%) | 2,128 (100%) | 3,752 (100%) | 835 (100%) | 6,835 (100%) |

| Defaulted | 0 | 0 (0%) | 54 (17%) | 181 (12%) | 24 (6%) | 259 (12%) |
| Censored  | 0 | 0 (0%) | 104 (32%) | 505 (34%) | 138 (36%) | 747 (34%) |
| Repaid    | 0 | 1 (100%) | 166 (51%) | 781 (53%) | 217 (57%) | 1,165 (54%) |
| **Total**  | 0 | 1 (100%) | 324 (100%) | 1,467 (100%) | 379 (100%) | 2,171 (100%) |

versus a loan that defaults after a much longer period of time: a loan that makes more payments before defaulting will be more profitable, ceteris paribus. Therefore, what we are after is a time-to-event distribution estimate, where the general event of interest is the end of a loan’s payments. Indeed, we require this information to adequately address our research question centered around analyzing a loan’s conditional probability of default given

---

15 The same may be said for prepayments. We note that all of our results adjust for prepayments, and these prepayment probabilities are estimable using these techniques. See Section III.B for additional details.
its survival. We are thus in the realm of \textit{survival analysis}, which is a branch of statistics dedicated to estimating a random time-to-event distribution.\footnote{As a nuanced but important theoretical point of emphasis, our data is from pools of consumer automobile loans found in publicly traded ABS (see Section I for details). Thus, we must consider a model and subsequent estimator consistent with working in both discrete-time and with incomplete data subject to random left-truncation and random right-censoring. For extended details on these incomplete data challenges, see the discrete-time work of Lautier et al. (2021) for the case of left-truncation and Lautier et al. (2023) for the discrete-time case of both left-truncation and right-censoring in the context of modeling a consumer lease asset-backed security.}

In addition to estimating a time-to-event random variable, we also desire to distinguish between the type of event. Again, from an economic perspective, this is natural; a loan that is repaid (or prepaid) in a given month is more profitable than a loan that defaults in the same month, ceteris paribus. Succinctly, we wish to differentiate between loans ending in default and loans ending in prepayment. To do so, we can define the problem in terms of a \textit{competing risks} framework, which is a specialized branch of survival analysis.

The literature in this field is substantial. It is common to specify a competing risk model using a cause-specific hazard rate with covariates, see for example Prentice et al. (1978) and the well-known semiparametric proportional hazards model for the cumulative incidence function in Fine and Gray (1999). The work has since been anthologized into textbooks, such as Crowder (2001), Pintilie (2006), and Kalbfleisch and Prentice (2011, Chapter 8). For related work in discrete-time, see Tutz and Schmid (2016, Chapter 8), Lee et al. (2018b), and Schmid and Berger (2021). For the incomplete data case of left-truncation and right-censoring with covariates, see Geskus (2011).

Our framing of the problem is different, however. We will be working in terms of a multistate process (e.g., Andersen et al. (1993, Example III.1.5) or Beyersmann et al. (2009)) and thus may estimate the conditional cause-specific hazard rates directly. This is similar to recovering a bivariate distribution function in the presence of left-truncation and right-censoring (e.g., Sankaran and Antony, 2007; Dai et al., 2016), but we do not assume two lifetimes. Instead, we will be using a multistate process adjusted for left-truncation and right-censoring (e.g., Andersen et al. (1993, Example IV.1.7) for the absolutely continuous lifetime distribution case) in discrete-time (e.g., Andersen et al. (1993, pg. 94) for the general discrete case without incomplete data) but over a finite time horizon for two competing events. Specifically, we will generalize the discrete-time, left-truncation and right-censoring work of Lautier et al. (2023) to the case of two competing events: default and repayment.
Competing risks have been used to model loans successfully in the past. For an early comparison of various survival analysis techniques applied to credit scoring of personal loans, including competing risks, see Banasik et al. (1999). Similarly, Stepanova and Thomas (2002) consider competing risks to examine early repayment behavior of personal loans. For a more recent benchmark study, see Dirick et al. (2017). In terms of more focused studies on competing risks to model credit risk, De Leonardis and Rocci (2008) use the Cox proportional hazard method adjusted for discrete-time and applied to small and middle-sized Italian firms. In our setting, however, we will not be assuming the Cox method. Zhang et al. (2019) provide a mixture cure model under competing risks for the purposes of credit scoring peer-to-peer consumer loans. The model uses a latent failure times approach and assumes independence between the two competing risks (prepayment and default), however, which does not fit our specifications. Wycinka (2019) uses regression analysis to model the probability of default over time for consumer loans, which differs from our setting as well. For a specific study on time-varying covariates within the context of competing risks for the semi-parametric Cox model, see Thackham and Ma (2022). More recently and in the machine learning space, Blumenstock et al. (2022) evaluate the predictive performance of DeepHit (Lee et al., 2018a), a deep learning-based competing risk model and random survival forests in the context of US mortgages, and Frydman and Matuszyk (2022) use random survival forests for competing risks (Ishwaran et al., 2014) in an application to automobile leases from a Polish financial institution. While the number and breadth of these studies is extensive, we were unable to find a previous method that precisely meets our theoretical statistical requirements.\footnote{For the statistically inclined reader, these requirements are the ability to handle two competing events in discrete-time over a finite time horizon adjusted for both random left-truncation and right-censoring.}

We now present the mathematical details of the estimator in the context of an automobile loan ABS. We will follow the notation of Lautier et al. (2023) and for completeness include some details regarding accounting for incomplete data. Define the random time until a loan contract ends by the random variable $X$. The classical quantity of interest in survival analysis is the hazard rate, which in discrete-time represents the probability of a loan contract terminating in month $x$, given a loan has survived until month $x$. We denote the hazard rate by the traditional, $\lambda$, and so formally,

$$\lambda(x) = \Pr(X = x \mid X \geq x) = \frac{\Pr(X = x)}{\Pr(X \geq x)}. \quad (1)$$
Because we desire to model the probability of loan payments terminating given a loan remains current, it is clear that (1) is the ideal quantity of interest. Additionally, let $F$ represent the cumulative distribution function (cdf) of $X$. If we can reliably estimate (1), we can recover the complete distribution of $X$ by the uniqueness of the cdf since

$$1 - F(x-) = \Pr(X \geq x) = \prod_{x_{\min} \leq k < x} \{1 - \lambda(k)\},$$

where $x_{\min}$ is the lower bound of the distribution of $X$ and we take the the convention $\prod_{k=x_{\min}+1}^{x_{\min}} \{1 - \lambda(k)\} = 1$.

We now account for incomplete data. To address random left-truncation, let $Y$ represent the left-truncation random variable, which is a shifted random variable derived from the random time a loan is originated and the securitized trust begins making monthly payments. That is, we observe $X$ if and only if $X \geq Y$. We further assume $X$ and $Y$ are independent, an important assumption we now briefly justify within a securitization context. The random variable $Y$ represents the time an ABS first starts making payments. Typically, the decision to issue a securitization is more related to investment market conditions and the financing needs of the parent company than the performance of the underlying assets, in this case automobile loans. In other words, the forming and subsequent issuance of an ABS bond has little to do with the time-to-event distribution of each individual loan, which is represented by $X$. Hence, the assumption that $X$ and $Y$ are independent is actually quite reasonable within the context of the securitization process. To account for right-censoring, define the censoring random variable as $C = Y + \tau$, where $\tau$ is a constant that depends on the last month the securitization is active and making monthly payments. Note that independence between $X$ and $C$ follows trivially from the assumed independence of $X$ and $Y$. We thus observe the exact loan termination time, $x$, if $x \leq C \mid X \geq Y$, and we only know that $x > C$ if $x > C \mid X \geq Y$.

For those familiar with incomplete data from observational studies, we can think of the period of time the ABS is active and paying as the observation window. Hence, random left-truncation occurs because we only observe loans that survive long enough to enter into the trust, and right-censoring occurs because we only observe the exact termination time of loans that end prior to end of the securitization. For completeness, we will assume discrete-time because a borrower’s monthly obligation is considered satisfied as long as the payment
is received before the due date. Therefore, we may assume the recoverable distribution of $X$ is integer-valued with a minimal time denoted by $\Delta + 1$ for nonrandom $\Delta \in \{\mathbb{N} \cup 0\}$, where $\mathbb{N}$ denotes the natural numbers, and a finite maximum end point, which we denote by $\xi \geq \Delta + \tau$, for nonrandom $\xi \in \mathbb{N}$. We emphasize the word *recoverable*, further discussion of which may be found in Lautier et al. (2021) and Lautier et al. (2023).

We now generalize Lautier et al. (2023) to the case of two competing risks as follows.\(^1\) First, consider two competing risks as a multistate process, such as in Section 3 of Beyersmann et al. (2009). Formally, let $\{Z_x\}_{\Delta+1 \leq x \leq \xi}$ denote a set of random variables with probability distributions that depend on $x$, $\Delta + 1 \leq x \leq \xi$. More specifically, given a loan terminates at time $x$, we assume the loan must be in one of two states, $Z_x \in \{1, 2\}\(^2\):

1. This is the *event of interest*. Loans move into this state if a default occurs. The probability of moving into state 1 at time $x$ is the cause-specific hazard rate for state 1, denoted $\lambda_0^1(x)$.

2. This is the *competing event*. Loans move into this state if a prepayment occurs. The probability of moving into state 2 at time $x$ is the cause-specific hazard rate for state 2, denoted $\lambda_0^2(x)$.

The discrete-time cause-specific hazard rate is then defined as

$$
\lambda_0^i(x) = \Pr(X = x, Z_x = i \mid X \geq x) = \frac{\Pr(X = x, Z_x = i)}{\Pr(X \geq x)}, \quad i = 1, 2.
$$

Conveniently, therefore, from the law of total probability, we have

$$
\lambda(x) = \frac{\Pr(X = x)}{\Pr(X \geq x)} = \frac{\Pr(X = x, Z_x = 1)}{\Pr(X \geq x)} + \frac{\Pr(X = x, Z_x = 2)}{\Pr(X \geq x)} = \lambda_0^1(x) + \lambda_0^2(x).
$$

Within a competing risk framework, $\lambda(x)$ may be referred to as the *all-cause hazard*\(^2\).

---

\(^1\)Indeed, these forthcoming results are a contribution to the statistical literature in their own right. For statistically inclined readers, Appendix A provides formal statements, and the Online Appendix A provides complete proofs.

\(^2\)It may be of help to see the related Beyersmann et al. (2009, Figure 1).

\(^3\)It may be illuminating to review Table 4 in the simulation study of the Online Appendix C for a numeric example of our competing risk model. To make the economic connection between loan default risk over time and the cause-specific hazard for default (cause 01), we will elucidate the probability that the cause-specific
Given this framework, it is not difficult to account for securitization data subject to right-censoring and left-truncation along the lines of Lautier et al. (2023). Formally, assume a trust consists of \( n > 1 \) consumer automobile loans. For \( 1 \leq j \leq n \), let \( Y_j \) denote the truncation time, \( X_j \) denote the loan ending time, and \( C_j = Y_j + \tau_j \) denote the loan censoring time. Because of the competing events, we also have the event-type random variable \( Z_{X_j} = i \), where we observe \( Z_{X_j} \) given \( X_j \) for \( i = 1, 2 \). In what follows, we will use a subscript of \( \tau \) where appropriate to remind us that right-censoring is present in the data.

If we assume independence between \( Y \) and the random vector \((X, Z_X)\) (not at all unreasonable given the securitization backdrop), then we may derive estimators for (3) along the same lines as Lautier et al. (2023). We demonstrate as follows. Let \( \alpha = \Pr(Y \leq X) \) and for \( i = 1, 2 \), define

\[
\begin{align*}
    f_{x,\tau}^0(x) &= \Pr(X_j = x, X_j \leq C_j, Z_{X_j} = i) \\
    &= \Pr(X = x, X \leq C, Z_x = i \mid X \geq Y) \\
    &= \Pr(X = x, x \leq C, Z_x = i, x \geq Y)/\Pr(X \geq Y) \\
    &= (1/\alpha)\{\Pr(X = x, Z_x = i)\Pr(Y \leq x \leq C)\},
\end{align*}
\]

and

\[
U_\tau(x) = \Pr(Y_j \leq x \leq \min(X_j, C_j)) = \Pr(Y \leq x \leq \min(X, C) \mid X \geq Y) = (1/\alpha)\{\Pr(Y \leq x \leq C)\Pr(X \geq x)\}.
\]

hazard rate represents. Suppose the current age of a loan is \( x \) months, where \( \Delta + 1 \leq x \leq \xi \). Then the quantity \( \lambda^0(x) \) denotes the probability that a loan will end in default in month \( x \), given it has survived at least \( x \) months. Therefore, if the default risk of a borrower changes over time, a plot of the month-by-month hazard rate for a given risk band will provide a current risk estimate for the borrowers within that risk band who have continued to make ongoing payments (i.e., “survived”). If the hazard rate remains constant, then the monthly default risk does not change as a loan matures and continues to remain actively paying. On the other hand, if the hazard rate declines (increases), this would suggest that the current month default risk declines (increases) as a loan matures.

\footnote{We emphasize that the observable data from a trust, \( \{X_j, Y_j, C_j, Z_{X_j}\}_{1 \leq j \leq n} \) differs from the random variables, \( \{X, Y, C, Z_X\} \). For example, the random variables \( X \) and \( (Y, C) \) are independent, whereas \( X_j \) and \( (Y_j, C_j) \) clearly are not. The reference Lautier et al. (2021) expounds on this point thoroughly.}
Thus,
\[
\lambda^0_i(x) = \frac{\Pr(X = x, Z_x = i)}{\Pr(X \geq x)} = \frac{f^0_i(x)}{U_{\tau}(x)}. \tag{4}
\]

In terms of our observable data, for a given loan \(j\), \(1 \leq j \leq n\), we observe \(Y_j, \min(X_j, C_j)\), and \(1_{X_j \leq C_j}\), where \(1_Q = 1\) if the statement \(Q\) is true and 0 otherwise. Further, if we observe an event for loan \(j\), we will also observe the information \(Z_{X_j} = i, i = 1, 2\). Therefore, using the standard estimators vis-à-vis the observed frequencies
\[
\hat{f}^{0_i}_{*\tau,n}(x) = \frac{1}{n} \sum_{j=1}^{n} 1_{X_j \leq C_j} 1_{Z_{X_j} = i} 1_{\min(X_j, C_j) = x},
\]

and
\[
\hat{U}_{\tau,n}(x) = \frac{1}{n} \sum_{j=1}^{n} 1_{Y_j \leq x \leq \min(X_j, C_j)},
\]

we obtain the estimate for (4)
\[
\hat{\lambda}^0_{\tau,n}(x) = \frac{\hat{f}^{0_i}_{*\tau,n}(x)}{\hat{U}_{\tau,n}(x)} = \frac{\sum_{j=1}^{n} 1_{X_j \leq C_j} 1_{Z_{X_j} = i} 1_{\min(X_j, C_j) = x}}{\sum_{j=1}^{n} 1_{Y_j \leq x \leq \min(X_j, C_j)}}. \tag{5}
\]

Pleasingly, (5) is equivalent to the related classical work of Huang and Wang (1995), despite our assumption of discrete-time at the problem’s onset.

