On the Convergence of Credit Risk in Current Consumer Automobile Loans

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

January 20, 2023

ABSTRACT

Risk-based pricing of loans is well-accepted. Left unstudied, however, is the conditional credit risk of a loan that remains current. Using large-sample statistics and asset-level consumer automobile asset-backed security data, we find that default risk conditional on survival converges for borrowers in disparate credit risk bands well before scheduled termination, a phenomenon we call *credit risk convergence*. We then use actuarial techniques to derive the conditional market-implied credit spread by credit risk band to estimate current deep subprime and subprime borrowers eventually overpay by annual percentage rates of 285-1,391 basis points. Our results are robust to various sensitivity tests.

JEL Codes: C58, D11, D12, D18, D53, G14, G32, G53, P46

Keywords— Adjustable premium loans, asset-level disclosures, competing risks, Coronavirus, COVID-19, economic inequality, financial literacy, market inefficiency, Reg AB II

FOR NEARLY ONE HUNDRED YEARS the American consumer has relied on installment credit to purchase automobiles and other expensive goods (e.g., Hughes and Cain, 2011). And for nearly as long, the discourse of political economy has contemplated the amount of protection required for individual consumers in financial transactions. After the economic calamities of the Great Depression, for example, Franklin D. Roosevelt famously pledged "a new deal for the American people" in his 1932 acceptance speech for the Democratic nomination (Roosevelt, 1932). Exactly thirty years later, the eventual Nobel laureate Milton Friedman wrote, "Underlying most arguments against the free market is a lack of belief in freedom itself"(Friedman, 2002). Forty-eight years later in 2010, shortly following the Global Financial Crisis amid a historical cyclicality befitting poetry, then President 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 literature has naturally followed suit, with noteworthy contributions on payday loans (e.g. Melzer, 2011; Bertrand and Morse, 2011), credit cards (e.g. Gross and Souleles, 2002; Agarwal et al., 2014), and automobile loans (e.g. Adams et al., 2009; Grunewald et al., 2020) to name but a few.

Much of the traditional focus in these studies is on the moments prior to and including contract signing or the aftereffects of such consumer contracts. We examine the well-accepted practice of risk-based pricing (Edelberg, 2006) but spanning the entire lifetime of a current loan. In other words, the risk-based cost of borrowing is traditionally a function of a singular point-in-time historical assessment of a consumer's credit risk profile (Phillips, 2013). A borrower's credit profile is dynamic, however, and this disconnect leads to a natural question: how many on-time payments does it take for a high-risk borrower to no longer be considered high-risk? We find that the answer is considerably less payments than the original term of the loan, which raises an inevitable second question: what are the financial and economic implications of loan contracts for borrowers originally charged a high-risk cost of borrowing that remain current? This paper is dedicated to providing the first empirical answers to both of these questions.

Our findings may be split into two major categories. The first is a statistical argument that borrowers of secured consumer automobile loans in disparate credit risk bands who remain current eventually converge in default risk. We call such a phenomenon *credit risk*

convergence.[1] Specifically, we find that the conditional monthly probability of default given survival to at least that month eventually converges for borrowers from disparate risk bands after approximately 28 to 49 months on consumer automobile loans with an original term of 72–73 months. The second major finding is that borrowers originally from high-risk credit bands (deep subprime, subprime, near-prime) who remain current and thus migrate to low-risk credit bands (near-prime, prime, super prime) eventually overpay in comparison to the prevailing market risk-based cost of borrowing by amounts ranging from approximately 285 to 1,391 basis points, as measured by an annual percentage rate (APR). Both of these results are robust to a variety of sensitivity tests covering the economic shock of the Coronavirus pandemic economic shutdown of Spring 2020, alternative definitions of consumer credit risk bands based on the contracted interest rate rather than credit score, and assumptions made beyond the estimable window of conditional probabilities.

The statistical analysis is based on 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). Recent 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. We thus compile a unique data set of monthly performance from 50,107 individual 72–73 month consumer automobile loans from four completed ABS bonds (i.e., CarMax, 2017; Ally, 2017; Santander, 2017b,a). We select loans to be as comparable as possible (i.e., no co-borrowers, used vehicles only, etc.). We assign borrowers to risk bands via the borrower's application credit score by using the same classification ranges as the Consumer Financial Protection Bureau (2019). We then write software to determine the outcome of a loan from the asset-level performance data: either a complete repayment or default.[2] Finally, we empirically estimated a recovery rate by default time. For details, see Section I and for computer code, see the online supplemental

---

[1]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 same may be said for two consumers that default, except convergence will occur to a probability of default of unity. The major analysis, therefore, is determining if such convergence occurs prior to a loan outcome and, if so, its financial implications.