The estimator (5) is itself a random variable, and it has a number of pleasing asymptotic properties. First, the complete vector of estimators over the recoverable space of \(X\), \(\hat{\Lambda}^0_{\tau,n} = (\hat{\lambda}^0_{\tau,n}(\Delta + 1), \ldots, \hat{\lambda}^0_{\tau,n}(\xi))^T\), is asymptotically unbiased for the true hazard rates. Further, \(\hat{\Lambda}^0_{\tau,n}\) is multivariate normal with a completely specifiable diagonal covariance structure (i.e., two estimators within \(\hat{\Lambda}^0_{\tau,n}\) are asymptotically independent). The formal statement of these properties may be found in Proposition 1 in Appendix A. Additionally, we may use Proposition 1 to produce asymptotic confidence intervals that are appropriately bounded within \((0, 1)\). These asymptotic confidence intervals are available in Lemma 1 in Appendix A.

Finally, the asymptotic confidence intervals of Lemma 1 may be used in a straightforward large sample statistical hypothesis test. Formally, for two risk bands \(a, a’\), where \(a \neq a’\) (i.e., \(a, a’\) would represent one of the risk bands deep subprime, subprime, near-prime, prime, or
super-prime), we may test

\[ H_0 : \lambda_{r,(a)}^{0_i}(x) = \lambda_{r,(a')}^{0_i}(x), \quad \text{v.s.} \quad H_1 : \lambda_{r,(a)}^{0_i}(x) \neq \lambda_{r,(a')}^{0_i}(x), \]

for each age \( x \) by determining if the \((1-\theta)\)% asymptotic confidence intervals of the estimators \( \hat{\lambda}_{r,n,(a)}^{01}(x) \) and \( \hat{\lambda}_{r,n,(a')}^{01}(x) \) overlap for \( \Delta + 1 \leq x \leq \xi \) and \( i = 1, 2 \). The decision rules and interpretations are as follows. Fix \( x \in \{\Delta + 1, \ldots, \xi\} \) and \( i = 1 \). If the asymptotic confidence intervals of \( \hat{\lambda}_{r,n,(a)}^{01}(x) \) and \( \hat{\lambda}_{r,n,(a')}^{01}(x) \) overlap, we fail to reject \( H_0 \), and we cannot claim \( \lambda_{r,(a)}^{01}(x) \neq \lambda_{r,(a')}^{01}(x) \). That is, conditional default risk given survival to time \( x \) has converged. On the other hand, if the asymptotic confidence intervals do not overlap, we reject \( H_0 \), and we may claim with \((1-\theta)\)% confidence that \( \lambda_{r,(a)}^{01}(x) \neq \lambda_{r,(a')}^{01}(x) \). That is, conditional default risk given survival to time \( x \) has not converged. In honor of the age-old cliché about the efficiency of examining a picture versus words, we refer to the reader to Figure 3.

II.B. Empirical Evidence

We now apply the statistical methods of Section II.A to the consumer loan data of Section I. Specifically, we both plot estimates of the cause-specific hazard rates for default by loan age and risk band, \( \hat{\lambda}_{r,n}^{01}(a) \), and perform the hypothesis test described by (6) to the filtered loan population summarized in Figure 2 and Table 1. For convenience of exposition, we will initially focus our discussion on two risk bands: subprime and prime borrowers. A plot of \( \hat{\lambda}_{r,n}^{01} \) by loan age may be found in Figure 3 for the 21,332 subprime loans (solid line) and 6,300 prime loans (dashed line).

As an initial observation, we can see that the estimated CSH rates for subprime loans are initially higher than the CSH rates for prime loans. This is expected given our expectations about credit risk, risk-based pricing, and the difference in APRs between the two risk bands (i.e., the subprime risk band has APRs between 15-20% and the prime risk band has APRs between 5-10%). This pattern does not maintain for the full lifetime of the loan, however. As the subprime loans continue to stay current (i.e., given survival), the hazard rate declines. Interestingly, the hazard rates for prime loans in this sample appear to increase slightly, though they remain generally stable even as loan age increases. We remark here that, due to left-truncation and right-censoring, we have recoverable estimates of the CSH rate and confidence intervals spanning approximately month 4 through month 54 and not the complete
Figure 3: **Credit Risk Convergence: Subprime and Prime Loans.** A plot of $\hat{\lambda}_{01,\tau,n}^1$ (defaults) defined in (5) by loan age for the subprime and prime risk bands within the sample of 58,188 loans (Table 1), plus 95% confidence intervals using Lemma 1. We may use the hypothesis test described in (6) by searching for the minimum age that the confidence intervals overlap between two disparate risk bands. In this case, we see the first evidence of credit risk convergence by approximately loan age 42 months and then consistently so by loan age 50 months on 72-73 month auto loans. The large upward spike in $\hat{\lambda}_{01,\tau,n}^1$ for the subprime risk band around loan age 40 is related to the economic impact of COVID-19, a point discussed more fully in Section II.C. Due to left-truncation and right-censoring, the recoverable range is approximately $4 \leq X \leq 54$ for these 72-73 month automobile loans.

This brings us to the major result of this paper, which is the lower right corner of Figure 3. In addition to plotting the point estimates, we also provide the asymptotic confidence intervals (shaded regions surrounding each line). Eventually, as the two lines slowly approach each other, the confidence intervals begin to overlap. The first evidence of this is around loan age 42, and it is consistent by approximately loan age 50 for these 72-73 month consumer auto loans. With the statistical test outlined in (6), therefore, for any age in which
we observe overlapping confidence intervals, we cannot claim the true CSH rates for default are different between the subprime and prime risk bands within this sample. It is this phenomenon in which two CSH rates for default between two different risk bands eventually become consistently statistically equivalent that we call credit risk convergence.\textsuperscript{22}

Table 2 (top) provides a summary matrix of the estimated month of credit risk convergence among the five risk bands considered for the sample of 58,118 filtering loans from the 2017 issuance. For conservatism, we defined the point of credit risk convergence as the first of three consecutive months of confidence interval overlap after a loan age of 10 months. Based on these results, we would say that a deep subprime loan eventually converges in risk to a subprime loan after three years, and it converges to a prime risk after 50 months and a super-prime risk after 52 months. Similarly, subprime loans converge in risk to prime loans after 42 months, and they become super-prime risks after four years. Near-prime loans become prime risks quite quickly, shortly after one year, and then become super-prime risks after 43 months. For completeness, we plot the full five-by-five matrix of CSH and confidence interval comparisons along the lines of Figure 3 in Figure 11 found in Appendix B.

We also see a large spike in the CSH rate for the subprime risk band by approximately loan age 40. Similarly, it appears the prime risk band also has a small increase in its CSH rate shortly after the same age. With some approximate date arithmetic from the first payment month of the ABS bonds (March-April-May 2017), we find that a loan age of 40 months corresponds to approximately Spring 2020 (when adjusted for left-truncation). If we recall the economic impact of the Coronavirus, which effectively stopped most economic activity in Spring 2020, it is not difficult to understand why so many loans defaulted around loan age 40. This also provides informal validation that the data sorting and estimation of the default CSH rate has been effective. It is interesting to compare the difference in impact to subprime and prime borrowers. That is, the economic shutdown brought on by the Coronavirus pandemic appears to have had a smaller impact on the prime risk band than the subprime risk band. In Section II.C, we provide much more discussion.

\textsuperscript{22}We also remark that measuring default risk conditional on survival gleans additional insight in comparison to a binary default analysis, such as that performed in Table 1. Indeed, 40\% of all subprime loans in the sample of 58,118 defaulted at some point, versus only 10\% of prime loans. Given just this analysis, it is not surprising the subprime borrowers received a higher APR than the prime borrowers. What we show in Figure 3 is that the default rates conditional on survival are not constant, however, and it implies that subprime borrowers that do not refinance are eventually overpaying based on an updated assessment of their risk profile. We come back to this point much more extensively in Section III.B.
Table 2: **Credit Risk Convergence: Matrix Summary.** A summary matrix of the estimated month of credit risk convergence for 72-73 month consumer automobile loans. The top matrix corresponds to the sample of 58,118 loans issued in 2017 (Table 1). The bottom matrix corresponds to the sample of 65,892 loans issued in 2019 (see Section II.C). For conservatism, the month of credit risk convergence is defined as the first of three consecutive months after ten months that the asymptotic confidence intervals for \( \hat{\lambda}_{\tau,n}^{01} \) overlap. Visually, it is helpful to compare Figure 3 with the top matrix (subprime, prime) and Figure 4 with the bottom matrix (subprime, prime). Full comparisons may be made with Figure 11 in Appendix B and Figure 12 in Appendix C. For the top matrix, the recoverable range is approximately \( 4 \leq X \leq 54 \). For the bottom matrix, the recoverable range is \( 4, 10 \leq X \leq 30, 35, 38 \leq X \leq 43 \). An entry of NA implies that convergence has not yet occurred by month 43.

|             | deep subprime | subprime | near-prime | prime | super-prime |
|-------------|---------------|----------|------------|-------|-------------|
| deep subprime | 10            | 36       | 50         | 50    | 52          |
| subprime    | 10            | 41       | 42         | 48    |             |
| near-prime  | 10            | 13       | 43         |       |             |
| prime       | 10            | 10       | 10         |       |             |
| super-prime | 10            |          |            |       |             |

|             | deep subprime | subprime | near-prime | prime | super-prime |
|-------------|---------------|----------|------------|-------|-------------|
| deep subprime | 10            | 39       | NA         | NA    | NA          |
| subprime    | 10            | 23       | 24         | 15    | NA          |
| near-prime  | 10            | 15       | 15         |       |             |
| prime       | 10            | 10       | 10         |       |             |
| super-prime | 10            |          |            |       |             |

**II.C. A Digression on COVID-19**

As alluded to in Section II.B, we have attributed the large increase around loan age 40 for the default CSH rate estimate observable in Figure 3 to the Spring 2020 economic shutdown resulting from the initial rapid spread of the Coronavirus disease. Because the point of credit risk convergence occurs after month 40 for some pairs of risk bands in Table 2 (e.g., deep subprime and prime credit risk convergence occurs by loan age 50), there is a concern that the observable phenomenon of default risk converging for disparate risk bands is due to
the filtering effect of the shock of the economic shutdown rather than due to some inherent property of loan risk behavior. In other words, only the strongest credits could survive such a shock, and credit risk convergence would not occur otherwise. While we feel the economic shutdown has played some role, we believe it is not adequate on its own to explain the credit risk convergence we observed in our sample. We argue as follows.

First, if we return again to Table 2, we can see that pairs of risk bands converge earlier than loan age 40 (e.g., deep subprime and subprime, near-prime and prime, and prime and super prime). Thus, we have examples of risk bands that converge in conditional monthly default risk prior to the onset of the Spring 2020 economic shutdown.

Second, if the phenomenon of credit risk convergence is completely driven by the Spring 2020 economic shutdown, we would expect to see it occur much earlier in a sample of bonds issued closer to Spring 2020 when subject to the same loan selection process and risk band definitions of Section I.A. Hence, we obtained loan level data from the same four consumer auto loan ABS issuers but from bonds issued closer to Spring 2020: SDART 2019-3 (Santander, 2019b), DRIVE 2019-4 (Santander, 2019a), CARMX 2019-4 (CarMax, 2019), and AART 2019-3 (Ally, 2019). These bonds began paying in late Summer 2019, whereas the bonds introduced in Section I began paying in Spring 2017.

Figure 4 is a repeat of Figure 3; it presents the estimated CSH rates for default plus asymptotic 95% confidence intervals for the 2019 sample. As expected, we see the large spike in the CSH rate for defaults in supprime loans around 10 months, which, when adjusted for left-truncation, corresponds to the Spring 2020 economic shutdown. We also display the estimated credit risk convergence matrix for the 2019 issuance in the bottom portion of Table 2. In reviewing the matrix, we see evidence of earlier convergence, and so the shock of the economic shutdown of Spring 2020 has played some role. It is not the whole story, however. For example, the subprime risk band in the 2019 issuance has not yet converged with the super-prime risk by loan age 43, which is the last recoverable age in the 2019 sample as of February 2023. In the 2017 issuance, the subprime risk band converged with the super-prime risk band by loan age 52. This suggests that loan age also plays a role.

23The filtered 2019 sample mirrors the distribution of the 2017 filtered sample summarized in Table 1. For example, there are 31,283 DRIVE 2019-4 loans, 19,989 SDART 2019-3 loans, 11,724 CARMX 2019-4 loans, and 2,896 AART 2019-3 loans, for a total of 65,892. By risk band, there are 24,140 (37%) deep subprime loans, 20,921 (32%) subprime loans, 9,940 (15%) near-prime loans, 8,625 (13%) prime loans, and 2,266 (3%) super prime loans.
Similarly, while convergence between risk bands occurs earlier for the 2019 sample, it takes more months after the shutdown shock for most disparate risk bands to converge than after the same shock in the 2017 sample. For example, the deep subprime and prime risk bands converge by loan 25 in the 2019 sample, which is 15 months after the economic shutdown shock. For the 2019 sample, however, the deep subprime risk band converges with the prime risk band by loan age 50, which is only 10 months after the economic shutdown. This again suggests that loan age, in addition to the economic shutdown of Spring 2020, plays a role.

Finally, we also remark that in the last twenty years it is difficult to find a span of 72 consecutive months in which there was not a large scale economic shock (e.g., September 11, 2001; 2007-2009 global financial crisis; 2009-2014 European sovereign debt crisis, COVID-19, etc.). Hence, in an economic environment in which one-in-a-hundred year events occur every decade or so, credit risk convergence may be perpetually present, even if it may be partially explained by the filtering effects of an economic crisis.

II.D. Additional Sensitivity Analysis

We consider two additional items of a robustness analysis in this section. The first is a third iteration of the statistical analysis of both Section II.B and Section II.C. We will sort the data for new cars at the point of sale. This will give us exposure to a potentially different borrower profile and collateral value pattern. It will also greatly reduce our exposure to the CARMX bond. Reduced exposure to CARMX is valuable because the parent company, CarMax, has an entirely different business model and therefore financing incentive than either Santander or Ally, the origination banks of the DRIVE, SDART, and AART ABS bonds. Hence, it is also possible that CARMX loans behave differently than loans originated by banks. The second item is a discussion on the potential to generalize the idea of credit risk convergence to other types of consumer loans or even fixed-income assets more generally. We discuss each in turn.

We again return to the original collective pool of over 275,000 consumer auto loans of the 2017 issuance of the four bonds introduced in Section I: CARMX, AART, DRIVE, and SDART. We then perform an identical risk band APR-based sorting and loan filtering of Section I.A, except rather than used cars we restrict our sample to new cars. This leaves a total sample of 16,412 loans, with bond exposures of DRIVE (7,692), SDART (7,369), ALLY
Figure 4: Credit Risk Convergence: Subprime and Prime Loans, COVID Sensitivity. A plot of $\hat{\lambda}_{r,n}^{01}$ (defaults) defined in (5) by loan age for the subprime and prime risk bands within the sample of 65,892 loans issued in 2019, plus 95% confidence intervals using Lemma 1. We may use the hypothesis test described in (6) by searching for the minimum age that the confidence intervals overlap between two disparate risk bands. Because the 2019 bonds were issued closer to Spring 2020, the large upward spike in $\hat{\lambda}_{r,n}^{01}$ occurs much earlier for the subprime risk band, closer to loan age 10 (compare with Figure 3). We see some evidence of earlier credit risk convergence around loan age 25 in comparison to Figure 3, but it does not appear to be consistent until after loan age 40. This suggests that the economic impact of COVID-19 has played some role in credit risk convergence, but loan age also plays a role. Due to left-truncation and right-censoring, the recoverable range is $4, 10 \leq X \leq 30, 35 \leq X \leq 43$ for these 72-73 month automobile loans. (1,342) and CMAX (9). As expected, restricting the sample to new cars has eliminated almost all loans from CMAX, whose parent company, CarMax, specializes in used auto sales. Thus, the current sample of 16,412 loans consists of loans originated by traditional banks, Santander and Ally. In terms of risk band, we maintain dispersed exposure with deep subprime (3,892), subprime (8,242), near-prime (2,132), prime (1,407), and super prime (739). Finally, all loans consist of a new vehicle at the point of sale, and so we are now
Figure 5: Credit Risk Convergence: Subprime and Prime Loans, Collateral Sensitivity. A plot of $\hat{\lambda}^{01}_{r,n}$ (defaults) defined in (5) by loan age for the subprime and prime risk bands within the sample of 16,412 loans issued in 2017 with new cars at the point of sale, plus 95% confidence intervals using Lemma 1. We may use the hypothesis test described in (6) by searching for the minimum age that the confidence intervals overlap between two disparate risk bands. Because of the smaller sample, the asymptotic confidence interval for the CSH rate of prime loans is wider. The overall pattern is very similar to Figure 3, however, and so the concept of credit risk convergence appears robust to collateral type at the point of sale (i.e., new or used). The sample of 16,412 new car loans also has minimal exposure to CARMX, and so credit risk convergence also appears to be robust to the potentially different business incentives of the loan originator. Due to left-truncation and right-censoring, the recoverable range is approximately $4 \leq X \leq 54$ for these 72-73 month automobile loans.

considering an entirely different collateral depreciation pattern and even potentially borrower profile. We present an update of both Figure 3 and Figure 4 in Figure 5.

Immediately, we see that the overall pattern of Figure 5 closely mirrors that of Figure 3. The subprime loans have a CSH rate estimate that is consistently higher than prime loans in the early months of a loan’s age. We also see the large increase in the CSH rate for subprime
loans around loan age 40, which correspond to the timing of the economic shutdown due to COVID-19 in Spring of 2020. As with the used cars-at-the-point-of-sale loans, there appears to be minimal impact from COVID-19 for prime loans. Importantly, the two CSH rates for the subprime and prime risk bands eventually converge, however, which we see at the lower right corner of Figure 5. The asymptotic confidence intervals begin to consistently overlap beginning shortly after loan age 40, which corresponds to row two, column four of the top matrix of Table 2. This suggests that the concept of credit risk convergence appears to be robust in consumer auto loans to the collateral type at the point of sale (i.e., new or used). Because the sample of 16,412 new car loans has such minimal exposure to CARMX, we also see that the concept of credit risk convergence appears to be robust to potentially different business incentives of the parent company to the loan originator (i.e., used car sales versus traditional banking).

As a final note on collateral type, a close inspection of Figure 5 in comparison to Figure 3 reveals wider asymptotic confidence intervals for the CSH rate for default in prime loans. This is driven by the smaller sample size, and it is exacerbated for super prime loans written on new cars (i.e., there are very few defaults for super prime loans written on new cars in our sample of 739). Hence, we have avoided reporting the credit risk convergence matrix of Table 2 for the sample of 16,412 new car loans to avoid potentially erroneously conclusions due to faulty asymptotic statistics stemming from a small default sample. Instead, we report the point CSH rate estimates for default for all five risk bands in Figure 6. In this case, a simple line plot speaks volumes. In the young ages of a loan, we see that the CSH rates for default is the highest for deep subprime loans, and it progresses sequentially downward by risk band until super prime loans, of which there are very few defaults. This pattern is expected. As the loans age, however, we see all CSH rates for default for each risk band converge together in the bottom right of Figure 6 near loan age 50. Even without the confidence intervals, therefore, we find additional evidence of credit risk convergence.

We close this section with a brief discussion on attempting to generalize the concept of credit risk convergence to debt instruments beyond secured consumer auto loans. We consider first other types of consumer loans. It is a reasonable concern that credit risk convergence may not occur in other types of consumer loans, particularly those not secured by collateral. Because a borrower faces repossession in the event of default, there is likely a continuously strong incentive to stay current. Further, the borrower is slowly building an
equity position as the loan matures, and so the declining default risk for riskier loans as the loans mature may also reflect a borrower’s gradually increasing cost of default or potentially increasing desire to own the vehicle outright. While this bodes well for observing credit risk convergence in residential mortgage loans, it does raise the question if it will still occur for unsecured forms of consumer lending, such as credit cards, peer-to-peer, or unsecured consumer loans. We also return to the high priority of payment of consumer automobile loans, which may not extend to income/vacation properties and even residential mortgages more generally. Thus, while we are optimistic credit risk convergence will extend to other
types of consumer loans, we believe more study is needed. Additionally, we postulate this “survivor bias” will extend to many other areas of fixed-income (e.g., high-yield or “junk bonds” in the corporate bond space), but we again believe more research is needed. Thus, we suggest cautious optimism before attempting to generalize these results outside of consumer auto loans, and we leave this open as an area of further study.

III. Financial Implications

We now build on the theoretical results of Section II to offer financial perspectives. The present section proceeds in two parts. In Section III.A, we demonstrate how the probability estimates for default may be used to visualize the back-loading of lender’s profits. Related details for estimating a recovery upon default assumption and extensions to the profitability calculations may be found in Appendices D and E. In Section III.B, we then focus our analysis on the individual consumer. We find that borrowers in all non-super prime risk bands delay prepayment inefficiently. In a surprise, we find that borrowers in lower risk bands, near-prime and prime, operate less efficiently than borrowers in higher risk bands, deep subprime and prime. Details may be found in Table 3. We also evaluate borrower prepayment behavior using the sibling estimator (5) for prepayment. In a visual analysis, we find that borrower’s prepayment decisions appear to be driven by economic stimulus payments and unusual used auto markets rather than by financial sophistication. The section closes with brief thoughts on why the market for mature consumer auto loans appears to be operating at a sub-optimal level of efficiency.

III.A. Lender Profitability Analysis

Conventional profitability wisdom of risk-based pricing from the perspective of a lender is that the high-returns of high-risk loans that don’t default help offset the losses from the high-risk loans that do default. In other words, there is an implied insurance arrangement in which the cost of the losses are dispersed among the individual borrowers. Furthermore, it can be argued that through its precision, risk-based pricing has been attributed to lowering the cost of credit for a majority of borrowers and expanding credit availability to higher risk
borrowers (Staten, 2015). These are positive economic outcomes, and we do not attempt to argue against the overall practice of risk-based pricing or imply lenders of such loans are operating unscrupulously. Indeed, all loans considered within our analysis have been sampled from pools of securitized bonds. That is, the risk of default has already been transferred off the lender’s balance sheet. We will argue that the consumer auto loan market is capable of operating more efficiently, however. As one component of this argument, it is illuminating to perform a profitability analysis, especially in light of the default and prepayment probabilistic estimates we obtained in Section II.B. This is the purpose of the present section.

A common term to describe the profit of a high-risk, high-interest-rate loan that remains current is back-loaded. Quite simply, a high-risk, high-interest-rate loan gradually becomes more profitable as it continues paying, and it is these increased profits later in the loan’s life that offset the losses taken on other similar loans that have defaulted. To provide some formality to this idea, we will utilize an actuarial approach to calculate an implied risk-adjusted return for each month a loan stays current. Specifically, we will examine a rolling monthly expected annualized rate of return assuming an investor purchases a risky fixed-income asset at a price of the outstanding balance of the consumer loan at age \( x \) for risk band \( a \), \( B_{a|x} \), with a one-month term. This hypothetical risky asset pays either (1) the outstanding balance at loan age \( x + 1 \) for risk band \( a \), \( B_{a|x+1} \), plus the next month’s payment due, \( P_a \), with probability \( 1 - \lambda_{r(a)}^0(x) \) or (2) the recovery amount at time \( x + 1 \) in the event of default, \( R_{x+1} \), with probability \( \lambda_{r(a)}^0(x) \). The subscript \( a \) denotes one of the five standard risk bands: deep subprime, subprime, near-prime, prime, and super prime. We illustrate this hypothetical asset in Figure 7.

To calculate the annualized risk-adjusted return by month, we first define the expected present value (EPV) of a \( B_{a|x} \) risky one-month loan depicted in Figure 7 as

\[
EPV_{a|x}^1 = \lambda_{r(a)}^0(x) \left[ \frac{R_{x+1}}{1 + \tilde{r}_{a|x}} \right] + (1 - \lambda_{r(a)}^0(x)) \left[ \frac{B_{a|x+1} + P_a}{1 + \tilde{r}_{a|x}} \right],
\]

(7)

where \( \tilde{r}_{a|x} \) is some unknown one-month effective rate of interest. To calculate the annualized risk-adjusted return, we can interpret the outstanding balance of an age \( x \) loan in risk band \( a \),

\(^{24}\)For readers seeking a more thorough introduction to risk-based pricing, Livshits (2015) presents a useful survey.

\(^{25}\)We thank Jonathan A. Parker for this concise descriptive term.
Figure 7: Hypothetical Risky Fixed-Income Asset and Path Probabilities. This hypothetical risky asset pays either (1) the outstanding balance at loan age $x + 1$ for risk band $a$, $B_{a|x+1}$, plus the next month’s payment due, $P_a$, with probability $1 - \lambda_{\tau,(a)}(x)$ or (2) the recovery amount at time $x + 1$ in the event of default, $R_{x+1}$, with probability $\lambda_{\tau,(a)}(x)$. The subscript $a$ denotes one of the five standard risk bands: deep subprime, subprime, near-prime, prime, or super prime.

$B_{a|x}$, as the market-implied price of a risky zero coupon bond following the payment pattern of Figure 7. Therefore, we can use (7) to solve for $\tilde{r}_{a|x}$ such that $\text{EPV}^1_{a|x} = B_{a|x}$. This rate, $\tilde{r}_{a|x}$, is then the monthly effective risk-adjusted return, which can then be annualized.\textsuperscript{26} The calculation in (7) also requires an estimate for the recovery upon default, $R_{x+1}$, for each age $x$. We performed this estimate separately for each filtered sample of loans: the 58,118 loans from the four ABS (CARMX, ALLY, SDART, DRIVE) issued in 2017 and

\textsuperscript{26}We remark that implicit in this analysis is the assumption that the remaining payments beyond month $x + 1$ are a tradable asset with no friction (i.e., the risky asset may be traded at time $x + 1$ for $B_{a|x+1}$). We can instead perform a risk-adjusted return calculation over the entire remaining lifetime of the loan (i.e., assuming uncertainty for each future payment following the estimates in Section II.B). For interested readers, the details of how to perform this full calculation may be found in Appendix E.
summarized in Section I.B and the 65,892 loans from the same four ABS bonds issued in 2019 and summarized in Section II.C. The complete details, including a depicted recovery curve, may be found in Appendix D and Figure 13, respectively. The probabilities, $\lambda_{r,(a)}^0(x)$, for each age, $x$, and risk band, $a$, may be estimated using the methods of Section II.

For ease of interpretation, we elected to consider a single loan of $100 for 72 months with a payment and amortization schedule determined by the average APR of each risk band: deep subprime (22.7%), subprime (18.0%), near-prime (12.8%), prime (7.79%), and super prime (3.63%). The estimated results for both the 2017 and 2019 issuance may be found in Figure 8. There are some interesting observations. For the 2017 issuance (top), we see that the deep subprime, subprime, near-prime, and prime risk bands generally group together around 7.5% during the earlier part of the loan’s lifetime. This demonstrates that the risk-adjusted pricing is generally accurate by risk band, as the higher APRs help offset the higher default risk. The super prime risk-band consistently hovers around a 2.5% annualized risk-adjusted return. We then see the negative impact of COVID-19 around loan age 40, which is consistent with the discussion in Sections II.B and II.C. It is notable that the impact on the risk-adjusted return for the super prime risk band due to COVID-19 is minimal. As the loans mature, however, and credit risk convergence begins, we see the risk-adjusted returns for the higher APR loans begin to accelerate. This is a visualization of back-loaded profits. For the 2019 issuance (bottom), there is a similar clustering in the early months of a loan’s lifetime. The impact of COVID-19 is much sooner, however, and we also estimated earlier credit risk convergence between risk bands (see Table 2). Thus, the risk-adjusted returns by risk band separate earlier.

III.B. Consumer Perspectives

If a borrower’s default risk conditional on survival declines as a loan stays current, but the loan’s original APR is a single point-in-time estimate of risk, then it is possible a gradual economic inefficiency from the perspective of the consumer may development. The purpose of the present section is an attempt to quantify this inefficiency and offer potential explanations for its appearance. The first part will be dedicated to estimating a dollar amount, and the

\footnote{These averages are for the 2017 sample of 58,118 loans. The averages for the 2019 sample of 65,892 loans are similar: deep subprime (22.7%), subprime (17.7%), near-prime (12.6%), prime (8.16%), and super prime (4.26%).}
second part will offer observations on consumer behavior and larger market behavior.

As an initial starting point, it may be tempting to trivialize the true dollar impact to consumers. We estimate credit risk convergence does not begin to appear until as late as 52 months into a 72-73 month auto loan in some cases (see Table 2), which suggests that any outstanding loan balances will be minimal. An examination of Figure 9 reveals this is not the case, however. For the sample of 51,118 filtered loans from the 2017 issuance, we
find average median balances still well over $5,000 for loans as old as 60 months. For a deep subprime consumer with an average APR north of 20% and an estimated annual income of just over $41,000 (median $34,638), this is not an insignificant loan. Hence, if a current borrower survives long enough to converge into a lower risk band and is able to refinance at the lower rate, the potential savings could be meaningful.

We estimate such potential savings in Table 3. Moving left-to-right along the column headings, we first report a count of the current loans by loan age. Next, of the active loans, we present an average outstanding balance, average payment, and average APR. The “Pmts (#)” column calculates the remaining payments needed to pay-off the average loan balance given the average payment. The next four columns represent the potential savings in monthly payment if a borrower refines at the average APR of the lower risk band, after the point of credit risk convergence. If two risk bands have not yet converged in credit risk, the numbers are not provided in the table. The convergence points may be found in Table 2. The calculations do not assume any upfront refinancing charge. There are some interesting observations.