[2]All 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 repayments.

material. Later robustness checks consider an additional 56,683 consumer loans selected from later offerings of the same ABS issuers (i.e., CarMax, 2019; Ally, 2019; Santander, 2019b,a) and redefining the credit risk bands by contracted interest rate, see Section IV for details.

Our probabilistic model follows from the field of survival analysis or time-to-event modeling. Time-to-event modeling is a logical tool in this application because both the time-until-default and time-until-repayment affect the profitability of a loan. Because our data come from ABS pools, it is subject to incompleteness due to left-truncation and right-censoring. Furthermore, the nature of a loan contract requires us to work with discrete-time. Extended detail may be found in the theoretical study of Lautier et al. (2021) for left-truncation and the asset-pricing analysis of Lautier et al. (2022), which considers both left-truncation and right-censoring in the context of modeling a consumer lease ABS. The desire to differentiate between default and prepayment requires us to generalize the work of Lautier et al. (2022) to a competing risk model. Competing risks have been used to model loans (e.g., Banasik et al., 1999; Stepanova and Thomas, 2002; Dirick et al., 2017), but we find previous approaches insufficient to address our requirements of modeling competing risks in discrete-time adjusted for both random left-truncation and right-censoring (see Section II for additional references and discussion details). Critically, we desire to estimate the *cause-specific hazard rate* by risk band for defaults (prepayments), which is the probability of default (prepayment) in the current month given that a loan has survived the previous month.

The crux of our analysis then relies on the asymptotic properties of these hazard rate estimators, which are stated in Proposition 1 and formally proved in Appendix A, a contribution to the statistical literature in its own right. The asymptotic normality of the hazard rate estimators allow us to construct 95% confidence intervals for the true hazard rate. Hence, we can formally test for the equivalence between the hazard rate of two disparate credit risk bands by searching for the month in which the confidence intervals between two disparate risk bands overlap. As stated earlier, we conjecture that the conditional monthly probability of default given survival to at least that month (i.e., the hazard rate) eventually converges for borrowers from disparate risk bands after approximately 28 to 49 months on consumer auto loans with an original term of 72–73 months. For details of the probabilistic model, see Section II.B, and for results see Section II.C. We also find that credit risk convergence of these consumer loans cannot be fully explained by the economic shock of the Coronavirus pandemic economic shutdown of Spring 2020 and is robust to an alternative definition of

consumer credit risk bands based on the contracted interest rate rather than credit score, see Sections IV.A and IV.B, respectively, for details.

Assessing the financial implications of credit risk convergence requires deconstructing the fundamentals of risk-based pricing (Edelberg, 2006; Phillips, 2013; Livshits, 2015). Specifically, we can compartmentalize a borrower's interest rate into a cost of capital, profit margin, and an added spread that varies by risk-band. It is this added spread that we desire to recover and analyze. That is, we can utilize the empirical estimates of the probabilistic model in conjunction with actuarial techniques to derive a fair interest rate based on the expected present value of future payments. Hence, any remaining spread over the fair actuarial rate is the market-implied credit spread demanded for the uncertainty of future payments. Better still, we can plot this market-implied credit spread as a function of a loan conditional on survival, which illustrates the profitability dynamics of a current loan. Because the loan interest rate is fixed for the lifetime of the contract, the shape of the conditional market-implied credit spread over time reveals the proportion of the rate attributable to expected future return for a current loan; that is, a dwindling (growing) conditional market-implied credit spread suggests a growing (dwindling) expected profit. Not surprisingly, borrowers that pay a higher interest rate and remain current are more profitable. What is noteworthy, however, is that we can find a borrower with the same conditional default risk that is paying a much lower rate, and so consumers originally from lower quality credit risk bands (deep subprime, subprime, near-prime) that remain current and thus migrate to higher quality risk bands (near-prime, prime, super prime) and do not refinance eventually overpay in comparison to the prevailing market risk-based cost of borrowing. As stated earlier, we conjecture that this unjustifiable overpayment ranges from approximately 285 to 1,391 basis points, as measured by an annual percentage rate (APR). Much greater detail may be found in Section III. We further find these these results are robust to assumptions made beyond the estimable window of conditional probabilities, see Section IV.C for details.