First, we find that borrowers in all four risk bands, deep subprime, subprime, prime, and super prime, leave money on the table. On a monthly payment basis, deep subprime borrowers begin to overpay between $11-62 per month around loan age 36, for a total potential savings between $136-1,212. Based on our estimates, deep subprime borrowers would benefit the most in terms of total savings by waiting until approximately loan age 50, when they converge in risk with prime borrowers. In terms of monthly payment savings, deep subprime borrowers should wait to refinance until they converge in credit risk with super prime borrowers. Encouragingly, we see that most deep subprime borrowers have prepaid or refinanced by loan age 60, which suggests some self-correction, albeit slower than our calculations would recommend. Shortly, we will opine on these borrower’s motivations to refinance, and we are uncertain if it is motivated by an updated personal risk assessment. The situation for subprime borrowers is similar; they benefit the most in total savings by

\[28\text{We present the (average; median) annual income by risk band for the filtered 51,118 loans issued in 2017: deep subprime ($41,093.18; $34,638.03), subprime ($46,063.52; $38,688.64), near-prime ($56,745.51; $44,152.97), prime ($65,559.14; $49,287.19), and super prime ($98,161.61; $73,900.87). If we perform the same calculations only for loans that made 40 or more payments, the figures are similar: deep subprime ($39,448.72; $34,090.03), subprime ($44,635.46; $37,457.02), near-prime ($55,807.83; $42,769.99), prime ($64,318.90; $48,760.87), and super prime ($97,386.01; $73,928.23).}\]
Figure 9: **Distributional Plots of Outstanding Loan Balance by Loan Age, Risk Band.** A standard box plot of the outstanding loan balance by loan age and risk band for the filtered sample of 51,118 loans from the ABS bonds CARMX, ALLY, SDART, and DRIVE issued in 2017 and summarized in Section I.B. Even for loans as old as 60 months with original terms of 72-73 months, the average outstanding balances exceed $5,000. This suggests that consumers have the potential for meaningful savings despite credit risk convergence not occurring until later in a loan’s lifetime.

refinancing by loan age 42, when they converge in credit risk with prime borrowers. Overall, the potential total savings over the life of the loan for subprime borrowers ranges between $283-1,616. In terms of monthly payment, subprime borrowers benefit the most by waiting until loan age 48, when they converge in credit risk to super prime borrowers. In total, the potential monthly payment savings for subprime borrowers ranges between $22-61. As with deep subprime borrowers, it seems most have refinanced by loan age 60. While this is slower than our calculations would suggest, it still indicates borrowers may be attempting to self-correct. On the other hand, we are uncertain as to the borrower’s motivation, a point we discuss more shortly.
In moving to discuss borrowers in lower risk bands, we find slightly different results. As with deep subprime and subprime borrowers, we also find evidence that near-prime and prime borrowers are leaving money on the table. We estimate that near-prime borrowers are eligible for a potential monthly payment savings of $12-40 for a potential total savings of $148-2,206. The figures for prime borrowers are similar; a potential $18-39 in monthly savings for a potential total savings of $324-2,327. On the other hand, we find that both near-prime and prime borrowers should refinance as soon as possible, after 15 months for near-prime borrowers when they converge in credit risk with prime borrowers and after 12 months for prime borrowers when they converge in credit risk with super-prime borrowers. We find that both near-prime and prime borrowers do not start refinancing in earnest until approximately loan age 60, similar to borrowers in the higher risk bands. This suggests that near-prime and prime borrowers manage their loans less efficiently than deep subprime and subprime borrowers, a result that is surprising given typical expectations about borrower sophistication and credit score. We note that the savings assuming the 2019 convergence matrix (Table 2), given its earlier convergence points, are generally more substantial for the recoverable estimates.\footnote{We have omitted these figures for brevity and conservatism, but the corresponding author may be contacted for details.}

It is also of interest to examine loan prepayment behavior. Conveniently, the CSH rate estimator defined in (5), $\hat{\lambda}_{\tau,n}$, is a direct estimator for prepayment behavior, also conditional on survival. Hence, we can report similar figures to Section II.B but instead focus on borrower prepayment behavior conditioning on the set of current loans. From this, we can attempt to explain consumer behavior and assess if borrowers are acting on the potential savings reported in Table 3. For context, we also overlay two additional economic variables. The first is the Manheim Used Auto Price Index\footnote{For reference, the Bloomberg ticker is: MUVVU.} (Manheim, 2023), which is a common industry assessment of the prevailing value of used automobiles. Given the unusual observations in the used auto market during the COVID-19 pandemic (Rosenbaum, 2020), it is possible that higher-than-expected trade-in values motivated consumers to prepay their loans. Additionally, the United States federal government provided individuals with three direct payments known as Economic Impact Payments (EIPs) and expanded the Childcare Tax Credit (CTC) during the observation period of our sample (U.S. Government Accountability Office, 2022).
Table 3: **Estimated Savings by Risk Band, Loan Age.** The potential savings for a borrower who refinances at the average interest rate of a lower risk band after the point of credit risk convergence in Table 2 (S = subprime, NP = near-prime, P = prime, SP = super-prime).

| Age | Cnt | Bal ($) | Pmt ($) | APR (%) | Pmts (#) | Mo Pmt Savings ($) | Total Savings ($) |
|-----|-----|---------|---------|---------|----------|-------------------|------------------|
|     |     |         |         |         |          | S | NP | P | SP | S | NP | P | SP |
| deep subprime | | | | | | | | | | | | | |
| 12  | 6.084  | 13,295 | 344 | 23.01 | 65 | 577 | 813 | 1,212 |
| 15  | 5.622  | 12,887 | 343 | 22.99 | 62 | 442 | 685 | 1,127 |
| 18  | 5.002  | 12,598 | 343 | 22.99 | 60 | 386 | 513 | 899 |
| 24  | 4.082  | 11,976 | 341 | 22.92 | 55 | 315 | 462 | 777 |
| 30  | 3.333  | 11,192 | 342 | 22.87 | 49 | 261 | 393 | 654 |
| 36  | 2.766  | 10,245 | 341 | 22.83 | 43 | 217 | 325 | 542 |
| 42  | 2.170  | 9,187 | 339 | 22.85 | 37 | 182 | 277 | 465 |
| 48  | 1.782  | 8,237 | 342 | 22.83 | 32 | 161 | 234 | 395 |
| 54  | 1.062  | 6,897 | 338 | 22.81 | 26 | 138 | 208 | 346 |
| 60  | 4      | 7,493 | 348 | 21.34 | 27 | 118 | 168 | 286 |
| subprime | | | | | | | | | | | | |
| 12  | 18.261 | 16,693 | 395 | 17.97 | 64 | 897 | 1,341 | 2,238 |
| 15  | 17.021 | 16,126 | 394 | 17.98 | 61 | 817 | 1,229 | 1,946 |
| 18  | 15.487 | 15,619 | 393 | 17.97 | 59 | 762 | 1,146 | 1,808 |
| 24  | 12.997 | 14,621 | 389 | 17.94 | 54 | 677 | 1,034 | 1,611 |
| 30  | 11.021 | 13,420 | 388 | 17.94 | 48 | 592 | 905 | 1,497 |
| 36  | 9.309  | 12,194 | 386 | 17.94 | 42 | 522 | 773 | 1,395 |
| 42  | 7.481  | 10,835 | 384 | 17.93 | 37 | 462 | 630 | 1,292 |
| 48  | 6.192  | 9,506 | 383 | 17.92 | 31 | 417 | 560 | 1,177 |
| 54  | 5.901  | 8,953 | 383 | 17.93 | 29 | 377 | 507 | 984 |
| 60  | 12     | 7,398 | 348 | 17.92 | 27 | 332 | 469 | 801 |
| near-prime | | | | | | | | | | | | |
| 12  | 5.807  | 19,111 | 411 | 12.79 | 64 | 490 | 678 | 1,168 |
| 15  | 5.587  | 18,245 | 407 | 12.76 | 60 | 457 | 626 | 1,083 |
| 18  | 5.315  | 17,617 | 405 | 12.74 | 58 | 428 | 586 | 1,014 |
| 24  | 4.692  | 16,204 | 402 | 12.72 | 52 | 399 | 549 | 948 |
| 30  | 4.146  | 14,694 | 400 | 12.71 | 47 | 371 | 498 | 869 |
| 36  | 3.592  | 13,187 | 398 | 12.71 | 41 | 343 | 446 | 790 |
| 42  | 3.041  | 11,446 | 394 | 12.67 | 35 | 316 | 381 | 697 |
| 48  | 2.622  | 9,862 | 394 | 12.68 | 29 | 289 | 327 | 516 |
| 54  | 2.455  | 9,283 | 395 | 12.69 | 27 | 263 | 297 | 440 |
| 60  | 1.663  | 8,218 | 400 | 12.69 | 24 | 237 | 255 | 382 |
| prime | | | | | | | | | | | | |
| 12  | 5.173  | 18,582 | 358 | 7.83 | 64 | 370 | 456 | 826 |
| 15  | 5.283  | 17,611 | 354 | 7.81 | 60 | 336 | 422 | 758 |
| 18  | 5.315  | 16,706 | 350 | 7.78 | 57 | 303 | 389 | 712 |
| 24  | 4.971  | 15,097 | 346 | 7.76 | 52 | 272 | 348 | 620 |
| 30  | 4.538  | 13,503 | 345 | 7.74 | 46 | 244 | 314 | 558 |
| 36  | 4.086  | 11,866 | 344 | 7.73 | 39 | 217 | 283 | 490 |
| 42  | 3.697  | 10,274 | 342 | 7.72 | 34 | 191 | 247 | 438 |
| 48  | 3.191  | 8,615 | 343 | 7.71 | 28 | 165 | 213 | 368 |
| 54  | 2.963  | 8,101 | 345 | 7.71 | 26 | 150 | 186 | 336 |
| 60  | 1.898  | 7,075 | 351 | 7.66 | 22 | 125 | 157 | 282 |
|     | 28     | 6,920 | 345 | 7.74 | 17 | 111 | 142 | 253 |
Figure 10: Consumer Prepayment Behavior, Used Autos, Economic Stimulus. (top) A plot of the Manheim Used Auto Index (price) (Manheim, 2023) by loan age for the sample of 58,118 filtered loans issued in 2017 (left) and 65,892 loans issued in 2019 (right). (bottom) A plot of $\hat{\lambda}_{02: \tau,n}$ (prepayments) defined in (5) by loan age for all risk bands within the sample of 58,118 filtered loans issued in 2017 (left) and 65,892 loans issued in 2019 (right), plus 95% confidence intervals using Lemma 1. By the hypothesis test defined in (6), there is very little difference in prepayment behavior conditional on survival by risk band. The labels E1, E2, E3, and C indicate the timing of the Economic Impact Payments and Childcare Tax Credit expansion (U.S. Government Accountability Office, 2022). It appears prepayment behavior was motivated by used auto prices and economic stimulus payments rather than a borrower’s updated assessment of their risk profile.

It is thus possible that borrowers, upon receiving these cash payments, made the decision to purchase a new vehicle and thus prepay. The results are presented in Figure 10.

We report some interesting observations. First, there appears to be very little difference in prepayment behavior by risk band throughout the life of the loan, which differs significantly from default rates. This is especially so for the sample of 58,118 loans issued in 2017. Visually, we see some differences in the sample of 65,892 loans issued in 2019, but many of
the asymptotic confidence intervals still overlap by risk band. Second, there does appear to be a meaningful connection between prevailing used auto prices and borrower prepayment behavior. That is, as the value of used autos rose, borrowers of current loans appear to increase prepayment frequency. Indeed, prepayments occur at a higher rate sooner in the 2019 issuance, when the value of used autos increased earlier in the loan’s lifetimes in comparison to the 2017 issuance. Furthermore, the timing of economic stimulus payments plotted against prepayment behavior is also telling. The prepayment rates for both the 2017 and 2019 issuance also increase shortly after individuals would have received the first direct EIP from the U.S. federal government. Because of the potential savings we observe in Table 3, it is possible that the EIPs may have also provided individuals with further implicit economic gains, if they used the EIPs to refinance at a lower interest rate. The results of Figure 10 in connection with Table 3 taken together suggest that individual borrowers may not consider their updated risk profile in deciding to prepay. Instead, the borrowers may be more motivated by economic indicators that are more tangible, such as direct cash payments or higher trade-in values.

Given the results in Table 3, it begs the question: why does the market for mature consumer auto loans appear to operate inefficiently? A natural starting point is a lack of borrower sophistication in performing an updated personal risk assessment as a loan remains current. Generally, the typical consumer has a poor reputation in making financial decisions (e.g. Gross and Souleles, 2002; Stango and Zinman, 2011; Lusardi and de Bassa Scherberger, 2013; Campbell, 2016; Heidhues and Köszegi, 2016; Dobbie et al., 2021), and the type of calculations we perform herein assume some advanced expertise, such as a working understanding of actuarial mathematics. An inability to self-assess creditworthiness within financial markets against a current APR seems to plague borrowers within all risk bands, as we find the surprising result that it is actually the near-prime and prime borrowers that leave the most money on the table by delaying prepayment.

It may not be fair to blame this perceived borrower inefficiency solely on the borrowers, however. A borrower’s main tool to assess creditworthiness is their credit score. While consumers have obtained better access to credit scores, they may update too slowly within the context of a 72-73 month consumer auto loan to motivate a borrower to seek out a lower rate. Additionally, such borrowers may face friction in attempting to refinance mature auto loans, either through limited options, refinance fees, or perceived hassle. Indeed, encour-
aging borrowers to self-correct has proven to be less effective in practice (e.g., Keys et al., 2016; Agarwal et al., 2017). From this point of view, we see an opportunity for lenders to target these mature loans from borrowers in higher risk bands. Because a borrower that stays current eventually outperforms their initial risk profile and loan APRs are constant throughout the life of the loan, it is not a leap in logic to suggest there exists a lower rate that would both lower this borrower’s financing cost and be profitable to a second lender. On the other hand, lenders themselves may face similar market frictions, such as an inability to identify these borrowers or unattractive returns after accounting for the full scope of origination costs. We are optimistic that continued increases in financial technology may lower these possible hurdles for both borrowers and lenders.

To spur future research, we offer two potential solutions. The first is that we see a market ripe for financial innovation. Specifically, we propose that lenders offer a loan structure with a reducing payment based on good performance, an adjustable payment loan. It is likely lenders already possess the data needed to provide pricing structures capable of adjusting for a borrower’s updated risk profile. We postulate that a lower future payment may act as an incentive for a borrower to remain active and paying, which could work to offset potential profit losses from lowering rates to these high-interest rate loans that perform well. We caution lenders from making opposite adjustments, however, as increasing payments in response to poor performance (i.e., sudden delinquencies) may further discourage a likely overwhelmed borrower or lead to adverse selection (though late payment penalties are common). Second, there is always the regulatory angle, which has been successful in other consumer lending spaces (e.g., Stango and Zinman, 2011; Agarwal et al., 2014).

---

31There are examples of specialty finance companies in the student loan space that attempt to refinance borrowers into lower interest rates (e.g., SoFi). The size (and potential profitability) of such loans may be larger than auto loans, however.

32President Barack Obama remarked during the signing of the Dodd-Frank Wall Street Reform and Consumer Protection Act that, “We all win when consumers are protected against abuse. And we all win when folks are rewarded based on how well they perform, not how well they evade accountability” (Obama, 2010). The terms “abuse” and “evade accountability” feel strong, but attempting to reward borrowers based on good performance feels aligned in spirit with an ideal of merit-based economic gains.

33It is possible that the process of securitization, whereby default risk is transferred off a lender’s balance sheet, creates a disincentive for lenders to maintain a continued interest in loan performance. At the same time, for loans not securitized, it is difficult to ask a for-profit lender to actively seek out lower profits. We suspect the branch with the most potential fruit is a second opportunistic lender, or perhaps some specialty finance companies connected with responsible investing (i.e., environmental, social, and governance (ESG), socially responsible investing (SRI), or impact investing.

34One very positive example is Reg AB II (Securities and Exchange Commission, 2014), which has made
ample, there is potentially minimal additional cost to lenders to require ongoing loans to be underwritten again after a set period of good performance, say 36 months, especially given the lender will already have most of the borrower’s information. Ideally, this update would not count as a formal inquiry against the borrower’s credit report. Further, given the results of Figure 10, an initial cash payment incentive to borrowers may provide sufficient motivation to get borrowers to refinance.\textsuperscript{35} On the other hand, regulatory intervention to increase the cost of lending may lead to these extra costs being pushed back to the borrowers. More research is needed.

\section*{IV. Conclusion}

This article tells a familiar financial story in a new way. We arrive at the familiar aspect, consumers behaving in a way that is financially inefficient, from the novel starting point of credit risk convergence: the phenomenon that all current loans eventually converge in default risk, regardless of each loan’s initial credit risk. We find empirical evidence of credit risk convergence using large-sample statistical tools from the field of survival analysis. We analyze over 140,000 consumer automobile loans from three different samples taken from ABS bonds spanning nearly six years: Spring 2017 through Winter 2023. We find that conditional credit risk converges between disparate risk bands after 10-52 months of current payments for 72-73 month auto loans. The rate of convergence depends on the two risk bands being compared, with the full matrix in Table 2. These results are robust to various sensitivity tests for the Coronavirus pandemic, loans secured with new or used vehicles, and the parent business model of the loan originator.

We follow the empirical evidence of credit risk convergence with a thorough financial analysis of these loans. We combine the empirical probabilistic estimates underlying the evidence of credit risk convergence with risk band APRs to estimate month-by-month annualized risk-adjusted returns for lenders. It assigns values to the typical back-loading of profits commonly found in high-risk loans: deep subprime borrowers have early unstable risk-adjusted returns around 5% before eventually increasing to nearly 15%. We then consider a

\textsuperscript{35}The overall economic impact of such a program may be mixed, given the results for the “cash for clunkers” program (Mian and Sufi, 2012). Alternatively, competing lenders themselves may offer cash to borrowers in exchange for refinancing.
consumer perspective. Because a borrower’s APR reflects a single point-in-time assessment of credit risk (i.e., risk-based pricing), a high-risk borrower that remains current eventually outperforms the stale APR. We find borrowers in all risk bands below super prime delay prepayment in a way that is economically inefficient. We find prime and near-prime borrowers leave up to $2,327 and $2,206 in total potential savings on the table, respectively. In a surprise, this outpaces subprime and deep subprime borrowers, who leave up to $1,616 and $1,212 in total potential savings on the table, respectively. We then utilize the survival analysis estimators to model a current borrower’s prepayment behavior. In a visual analysis, it appears borrowers’ prepayment behavior was motivated by sharp increases in used auto values over the observation period and economic stimulus payments. This suggests that current borrowers may not look to refinance into lower rates based on a changing risk profile. We then opine that market frictions may exist both from the perspective of the borrower and lender that lead to these inefficiencies to persist.

In closing, our two contributions are intertwined. We suspect the phenomenon of credit risk convergence will appear in other forms fixed-income debt, such as other types of consumer loans but also more broadly, even corporate and sovereign debt. To make these claims formally, more study is needed, however. The credit risk convergence results than bleed into an analysis of consumer automobile loans, where we find borrowers delay prepayment in a manner that is economically inefficient. We suspect these results will extend to other forms of consumer debt, such as residential mortgages, but more study is needed.

Appendix

A. Asymptotic Properties of the Cause-Specific Estimator

The vector of estimators using (5) for $\Delta + 1 \leq x \leq \xi$ has convenient asymptotic properties, which we now summarize. For proofs of these properties, see Online Appendix A.

**Proposition 1 (\(\hat{\Lambda}_{\tau,n}^{0i}\) Asymptotic Properties).** For $i \in \{1, 2\}$, define $\hat{\Lambda}_{\tau,n}^{0i} = (\hat{\lambda}_{\tau,n}^{0i}(\Delta + 1), \ldots, \hat{\lambda}_{\tau,n}^{0i}(\xi))^\top$, where

$$\hat{\lambda}_{\tau,n}^{0i}(x) = \frac{\hat{f}_{*,\tau,n}^{0i}(x)}{U_{\tau,n}(x)} = \frac{\sum_{j=1}^{n} 1_{X_j \leq C_j} 1_{Z_{X_j} = i} 1_{\min(X_j, C_j) = x}}{\sum_{j=1}^{n} 1_{Y_j \leq x \leq \min(X_j, C_j)}}.$$
Then,

(i) \[ \hat{\Lambda}^{0i}_{\tau,n} \xrightarrow{P} \Lambda^0_{\tau}, \text{ as } n \to \infty; \]

(ii) \[ \sqrt{n}(\hat{\Lambda}^{0i}_{\tau,n} - \Lambda^0_{\tau}) \xrightarrow{D} N(0, \Sigma^{0i}), \text{ as } n \to \infty, \]

where \[ \Lambda^0_{\tau} = (\lambda^0_{\tau}(\Delta + 1), \ldots, \lambda^0_{\tau}(\xi))^\top \] with \[ \lambda^0_{\tau,n}(x) = f^0_{i,*}(x)/U^0_{\tau}(x) \] and

\[ \Sigma^{0i} = \text{diag}\left(f^0_{i,*}(\Delta + 1)\{U^0_{\tau}(\Delta + 1) - f^0_{i,*}(\Delta + 1)\}, \ldots, f^0_{i,*}(\xi)\{U^0_{\tau}(\xi) - f^0_{i,*}(\xi)\}\right). \]

That is, the cause-specific hazard rate estimators \[ \hat{\lambda}^{0i}_{\tau,n}(\Delta + 1), \ldots, \hat{\lambda}^{0i}_{\tau,n}(\xi) \] are consistent, asymptotically normal, and independent.

Often, it is of interest to construct confidence intervals for the cause-specific hazard rate estimators such that the confidence intervals have the desirable property of falling within the interval (0, 1). We may do so as follows.

**Lemma 1** \((\lambda^0_{\tau}(x) (1-\theta)% Confidence Interval).** The \((1-\theta)% asymptotic confidence interval bounded within \((0, 1)\) for \(\lambda^0_{\tau}(x), x \in \{\Delta + 1, \ldots, \xi\}, i = 1, 2\) is

\[ \exp\left\{ \ln \hat{\lambda}^{0i}_{\tau,n}(x) \pm Z_{(1-\theta/2)} \sqrt{\frac{\hat{U}^0_{\tau,n}(x) - \hat{f}^0_{i,*}(x)}{n\hat{U}^0_{\tau,n}(x)\hat{f}^0_{i,*}(x)}} \right\}, \] (8)

where \(Z_{(1-\theta/2)}\) represents the \((1 - \theta/2)th percentile of the standard normal distribution.**

**B. Empirical Evidence: Additional Details**

The purpose of this section is to provide additional details related to Section II.B. We plot the full five-by-five matrix of CSH rate estimates for default in Figure 11 for the sample of 58,118 loans issued in 2017. It is a complete extension of the subprime versus prime plot in Figure 3. That is, Figure 3 is a zoomed-in view of the subprime-prime cell (row 4, column 2) in Figure 11. There are a few consistent observations. First, we generally see that the
monthly conditional default rate declines as the credit quality of the risk band improves, as expected. We again see a large increase in the hazard rate for the deep subprime, subprime, and near prime risk bands around loan age 40. With some approximate date arithmetic from the first payment month of the ABS bonds (March-April-May 2017), we find that a loan age of 40 months corresponds to roughly Spring 2020 (when adjusted for left-truncation). This corresponds to the economic impact of the Coronavirus pandemic, which effectively stopped most economic activity in Spring 2020. It is interesting that the economic shutdown of the Coronavirus appears to have had minimal impact on the prime risk band and almost no notable impact on the super-prime risk band. We remark here that, due to left-truncation and right-censoring, we have recoverable estimates of the hazard rate spanning approximately the full range $4 \leq X \leq 54$. By comparing the asymptotic confidence intervals within each risk band comparison by loan age, we find many examples of credit risk convergence. Figure 11 may be compared with the matrix in the top row of Table 2.

C. A Digression on COVID-19: Additional Details

The purpose of this section is to provide additional details related to Section II.C. We plot the full five-by-five matrix of CSH rate estimates for default in Figure 12 for the sample of 65,892 loans issued in 2019. It is a complete extension of the subprime versus prime plot in Figure 4. That is, Figure 4 is a zoomed-in view of the subprime-prime cell (row 4, column 2) in Figure 12. The purpose of considering the 2019 issuance is for the sensitivity testing related to COVID-19 (see Section II.C). If the phenomenon of credit risk convergence is completely driven by the Spring 2020 economic shutdown, we would expect to see it occur much earlier in the 2019 sample of bonds when subject to the same loan selection process and risk band definitions of Section I.A.

As expected, we see the large spike in the cause-specific hazard rate for defaults around loan age 10, which, when adjusted for left-truncation, corresponds to the Spring 2020 economic shutdown. It occurs much sooner in the comparison to the 2017 issuance. Overall, we see evidence of earlier convergence, and so the shock of the economic shutdown of Spring 2020 has played some role. Not all risk bands have converged by the end of the recoverable window of the 2019 sample (43 months), however, and so credit risk convergence is not solely a product of COVID-19. In other words, loan age, in addition to the economic shutdown of
Figure 11: Credit Risk Convergence: All Risk Bands (2017). A plot of $\hat{\lambda}_{01}^{\tau,n}$ (defaults) defined in (5) by loan age for all five risk bands within the sample of 58,188 loans (Table 1), plus 95% confidence intervals using Lemma 1. We may use the hypothesis test described in (6) by searching for the minimum age that the confidence intervals overlap between two disparate risk bands. The large upward spike in $\hat{\lambda}_{01}^{\tau,n}$ for the deep subprime, subprime, and near-prime risk bands around loan age 40 is related to the economic impact of COVID-19, a point discussed more fully in Section II.C. Due to left-truncation and right-censoring, the recoverable range is approximately $4 \leq X \leq 54$. Spring 2020, plays a role in credit risk convergence.
Figure 12: Credit Risk Convergence: All Risk Bands (2019). A plot of $\hat{\lambda}_{01}^{\tau,n}$ (defaults) defined in (5) by loan age for all five risk bands within the sample of 65,892 loans (Section II.C), plus 95% confidence intervals using Lemma 1. It is a repeat of Figure 11 for the 2019 issuance as a sensitivity check that the economic shock of COVID-19 is not the sole reason for credit risk convergence between disparate risk bands. The earlier convergence for the 2019 issuance versus the 2017 issuance suggests COVID-19 has played a role, but convergence does not occur immediately following COVID-19. This suggests that loan age also plays a role. Due to left-truncation and right-censoring, the recoverable range is $4, 10 \leq X \leq 30, 35, 38 \leq X \leq 43$.

D. Estimating Recovery Upon Default

Consumer auto loans are secured with the collateral of the attached automobile. In the event of a defaulted loan, the lender has legal standing to repossess the vehicle to make up the
outstanding balance of the loan. In most cases, particularly for deep subprime and subprime borrowers, the estimated value of a repossessed automobile in the event of default is an important component in the initial pricing of a loan. In this section, therefore, we briefly discuss our process to estimate a recovery assumption by loan age, which is ultimately defined as a percentage of the initial loan balance. Our estimates will be used in the analysis of Section III.A, but we acknowledge the empirical results may also be of interest to readers more generally. We thus present our estimated recovery curve for the 2017 issuance (see Section I.B) in Figure 13.

The results of Figure 13 utilize the detailed reporting of the loan level data of Securities and Exchange Commission (2016) to perform the estimation for both the filtered sample of 58,118 loans issued in 2017 and summarized in Section I.B and the filtered sample of 65,892 loans issued in 2019 and summarized in Section II.C. Specifically, we calculate a sum total of the \texttt{recoveredAmount} field for all loans that ended in default. The \texttt{recoveredAmount} field includes any additional loan payments made by the borrower after defaulting, legal settlements, and repossession proceeds (Securities and Exchange Commission, 2016). We then divide the total \texttt{recoveredAmount} by the \texttt{originalLoanAmount} for each defaulted loan. Finally, we take an average of these recovery percentages by age of default in months. The point estimates may be found in Figure 13. Next, for convenient use within the lender profitability analysis of Section III.A, we nonparametrically smooth the point estimates using the \texttt{loess()} function in R (R Core Team, 2022). See the dashed line in Figure 13. This nonparametric \texttt{loess} curve is then fitted to a gamma-kernel via ordinary minimization of a sum-of-squared differences, which allows for extrapolation beyond the recoverable sample space. See the solid line in Figure 13.

The shape of the recovery curve warrants some commentary. Loans that default shortly after origination generally have a low recovery amount as a percentage of the initial loan balance, between 10-20%. This is likely because a loan that defaults so quickly after origination may be due to fraud in the initial loan application\footnote{There is some anecdotal evidence that we share, as we find it illustrative. Lautier met with financial representatives from a high-end luxury auto manufacturer in late 2016. The representative noted they only had eight total defaults on their U.S. portfolio in the past year, and could thus recall what happened in each instance. He reported two cases of fraud: a new car was purchased with a down payment and first payment, and the car was immediately loaded into a shipping container and shipped outside the continental United States; no future loan payments were made. Any recovery in such an instance will likely be minimal.}, extreme circumstances for the
borrower (i.e., rapid decline in physical health), or severe damage to the vehicle. In the case of damage to the vehicle, it is possible the borrower has also lapsed on auto insurance. Overall, it can be difficult to recover a meaningful amount in these circumstances. The recovery percentage then peaks at month 12 at just over 42% before declining towards zero as the loan age approaches termination (72–73 months). Since all vehicles in our sample are used, the decline in recoveries reflects the typical depreciating value of the automobile over time.

We close this section by noting the economic welfare of an automobile repossession has
attracted the attention of researchers. Generally, the results are mixed. On the one hand, Pollard et al. (2021) discuss a vicious cycle of subprime auto lending where the same car may be bought, sold, and repossessed 20-30 times. This suggests repossessions may negatively impact economic welfare. A earlier result by Cohen (1998) finds that manufacturers prefer to offer prospective borrowers interest discounts over equivalent cash rebates because a legal technicality finds such a discount is financially beneficial to the lender in the event of repossession. In this case, the legal circumstances of a repossession may influence market behavior. Along the same lines and an argument for the potential economic benefits of repossession, Assunção et al. (2013) find that a 2004 credit reform in Brazil, which simplified the sale of repossessed cars, lead to an expansion of credit for riskier, self-employed borrowers. In other words, a reform designed to make recouping money from a repossessed automobile easier for lenders improved the ability of riskier borrowers to access credit. It is noteworthy, however, that the reform also lead to increased incidences of delinquencies and default.

E. Estimating the Remaining Lifetime Risk-Adjusted Return

Denote the risk-adjusted rate of return for a loan in risk band \( a \) as \( \rho_a \). Given reliable estimates of borrower default and prepayment probabilities, such as those in Section II.A, we may estimate \( \rho_a \) for a given loan in risk band \( a \). In particular, we may estimate \( \rho_a \) for each month a loan is still active and paying to find a conditional risk-adjusted rate of return over a loan’s full remaining lifetime.\(^{37}\) Pleasingly, \( \rho_a \) equals the loan contract effective rate of return in the event the future loan payments will proceed as scheduled with no uncertainty, which we state formally in Proposition 2. For proof, see Online Appendix B.

**Proposition 2 (Risk-Adjusted Rate of Return with No Payment Uncertainty).** Suppose a loan is originated with an initial balance, \( B \), a monthly rate of interest, \( r_a \), and a term of \( \psi \) months. Let \( \rho_{a|x} \) denote the risk-adjusted rate of return given the loan has survived to month \( x \). If the probability that all payments will follow the amortization schedule exactly is unity (i.e., no payment uncertainty), then \( \rho_{a|x} = r_a \) for all \( x \in \{1, \ldots, \psi\} \).

We now formalize the estimation of \( \rho_{a|x} \), as defined in Proposition 2. For convenience of notation, we will drop \( a \) to denote the arbitrary risk band and assume the proceeding

\(^{37}\)Contrast this with Section III.A, in which we calculate a one-month risk-adjusted return.
calculations will be performed entirely within one risk band. Assume we consider a loan with a 𝜨-month schedule. Denote the current age of a loan by \( x \), \( 1 \leq x \leq \psi \).\(^{38}\) Let the cause-specific hazard rate for default at time \( x \) be denoted by \( \lambda_0^1(x) \) and the cause-specific hazard rate for repayment at time \( x \) be denoted by \( \lambda_0^2(x) \). Assuming no other causes for a loan termination, the all-cause hazard rate is then \( \lambda(x) = \lambda_0^1(x) + \lambda_0^2(x) \). Further, recall (2) and observe for \( i = 1, 2, x \leq j \leq \psi \),

\[
\Pr(X = j, Z_x = i) = \frac{\Pr(X = j, Z_x = i)}{\Pr(X \geq x)} \Pr(X \geq x)
\]

\[
= \Pr(X = j, Z_x = i \mid X \geq x) \Pr(X \geq x)
\]

\[
= \lambda_0^i(j) \prod_{k=x}^{j-1} \{1 - \lambda(k)\},
\]

again with the convention \( \prod_{k=x}^{x-1} \{1 - \lambda(k)\} = 1 \). For convenience, denote \( p_0^i(j) = \Pr(X = j, Z_x = i \mid X \geq x) \) for \( i = 1, 2, x \leq j \leq \psi \). Hence,

\[
p_0^i(j) = \begin{cases} 
\lambda_0^i(x), & j = x \\
\lambda_0^i(j) \prod_{k=x}^{j-1} \{1 - \lambda(k)\}, & j > x,
\end{cases} \quad i = 1, 2.
\]

One may verify \( \sum_{j=x}^{\psi} \sum_{i=1}^{2} p_0^i(j) = 1 \) for every \( x \).\(^{39}\)

We estimate \( \rho_x \) as follows. Let the scheduled amortization loan balance of a consumer auto loan at month \( x \), \( 1 \leq x \leq \psi \) be denoted by \( B_x \), where \( B_\psi = 0 \). Denote the scheduled monthly payment by \( P \). If we denote the recovery of a defaulted consumer auto loan at month \( x \), \( 1 \leq x \leq \psi \), by \( R_x \), then the default matrix at loan age \( x \leq \psi - 1 \) for the possible

\(^{38}\) Depending on the impact of left-truncation and right-censoring, the recoverable range of \( X \) may not be the entire original loan termination schedule (see Section II.A for details). In such an instance, assumptions about the probability distribution may be necessary. Assuming a geometric right-tail (i.e., a constant hazard rate that follows the last recoverable value) is common in survival analysis (Klugman et al., 2012, Section 12.1). We will proceed as though the full distribution is recoverable and allow readers to adjust as needed.

\(^{39}\) It may be of help to review the numeric example of Table 4 in Online Appendix C.
future default paths is

\[
\text{DEF}_{(ψ−x+1)×(ψ−x+1)} = \begin{bmatrix}
R_x & 0 & 0 & \ldots & 0 & 0 \\
 P & R_{x+1} & 0 & \ldots & 0 & 0 \\
 P & P & R_{x+2} & \ldots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
 P & P & P & \ldots & R_{ψ−1} & 0 \\
 P & P & P & \ldots & P & R_{ψ}
\end{bmatrix}.
\]

Note that row 1 of DEF would be the cash flows assuming a default at loan age \( x \), which occurs with probability \( p_{0x}^{(1)}(x) \). Similarly, row 2 of DEF would be the cash flows assuming a default at loan age \( x + 1 \), which occurs with estimated probability \( p_{0x}^{(1)}(x + 1) \), and so on and so forth. In the same way, we can define the prepayment matrix at loan age \( x ≤ ψ − 1 \) as

\[
\text{PRE}_{(ψ−x+1)×(ψ−x+1)} = \begin{bmatrix}
B_x + P & 0 & 0 & \ldots & 0 & 0 \\
 P & B_{x+1} + P & 0 & \ldots & 0 & 0 \\
 P & P & B_{x+2} + P & \ldots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
 P & P & P & \ldots & B_{ψ−1} + P & 0 \\
 P & P & P & \ldots & P & P
\end{bmatrix}.
\]

As with defaults, row 1 of PRE would be the cash flows assuming a prepayment at loan age \( x \), which occurs with estimated probability \( p_{0x}^{(2)}(x) \). Similarly, row 2 of PRE would be the cash flows assuming a prepayment at loan age \( x + 1 \), which occurs with estimated probability \( p_{0x}^{(2)}(x + 1) \), and so on and so forth. Therefore, if we denote the \((ψ − x + 1) \times 1\) dimensional discount vector assuming the unknown monthly rate of \( ρ_x \) as

\[
(\nu_x)^\top = \begin{pmatrix}
(1 + ρ_x)^{-1} & (1 + ρ_x)^{-2} & \ldots & (1 + ρ_x)^{-(ψ−x+1)}
\end{pmatrix}^\top,
\]

and the \((ψ − x + 1) \times 1\) dimensional cause-specific probability vector as

\[
(p_{x}^{(0)})^\top = \begin{pmatrix}
 p_{x}^{(0)}(x) & p_{x}^{(0)}(x+1) & \ldots & p_{x}^{(0)}(ψ)
\end{pmatrix}^\top,
\]

53
then the expected present value (EPV) of a loan at age \( x \leq \psi - 1 \) is

\[
\text{EPV}_x = (p^{01}_x)^\top \text{DEF}_x \nu_x + (p^{02}_x)^\top \text{PRE}_x \nu_x.
\]

Therefore, \( \rho_x \) is the interest rate such that \( B_x = \text{EPV}_x \); that is,

\[
\{ \rho_x : B_x = \text{EPV}_x \}. \tag{9}
\]

In words, \( \rho_x \) represents the expected return realized by lending \( B_x \) and taking into account the original monthly payments \( P \) and default and prepayment risk over the remaining lifetime of the loan. We have \( \rho_x \leq r \) for a given contract, with equality only in the circumstances of Proposition 2. Finally, we of course do not know the true distribution of \( X \). We do have the estimators in (5), however, and Proposition 1. Thus, we may estimate \( \rho_x \) by replacing the cause-specific hazard rates \( \lambda_{ii}^{0} \) with the estimate in (5). For completeness, we close this section with the following lemma.

**Lemma 2** (\( \hat{\rho}_{n,x} \) Asymptotic Properties). Replace the cause-specific hazard rates in (9) with the estimators from (5). Define the estimated risk-adjusted rate of return over the remaining lifetime given a loan has survived to month \( x \) as \( \hat{\rho}_{n,x} \). Then,

\[
\hat{\rho}_{n,x} \xrightarrow{P} \rho_x, \text{ as } n \to \infty.
\]

**REFERENCES**

Adams, William, Liran Einav, and Jonathan Levin, 2009, Liquidity constraints and imperfect information in subprime lending, *American Economic Review* 99, 49–84.

Agarwal, Sumit, Itzhak Ben-David, and Vincent Yao, 2017, Systematic mistakes in the mortgage market and lack of financial sophistication, *Journal of Financial Economics* 123, 42–58.

Agarwal, Sumit, Souphala Chomsisengphet, Neale Mahoney, and Johannes Stroebel, 2014, Regulating consumer financial products: Evidence from credit cards, *The Quarterly Journal of Economics* 130, 111–164.
Alan, Sule, and Gyongyi Loranth, 2013, Subprime consumer credit demand: Evidence from a lender’s pricing experiment, The Review of Financial Studies 26, 2353–2374.

Allcott, Hunt, Joshua Kim, Dmitry Taubinsky, and Jonathan Zinman, 2021, Are high-interest loans predatory? Theory and evidence from payday lending, The Review of Economic Studies 89, 1041–1084.

Ally, 2017, Ally Auto Receivables Trust, Prospectus 2017-3, Ally Auto Assets LLC.

Ally, 2019, Ally Auto Receivables Trust, Prospectus 2019-3, Ally Auto Assets LLC.

Andersen, Per Kragh, Ørnulf Borgan, Richard D. Gill, and Niels Keiding, 1993, Statistical Models Based on Counting Processes (Springer).

Assunção, Juliano J., Efraim Benmelech, and Fernando S. S. Silva, 2013, Repossession and the democratization of credit, The Review of Financial Studies 27, 2661–2689.

Ausubel, Lawrence M., 1991, The failure of competition in the credit card market, The American Economic Review 81, 50–81.

Ayres, Ian, and Peter Siegelman, 1995, Race and gender discrimination in bargaining for a new car, The American Economic Review 85, 304–321.

Banasik, John, Jonathan N. Crook, and L. C. Thomas, 1999, Not if but when will borrowers default, Journal of the Operational Research Society 50, 1185–1190.

Bertrand, Marianne, and Adair Morse, 2011, Information disclosure, cognitive biases, and payday borrowing, The Journal of Finance 66, 1865–1893.

Beyersmann, Jan, Aurélien Latouche, Anika Buchholz, and Martin Schumacher, 2009, Simulating competing risks data in survival analysis, Statistics in Medicine 28, 956–971.

Blumenstock, Gabriel, Stefan Lessmann, and Hsin-Vonn Seow, 2022, Deep learning for survival and competing risk modelling, Journal of the Operational Research Society 73, 26–38.

Butler, Alexander W, Erik J Mayer, and James P Weston, 2022, Racial disparities in the auto loan market, The Review of Financial Studies hhac029.

Calem, Paul S., and Loretta J. Mester, 1995, Consumer behavior and the stickiness of credit-card interest rates, The American Economic Review 85, 1327–1336.

Campbell, John Y., 2016, Restoring rational choice: The challenge of consumer financial regulation, American Economic Review 106, 1–30.
CarMax, 2017, CarMax Auto Owner Trust, Prospectus 2017-2, CarMax Business Services LLC.

CarMax, 2019, CarMax Auto Owner Trust, Prospectus 2019-4, CarMax Business Services LLC.

Cohen, Lloyd, 1998, The puzzling phenomenon of interest-rate discounts on auto loans, *The Journal of Legal Studies* 27, 483–501.

Consumer Financial Protection Bureau, 2019, Borrower risk profiles, url: https://www.consumerfinance.gov/data-research/consumer-credit-trends/auto-loans/borrower-risk-profiles/ (Accessed: 2022-06-15).

Crowder, Martin J, 2001, *Classical Competing Risks* (Chapman and Hall/CRC).

Dai, Hongsheng, Marialuisa Restaino, and Huan Wang, 2016, A class of nonparametric bivariate survival function estimators for randomly censored and truncated data, *Journal of Nonparametric Statistics* 28, 736–751.

De Leonardis, Daniele, and Roberto Rocci, 2008, Assessing the default risk by means of a discrete-time survival analysis approach, *Appl. Stoch. Model. Bus. Ind.* 24, 291–306.

Dirick, Lore, Gerda Claeskens, and Bart Baesens, 2017, Time to default in credit scoring using survival analysis: A benchmark study, *Journal of the Operational Research Society* 68, 652–665.

Dobbie, Will, Andres Liberman, Daniel Paravisini, and Vikram Pathania, 2021, Measuring bias in consumer lending, *The Review of Economic Studies* 88, 2799–2832.

Edelberg, Wendy, 2006, Risk-based pricing of interest rates for consumer loans, *Journal of Monetary Economics* 53, 2283–2298.

Edelberg, Wendy, 2007, Racial dispersion in consumer credit interest rates, Finance and Economics Discussion Series 2007-28, Board of Governors of the Federal Reserve System (U.S.).

Einav, Liran, Mark Jenkins, and Jonathan Levin, 2012, Contract pricing in consumer credit markets, *Econometrica* 80, 1387–1432.

Fine, Jason P, and Robert J Gray, 1999, A proportional hazards model for the subdistribution of a competing risk, *Journal of the American Statistical Association* 94, 496–509.

Frydman, Halina, and Anna Matuszyc, 2022, Random survival forest for competing credit risks, *Journal of the Operational Research Society* 73, 15–25.
Fulford, Scott L., 2015, How important is variability in consumer credit limits?, *Journal of Monetary Economics* 72, 42–63.

Geskus, Ronald B., 2011, Cause-specific cumulative incidence estimation and the Fine and Gray model under both left truncation and right censoring, *Biometrics* 67, 39–49.

Gross, David B., and Nicholas S. Souleles, 2002, Do liquidity constraints and interest rates matter for consumer behavior? Evidence from credit card data, *The Quarterly Journal of Economics* 117, 149–185.

Grunewald, Andreas, Jonathan A Lanning, David C Low, and Tobias Salz, 2020, Auto dealer loan intermediation: Consumer behavior and competitive effects, Working Paper 28136, National Bureau of Economic Research.

Heidhues, Paul, and Botond Kőszegi, 2016, Naïeté-based discrimination, *The Quarterly Journal of Economics* 132, 1019–1054.

Huang, Ying, and Mei-Cheng Wang, 1995, Estimating the occurrence rate for prevalent survival data in competing risks models, *Journal of the American Statistical Association* 90, 1406–1415.

Ishwaran, Hemant, Thomas A. Gerds, Udaya B. Kogalur, Richard D. Moore, Stephen J. Gange, and Bryan M. Lau, 2014, Random survival forests for competing risks, *Biostatistics* 15, 757–773.

Kalbfleisch, John D, and Ross L Prentice, 2011, *The Statistical Analysis of Failure Time Data* (John Wiley & Sons).

Karger, Howard Jacob, 2003, No deals on wheels: How and why the poor pay more for basic transportation, *Journal of Poverty* 7, 93–112.

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

Klugman, Stuart A., Harry H. Panjer, and Gordon E. Willmot, 2012, *Loss Models: From Data to Decisions, Fourth Edition* (John Wiley & Sons, Inc., Hoboken, New Jersey).

Lautier, Jackson, Vladimir Pozdnyakov, and Jun Yan, 2021, Estimating a distribution function for discrete data subject to random truncation with an application to structured finance, *arXiv* 1, 1–26.

Lautier, Jackson P., Vladimir Pozdnyakov, and Jun Yan, 2023, Pricing time-to-event contingent cash flows: A discrete-time survival analysis approach, *Insurance: Mathematics and Economics* 110, 53–71.
Lee, Changhee, William Zame, Jinsung Yoon, and Mihaela van der Schaar, 2018a, Deephit: A deep learning approach to survival analysis with competing risks, *Proceedings of the AAAI Conference on Artificial Intelligence* 32.

Lee, Minjung, Eric J. Feuer, and Jason P. Fine, 2018b, On the analysis of discrete time competing risks data, *Biometrics* 74, 1468–1481.

Lehmann, E.L., and George Casella, 1998, *Theory of Point Estimation, 2nd Edition* (Springer).

Lim, Younghee, Trey Bickham, Cassie M. Dinecola, Julia Broussard, Brittany E. Weber, and Alethia Gregory, 2014, Payday loan use and consumer well-being: What consumers and social workers need to know about payday loans, *Journal of Poverty* 18, 379–398.

Livshits, Igor, 2015, Recent developments in consumer credit and default literature, *Journal of Economic Surveys* 29, 594–613.

Lusardi, Annamaria, and Carlo de Bassa Scheresberg, 2013, Financial literacy and high-cost borrowing in the United States, Working Paper 18969, National Bureau of Economic Research.

Manheim, 2023, Used vehicle value index, url: https://publish.manheim.com/en/services/consulting/used-vehicle-value-index.html (Accessed: 2023-03-14).

Melzer, Brian T., 2011, The real costs of credit access: Evidence from the payday lending market, *The Quarterly Journal of Economics* 126, 517–555.

Mian, Atif, and Amir Sufi, 2012, The effects of fiscal stimulus: Evidence from the 2009 cash for clunkers program, *The Quarterly Journal of Economics* 127, 1107–1142.

Morse, Adair, 2011, Payday lenders: Heroes or villains?, *Journal of Financial Economics* 102, 28–44.

Mukhopadhyay, Nitis, 2000, *Probability and Statistical Inference* (Marcel Dekker, New York, NY).

Obama, Barack, 2010, Remarks by the President at Signing of Dodd-Frank Wall Street Reform and Consumer Protection Act, Office of the Press Secretary, The White House.

Phillips, Robert, 2013, Optimizing prices for consumer credit, *Journal of Revenue & Pricing Management* 12.

Pintilie, Melania, 2006, *Competing Risks: A Practical Perspective* (John Wiley & Sons).
Pollard, Jane, Evelyn Blumenberg, and Stephen Brumbaugh, 2021, Driven to debt: Social reproduction and (auto)mobility in los angeles, *Annals of the American Association of Geographers* 111, 1445–1461.

Prentice, R. L., J. D. Kalbfleisch, A. V. Peterson, N. Flournoy, V. T. Farewell, and N. E. Breslow, 1978, The analysis of failure times in the presence of competing risks, *Biometrics* 34, 541–554.

Pressman, Steven, and Robert Scott, 2009, Consumer debt and the measurement of poverty and inequality in the US, *Review of Social Economy* 67, 127–148.

R Core Team, 2022, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

Robb, Cliff, Patryk Babiarz, Ann Woodyard, and Martin Seay, 2015, Bounded rationality and use of alternative financial services, *Journal of Consumer Affairs* 49, 407–435.

Rosenbaum, Eric, 2020, The used car boom is one of the hottest, and trickiest, coronavirus markets for consumers, url: https://www.cnbc.com/2020/10/15/used-car-boom-is-one-of-hottest-coronavirus-markets-for-consumers.html (Accessed: 2023-03-14).

Sankaran, P.G., and Ansa Alphonsa Antony, 2007, Bivariate competing risks models under random left truncation and right censoring, *Sankhyā: The Indian Journal of Statistics (2003-2007)* 69, 425–447.

Santander, 2017a, Drive Auto Receivables Trust, Prospectus 2017-1, Santander Drive Auto Receivables LLC.

Santander, 2017b, Santander Drive Auto Receivables Trust, Prospectus 2017-2, Santander Drive Auto Receivables LLC.

Santander, 2019a, Drive Auto Receivables Trust, Prospectus 2019-4, Santander Drive Auto Receivables LLC.

Santander, 2019b, Santander Drive Auto Receivables Trust, Prospectus 2019-3, Santander Drive Auto Receivables LLC.

Schmid, Matthias, and Moritz Berger, 2021, Competing risks analysis for discrete time-to-event data, *Wiley Interdisciplinary Reviews: Computational Statistics* 13, e1529.

Securities and Exchange Commission, 2014, 17 CFR Parts 229, 230, 232, 239, 240, 243, and 249 Asset-Backed Securities Disclosure and Registration.
Securities and Exchange Commission, 2016, 17 CFR §229.1125 (Item 1125) Schedule AL - Asset-level information.

Stango, Victor, and Jonathan Zinman, 2011, Fuzzy math, disclosure regulation, and market outcomes: Evidence from Truth-in-Lending reform, *The Review of Financial Studies* 24, 506–534.

Staten, Michael, 2015, Risk-based pricing in consumer lending, *Journal of Law, Economics & Policy* 11, 33–58.

Stepanova, Maria, and Lyn Thomas, 2002, Survival analysis methods for personal loan data, *Operations Research* 50, 277–289.

Thackham, Mark, and Jun Ma, 2022, On maximum likelihood estimation of competing risks using the cause-specific semi-parametric cox model with time-varying covariates – an application to credit risk, *Journal of the Operational Research Society* 73, 5–14.

Tutz, Gerhard, and Matthias Schmid, 2016, *Modeling Discrete Time-to-Event Data* (Springer).

U.S. Government Accountability Office, 2022, Stimulus checks: Direct payments to individuals during the covid-19 pandemic, url: https://www.gao.gov/assets/gao-22-106044.pdf (Accessed: 2023-03-02).

Wycinka, Ewa, 2019, Competing risk models of default in the presence of early repayments, *Econometrics* 23, 99–120.

Zhang, Nailong, Qingyu Yang, Aidan Kelleher, and Wujun Si, 2019, A new mixture cure model under competing risks to score online consumer loans, *Quantitative Finance* 19, 1243–1253.

Zingales, Luigi, 2015, Presidential address: Does finance benefit society?, *The Journal of Finance* 70, 1327–1363.
A. Proofs: Section II

Proof of Proposition 1. Statement (i) follows from (ii), so it is enough to show (ii). Let \( \Delta + 1 \leq k \leq \xi \) and observe

\[
\hat{\lambda}^{0i}_{\tau,n}(k) - \lambda^{0i}_\tau(k) \text{=} \frac{1}{n} \sum_{j=1}^{n} 1_{X_j \leq C_j} 1_{Z_{X_j} = i} 1_{\min(X_j, C_j) = k} \frac{f^{0i}_{\tau,*}(k)}{U_\tau(k)} \text{=} \frac{\{\sum_{j=1}^{n} 1_{X_j \leq C_j} 1_{Z_{X_j} = i} 1_{\min(X_j, C_j) = k}\} U_\tau(k) - f^{0i}_{\tau,*}(k) \hat{U}_{\tau,n}(k)}{U_\tau(k) \hat{U}_{\tau,n}(k)} \text{=} \left[ \frac{1}{U_\tau(k) \hat{U}_{\tau,n}(k)} \right] \frac{1}{n} \sum_{j=1}^{n} \{1_{X_j \leq C_j} 1_{Z_{X_j} = i} 1_{\min(X_j, C_j) = k}\} U_\tau(k) - f^{0i}_{\tau,*}(k) 1_{Y_j \leq k \leq \min(X_j, C_j)}.
\]

Define

\[
H^{0i}_{\tau,k(j)} = 1_{X_j \leq C_j} 1_{Z_{X_j} = i} 1_{\min(X_j, C_j) = k} U_\tau(k) - f^{0i}_{\tau,*}(k) 1_{Y_j \leq k \leq \min(X_j, C_j)},
\]

for \( 1 \leq j \leq n \) and

\[
A_{\tau,n} = \text{diag}([\hat{U}_{\tau,n}(\Delta + 1)U_\tau(\Delta + 1)]^{-1}, \ldots, [\hat{U}_{\tau,n}(\xi)U_\tau(\xi)]^{-1}).
\]

Then,

\[
\hat{\Lambda}^{0i}_{\tau,n} - \Lambda^{0i}_\tau = A_{\tau,n} \frac{1}{n} \sum_{j=1}^{n} \begin{bmatrix} H^{0i}_{\tau,\Delta+1(j)} \\ \vdots \\ H^{0i}_{\tau,\xi(j)} \end{bmatrix},
\]

or, letting \( H^{0i}_{\tau,(j)} = (H^{0i}_{\tau,\Delta+1(j)}, \ldots, H^{0i}_{\tau,\xi(j)})^\top \) denote independent and identically distributed random vectors, we have compactly

\[
\hat{\Lambda}^{0i}_{\tau,n} - \Lambda^{0i}_\tau = A_{\tau,n} \frac{1}{n} \sum_{j=1}^{n} H^{0i}_{\tau,(j)}.
\]
It is noteworthy the components of $H_{\tau,(j)}^0$ are uncorrelated. More specifically, 

$$\text{Cov}[H_{\tau,k(j)}^0, H_{\tau,k'(j)}^0] = \begin{cases} U_\tau(k) f_{k,\tau}^0(k)[U_\tau(k) - f_{k,\tau}^0(k)], & k = k' \\ 0, & k \neq k'. \end{cases} \tag{10}$$

To see this, first notice the indicator functions $1_{X_j \leq C_j} 1_{Z_{X_j} = \tau \text{min}(X_j, C_j)} = k$ and $1_{Y_j \leq \text{min}(X_j, C_j)}$ are Bernoulli random variables with probability parameters $f_{\tau,\tau}^0(k)$ and $U_\tau(k)$, respectively. Hence,

$$E H_{\tau,k(j)}^0 = E 1_{X_j \leq C_j} 1_{Z_{X_j} = \tau \text{min}(X_j, C_j)} = k U_\tau(k) - f_{\tau,\tau}^0(k) E 1_{Y_j \leq \text{min}(X_j, C_j)}$$

$$= f_{\tau,\tau}^0(k) U_\tau(k) - f_{\tau,\tau}^0(k) U_\tau(k)$$

$$= 0.$$

Therefore,

$$\text{Cov}[H_{\tau,k(j)}^0, H_{\tau,k'(j)}^0] = E H_{\tau,k(j)}^0 H_{\tau,k'(j)}^0 = E \{1_{X_j \leq C_j} 1_{Z_{X_j} = \tau \text{min}(X_j, C_j)} = k U_\tau(k) - f_{\tau,\tau}^0(k) E 1_{Y_j \leq \text{min}(X_j, C_j)} \}$$

$$\times \{1_{X_j \leq C_j} 1_{Z_{X_j} = \tau \text{min}(X_j, C_j)} = k' U_\tau(k') - f_{\tau,\tau}^0(k') E 1_{Y_j \leq \text{min}(X_j, C_j)} \}$$

$$= U_\tau(k) U_\tau(k') E 1_{X_j \leq C_j} 1_{Z_{X_j} = \tau \text{min}(X_j, C_j)} = k 1_{X_j \leq C_j} 1_{Z_{X_j} = \tau \text{min}(X_j, C_j)}$$

$$- U_\tau(k) f_{\tau,\tau}^0(k') E 1_{X_j \leq C_j} 1_{Z_{X_j} = \tau \text{min}(X_j, C_j)} = k 1_{Y_j \leq \text{min}(X_j, C_j)}$$

$$- U_\tau(k') f_{\tau,\tau}^0(k) E 1_{X_j \leq C_j} 1_{Z_{X_j} = \tau \text{min}(X_j, C_j)} = k' 1_{Y_j \leq \text{min}(X_j, C_j)}$$

$$+ f_{\tau,\tau}^0(k) f_{\tau,\tau}^0(k') E 1_{Y_j \leq \text{min}(X_j, C_j)} 1_{Y_j \leq \text{min}(X_j, C_j)}.$$

We proceed by cases.

Case 1: $k = k'$.

Working through each expectation in $\text{Cov}[H_{\tau,k(j)}^0, H_{\tau,k'(j)}^0]$, we have

$$E 1_{X_j \leq C_j} 1_{Z_{X_j} = \tau \text{min}(X_j, C_j)} = k 1_{X_j \leq C_j} 1_{Z_{X_j} = \tau \text{min}(X_j, C_j)} = k' = E 1_{X_j \leq C_j} 1_{Z_{X_j} = \tau \text{min}(X_j, C_j)} = k$$

$$= f_{\tau,\tau}^0(k),$$

$$= f_{\tau,\tau}^0(k).$$
\[ \begin{align*}
\mathbb{E}_1 X_j \leq C_j \mathbf{1}_{Z_{X_j} = 1} \min(X_j, C_j) = k \mathbf{1}_{Y_j \leq k' \leq \min(X_j, C_j)} &= \mathbb{E}_1 X_j \leq C_j \mathbf{1}_{Z_{X_j} = 1} \min(X_j, C_j) = k' \mathbf{1}_{Y_j \leq k' \leq \min(X_j, C_j)} \\
&= \mathbb{E}_1 X_j \leq C_j \mathbf{1}_{Z_{X_j} = 1} \min(X_j, C_j) = k \\
&= \mathbb{E}_1 Y_j \leq k \leq \min(X_j, C_j) = \mathbb{E}_1 Y_j \leq k' \leq \min(X_j, C_j) = f_{\tau, \tau}(k),
\end{align*} \]

and
\[ \begin{align*}
\mathbb{E}_1 Y_j \leq k \leq \min(X_j, C_j) \mathbf{1}_{Y_j \leq k' \leq \min(X_j, C_j)} &= \mathbb{E}_1 Y_j \leq \min(X_j, C_j) = U_{\tau}(k). 
\end{align*} \]

Thus,
\[ \text{Cov}[H_{\tau, k(j)}^{0i}, H_{\tau, k'(j)}^{0i}] = U_{\tau}(k) f_{\tau, \tau}^{0i}(k) [U_{\tau}(k) - f_{\tau, \tau}^{0i}(k)]. \]

Case 2: \( k \neq k' \).

Working through each expectation in \( \text{Cov}[H_{\tau, k(j)}^{0i}, H_{\tau, k'(j)}^{0i}] \), we have
\[ \begin{align*}
\mathbb{E}_1 X_j \leq C_j \mathbf{1}_{Z_{X_j} = 1} \min(X_j, C_j) = k \mathbf{1}_{X_j \leq C_j} \\
&= \left\{ \begin{array}{ll}
\text{Pr}(X_j \leq C_j, Z_{X_j} = i, \min(X_j, C_j) = k, Y_j \leq k'), & k > k' \\
0, & k < k',
\end{array} \right.
\end{align*} \]

and
\[ \begin{align*}
\mathbb{E}_1 X_j \leq C_j \mathbf{1}_{Z_{X_j} = 1} \min(X_j, C_j) = k' \mathbf{1}_{Y_j \leq k \leq \min(X_j, C_j)} \\
&= \left\{ \begin{array}{ll}
0, & k > k' \\
\text{Pr}(X_j \leq C_j, Z_{X_j} = i, \min(X_j, C_j) = k', Y_j \leq k), & k < k',
\end{array} \right.
\end{align*} \]

and
\[ \begin{align*}
\mathbb{E}_1 Y_j \leq k \leq \min(X_j, C_j) \mathbf{1}_{Y_j \leq k' \leq \min(X_j, C_j)} = \text{Pr}(Y_j \leq k \leq \min(X_j, C_j), Y_j \leq k' \leq \min(X_j, C_j)).
\end{align*} \]
Thus,
\[
\text{Cov}[H_{\tau,k(j)}, H_{\tau,k'(j)}] = f_{\tau}^{\text{oi}}(\min(k, k')) \left\{ -U_{\tau}(\max(k, k')) \Pr(X_j \leq C_j, Z_{X_j} = i, \min(X_j, C_j) = \max(k, k'), Y_j \leq \min(k, k')) \\
+ f_{\tau}^{\text{oi}}(\max(k, k')) \Pr(Y_j \leq k \leq \min(X_j, C_j), Y_j \leq k' \leq \min(X_j, C_j)) \right\}.
\]

However, because of the independence between $Y$ and $(X, Z_X)$,

\[
U_{\tau}(\max(k, k')) = \Pr(Y_j \leq \max(k, k') \leq \min(X_j, C_j))
= \Pr(Y \leq \max(k, k'), X \geq \max(k, k'), C \geq \max(k, k') \mid Y \leq X)
= \{\Pr(Y \leq \max(k, k') \leq C) \Pr(X \geq \max(k, k'))\}/\alpha,
\]

\[
\Pr(X_j \leq C_j, Z_{X_j} = i, \min(X_j, C_j) = \max(k, k'), Y_j \leq \min(k, k'))
= \Pr(C \geq \max(k, k'), Z_X = i, X = \max(k, k'), Y \leq \min(k, k') \mid Y \leq X)
= \{\Pr(X = \max(k, k'), Z_X = i) \Pr(Y \leq \min(k, k'), C \geq \max(k, k'))\}/\alpha,
\]

\[
f_{\tau}^{\text{oi}}(\max(k, k')) = \Pr(X = \max(k, k'), Z_X = i \mid Y \leq X)
= \{\Pr(X = \max(k, k'), Z_X = i) \Pr(Y \leq \min(k, k') \leq C)\}/\alpha,
\]

and

\[
\Pr(Y_j \leq k \leq \min(X_j, C_j), Y_j \leq k' \leq \min(X_j, C_j))
= \Pr(Y \leq \min(k, k'), C \geq \max(k, k'), X \geq \max(k, k') \mid Y \leq X)
= \{\Pr(Y \leq \min(k, k'), C \geq \max(k, k')) \Pr(X \geq \max(k, k'))\}/\alpha.
\]

Therefore,

\[
U_{\tau}(\max(k, k')) \Pr(X_j \leq C_j, Z_{X_j} = i, \min(X_j, C_j) = \max(k, k'), Y_j \leq \min(k, k'))
= f_{\tau}^{\text{oi}}(\max(k, k')) \Pr(Y_j \leq k \leq \min(X_j, C_j), Y_j \leq k' \leq \min(X_j, C_j)),
\]

64
and so Cov\[H^{0i}_{\tau,k(j)}, \ H^{0i}_{\tau,k'(j)}\] = 0 when \(k \neq k'\). This confirms (10). Now define
\[
D^{0i}_\tau = \text{diag}(U_\tau(\Delta + 1)f^{0i}_{*,\tau}(\Delta + 1)[U_\tau(\Delta + 1) - f^{0i}_{*,\tau}(\Delta + 1)], \ldots, U_\tau(\xi)f^{0i}_{*,\tau}(\xi)[U_\tau(\xi) - f^{0i}_{*,\tau}(\xi)]),
\]
and
\[
\bar{H}^{0i}_{\tau,n} = \frac{1}{n} \sum_{j=1}^{n} H^{0i}_{\tau,(j)}.
\]
By the multivariate Central Limit Theorem (Lehmann and Casella, 1998, Theorem 8.21, pg. 61), therefore,
\[
\sqrt{n}(\bar{H}^{0i}_{\tau,n} - 0) \xrightarrow{\mathcal{L}} N(0, D^{0i}_\tau), \text{ as } n \to \infty.
\]
Next, define \(V_\tau = \text{diag}(U_\tau(\Delta + 1)^{-2}, \ldots, U_\tau(\xi)^{-2})\). By Lemma 1 (Lautier et al., 2023), \(A_{\tau,n} \xrightarrow{p} V_\tau\), as \(n \to \infty\). Thus, by multivariate Slutsky’s Theorem (Lehmann and Casella, 1998, Theorem 5.1.6, pg. 283),
\[
\sqrt{n}(A_{\tau,n}\bar{H}^{0i}_{\tau,n}) \xrightarrow{\mathcal{L}} N(0, V_\tau D^{0i}_\tau V_\tau^T), \text{ as } n \to \infty.
\]
We may complete the proof by observing \(V_\tau D^{0i}_\tau V_\tau^T = \Sigma^{0i}\) and \(A_{\tau,n}\bar{H}^{0i}_{\tau,n} = \hat{\Lambda}^{0i}_{\tau,n} - \Lambda^{0i}\).

**Proof of Lemma 1.** The classical method dictates first finding a \((1 - \theta)\)% confidence interval on a log-scale and then converting back to a standard-scale to ensure the estimated confidence interval for the hazard rate, which is a probability, remains in the interval \((0, 1)\). By an application of the Delta Method (Lehmann and Casella, 1998, Theorem 8.12, pg. 58), we have for \(x \in \{\Delta + 1, \ldots, \xi\}\) and \(i = 1, 2\),
\[
\sqrt{n}(\ln \hat{\lambda}^{0i}_{\tau,n}(x) - \ln \lambda^{0i}(x)) \xrightarrow{\mathcal{L}} N\left(0, \frac{\int f^{0i}_{*,\tau}(x)\{U_\tau(x) - f^{0i}_{*,\tau}(x)\}}{U_\tau(x)^3}\frac{1}{\lambda^{0i}(x)^2}\right).
\]
The result follows from (4), the Continuous Mapping Theorem (Mukhopadhyay, 2000, Theorem 5.2.5, pg. 249), the pivotal approach (Mukhopadhyay, 2000, §9.2.2), and converting back to the standard scale.
B. Proofs: Section E

Proof of Proposition 2. For a loan with initial balance, $B$, monthly interest rate, $r_a$, and initial term of $\xi$, the monthly payment, $P$, is

$$P = B \left[1 - \frac{(1 + r_a)^{-\xi}}{r_a}\right]^{-1}.$$  

Assume $x \in \{1, \ldots, \xi\}$. The balance at month $x$, $B_x$ is

$$B_x = B(1 + r_a)^x - P \frac{(1 + r_a)^x - 1}{r_a}$$

$$= B(1 + r_a)^x - B \frac{1 - (1 + r_a)^{-\xi}}{r_a} \frac{(1 + r_a)^x - 1}{r_a}.$$  

(11)

Thus, $\rho_{a|x}$ is the rate such that the expected present value of the future monthly payments equals $B_x$. The payment stream is constant, however, and so

$$B_x = P \frac{1}{1 + \rho_{a|x}} + \cdots + \frac{1}{(1 + \rho_{a|x})^{\xi-x}}$$

$$= B \frac{1 - (1 + r_a)^{-\xi}}{r_a} \frac{1 - (1 + \rho_{a|x})^{-(\xi-x)}}{\rho_{a|x}}.$$  

Use (11) and solve for $\rho_{a|x}$ to complete the proof.

Proof of Lemma 2. The result follows by Proposition 1, part (i) and the Continuous Mapping Theorem (Mukhopadhyay, 2000, Theorem 5.2.5, pg. 249).

C. Large Sample Simulation Study

We present a simulation study in support of Proposition 1 and Lemma 1. Let the true distribution for the lifetime random variable $X$ and bivariate distribution of $(X, Z_X)$ be as in Table 4. The column $p(x)$ denotes the probability of event type 1 given an event at time $X$. This allows us to populate the joint distribution for $Pr(X = x, Z_X = i)$ for $i = 1, 2$. The cause-specific hazard rates then follow from (3), and we also report the all-cause hazard rate
Table 4: **Simulation Study Lifetime of Interest Probabilities.** The true probabilities of the lifetime random variable, $X$, for the simulation study results of Figure 14. The probabilities $p(x)$ and $Pr(X = x)$ for $x \in \{1, \ldots, 10\}$ are selected at onset, and the remaining probabilities in this table may be derived from these quantities. Not summarized here is the truncation random variable, $Y$, which was assumed to be discrete uniform over the integers $\{1, \ldots, 5\}$.

| $p(x)$ | $X$ | $Pr(X = x)$ | $Pr(X = x, Z_x = 1)$ | $Pr(X = x, Z_x = 2)$ | $\lambda^{01}(x)$ | $\lambda^{02}(x)$ | $\lambda(x)$ |
|--------|-----|-------------|----------------------|----------------------|-----------------|-----------------|----------------|
| 0.66   | 1   | 0.04        | 0.026                | 0.014                | 0.026           | 0.014           | 0.04           |
| 0.20   | 2   | 0.06        | 0.012                | 0.048                | 0.013           | 0.050           | 0.06           |
| 0.45   | 3   | 0.10        | 0.045                | 0.055                | 0.050           | 0.061           | 0.11           |
| 0.87   | 4   | 0.14        | 0.122                | 0.018                | 0.152           | 0.023           | 0.18           |
| 0.20   | 5   | 0.09        | 0.018                | 0.072                | 0.027           | 0.109           | 0.14           |
| 0.81   | 6   | 0.06        | 0.049                | 0.011                | 0.085           | 0.020           | 0.11           |
| 0.05   | 7   | 0.14        | 0.007                | 0.133                | 0.014           | 0.261           | 0.27           |
| 0.78   | 8   | 0.18        | 0.140                | 0.040                | 0.379           | 0.107           | 0.49           |
| 0.25   | 9   | 0.07        | 0.018                | 0.053                | 0.092           | 0.276           | 0.37           |
| 0.42   | 10  | 0.12        | 0.050                | 0.070                | 0.420           | 0.580           | 1.00           |

in the final column. Notice that, for each $x$,

$$p(x) = \frac{\lambda^{01}(x)}{\lambda^{01}(x) + \lambda^{02}(x)}.$$  

For the truncation random variable, we assume $Y$ is discrete uniform with sample space $\mathcal{Y} \in \{1, 2, 3, 4, 5\}$. This results in $\alpha = 0.864$. For the purposes of the simulation, we further assume $\tau = 5$. We use the simulation procedure of Beyersmann et al. (2009) but modified for random truncation. Specifically,

1. Simulate the truncation time, $Y$.
2. Set the censoring time to be $Y + \tau$.
3. Simulate the event time, $X$.
4. Simulate a Bernoulli event with probability $p(x)$ to determine if the event $X$ was caused by type 1 with probability $p(x)$ or type 2 with probability $1 - p(x)$.

We simulated $n = 10,000$ lifetimes using the above algorithm. We then tossed any observations that were truncated (i.e., $Y_j > X_j$, for $j = 1, \ldots, n$). This left a sample
of competing risk events subject to censoring, which would be the same incomplete data conditions as a trust of securitized loans. We then used the results of Section II.A to estimate $\hat{f}_{0i}(x)$, $\hat{U}_{0i}(x)$, and $\hat{\lambda}_{0i}(x)$ for $i = 1, 2$ and $x \in \{1, \ldots, 10\}$ over $r = 1,000$ replicates.

To validate the asymptotic results of Proposition 1, we compare the empirical covariance matrix against the derived asymptotic covariance matrix, $\Sigma_{0i}$, by examining estimates of the confidence intervals using Lemma 1. Figure 14 presents the results for the cause-specific hazard rate for cause 01 and 02, respectively. The empirical estimates and 95% confidence intervals are indistinguishable from the true quantities using Proposition 1 and estimated quantities using Proposition 1 but replacing all quantities with their respective estimates from Section II.A.

D. Determination of Loan Outcome

The detail of the loan-level data is extensive, but it remains up to the data analyst to use the provided fields to determine the outcome of an individual loan (see Securities and Exchange Commission (2016) for detail on available field names). To do so, we have aggregated each month of active trust data into a single source file. This allows us to review each bond’s monthly outstanding principal balance, monthly payment received from the borrower, and the portion of each monthly payment applied to principal. Our algorithm to determine a loan outcome proceeded as follows. For each remaining bond after the filtering of Section I.A, we extracted three vectors, each of which was the same length as the number of months a trust was active and paying. The first vector represented the ordered monthly balance, the second was the ordered monthly payments, and the third was the ordered monthly amount of payment applied to principal. We then considered a loan to be repaid if the sum total principal received was greater than the outstanding loan balance as of the first month the trust was actively paying. In this case, the timing of a repayment was set to be the first month with a zero outstanding principal balance. Note that we do not differentiate between a prepayment or naturally scheduled loan amortization; i.e., all repayments have been treated as a “non-default”. If the sum total principal received was less than the first month’s outstanding loan balance, we then considered a loan outcome to be either right-censored or defaulted. To make this determination, we searched the monthly payments received vector for three consecutive zeros (i.e., three straight months of missed payments).
Figure 14: **Simulation Study Results.** A comparison of true $\lambda^0_i(x)$ and estimated $\hat{\lambda}^0_{i,\tau,\tau,n}(x)$, including confidence intervals, for the distribution in Table 4 and $i = 1, 2$. The “true” values are from Proposition 1 and Lemma 1. The “estimate” values use the formulas from Proposition 1 and Lemma 1 but replace the true values with the estimates from Section II.A calculated from the simulated data. The “empirical” values are empirical confidence interval and mean calculations directly from the simulated data. All three quantities are indistinguishable for $n = 10,000$ and 1,000 replicates, which indicates the asymptotic properties hold in this instance.

If we found three consecutive missed payments, we assumed the loan to be defaulted with a time-of-default set to be the month in which the first of three zeros was observed. If we did not find three consecutive months of missed payments, the loan was assumed to be a right-censored observation and assigned an event time as of the last month the trust was actively paying. For the pseudo-code of this algorithm, see Figure 15.
Figure 15: Determination of Loan Outcome. We first extracted three vectors, each of which was the same length as the number of months a trust was active and paying. The first vector (\texttt{bal_vec}) represented the ordered monthly balance, the second (\texttt{pmt_vec}) was the ordered monthly payments, and the third (\texttt{prc_vec}) was the ordered monthly amount of payment applied to principal. We then considered a loan to be repaid if the sum total principal received was greater than the outstanding loan balance as of the first month the trust was actively paying. In this case, the timing of a repayment was set to be the first month with a zero outstanding principal balance. If the sum total principal received was less than the first month’s outstanding loan balance, we then considered a loan outcome to be either right-censored or defaulted. To make this determination, we searched the monthly payments received vector for three consecutive zeros (i.e., three straight months of missed payments). If we found three consecutive missed payments, we assumed the loan to be defaulted with a time-of-default set to be the month in which the first of three zeros was observed. If we did not find three consecutive months of missed payments, the loan was assumed to be a right-censored observation and assigned an event time as of the last month the trust was actively paying.