We also speak briefly on the economic implications and potential remedies. Because our study focuses on secured automobile loans from within a core economic lending space by major financial institutions, such large market inefficiencies are troubling. Encouragingly, our estimates of prepayment behavior suggest that borrowers across all risk profiles increase prepayment behavior as a loan matures. This suggests that some overpaying consumers do refinance these auto loans and self-correct. On the other hand, a sizable portion continue

to make payments at a rate that is much higher than the market would dictate given an improved credit profile. In our concluding remarks, we propose a consumer loan product that is designed to lower future payments based on performance of the borrower. We speculate that such a feature may act as an incentive for borrowers to keep making payments. We also suggest competing lenders target these overpaying borrowers to lower the overall market inefficiency. Finally, we suggest possible regulatory action.

Generally, this paper shares genetic material with financial research related to the perceived machinations and true economic welfare of the consumer lending space. A natural target 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 close 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 Kőszegi (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 leads 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. Lenders knowingly hide interest rates to select for borrowers who make decisions based on payment size (Stango and Zinman, 2011). 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

7

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. We have thus focused on the borrower with poor credit who signs a contract with a major financial institution to purchase an essential economic asset: an automobile. The rate for this contract may be upwards of 20% APR to reflect a perceived high level of risk. Subsequently, this borrower makes payments on time to the point an updated assessment of credit risk in combination without refinancing would suggest this borrower eventually overpays by potentially over 1,000 basis points of APR. We thus write on behalf of these borrowers.

The paper proceeds as follows. We first introduce and detail the supporting consumer auto loan data from securitization pools in Section I. The consequential statistical analysis occurs in Section II, which includes estimating the conditional default probabilities based on survival time along with large sample confidence intervals. The financial analysis then follows in Section III, which utilizes the probabilistic estimates of Section II to calculate a conditional expected rate of return to ultimately derive the market-implied credit spread. We perform a series of robustness checks in Section IV, and Section V concludes. Appendices A and B contain proofs, and Appendix C provides a large sample simulation study for reference. For interested readers, the online supplemental material contains data details and replication code.

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. The bonds 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. The bonds were selected to span approximately the same months to ensure all underlying loans were subject to the same macroeconomic environment. By count, the total number of loans for CARMX, AART, SDART, and DRIVE were 55,000, 67,797, 80,636, and 72,515, 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.[3] 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 were as comparable as possible, we employed a number of filtering mechanisms. First, we removed any loan contracts that included a co-borrower. Second, we required each loan to have been underwritten to the level of "stated not verified", 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 removed 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 required all loans to correspond to the sale of a used vehicle. We further dropped 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 considered loans younger than 18 months as of the first available ABS reporting month. In addition, we removed any loans with a "Not Available" or `NA` for the credit score field. For loan term, we only

---

[3]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".

included loans with an original term of 72 or 73 months. As a final data integrity check, we removed any loans that did not pay enough total principle to pay-off 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 only impacted 2,338 or 4.5% of the filtered loan population. This left 50,107 individual consumer auto loan contracts, summary details of which may be found in Section I.C. Complete replication code and other data details may be found in the online supplementary material.

Next, we assigned each loan into a credit risk category or *risk band* depending on a borrower's credit score as of the initial loan application. A credit score is a numerical representation of a borrower's likelihood of repaying a loan in full and on-time. While multiple companies may compute a credit score, scores will typically range from 300-900, with a higher score representing a stronger credit. To be consistent with industry standards, we have used the same borrower risk classification categories as the Consumer Financial Protection Bureau (2019). Specifically, going from highest risk to lowest risk, 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". An examination of the third row of Figure 2 shows that, generally speaking, the borrower's interest rate increases as the risk band credit quality decreases. This suggests that, under an assumption of risk-based pricing as described in Section III, defining the risk bands by consumer credit score is a reasonable approximation of the true lender risk bands. We remark here that we also consider a robustness check of our analysis by instead assigning risk bands by interest rate, details of which may be found in Section IV.B.

I.B.   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 borrowers, the 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, which is ultimately defined as a monthly percentage of the initial loan balance.

The loan level data of Securities and Exchange Commission (2016) is detailed, and so we may utilized the associated fields to perform the estimation. Specifically, we calculated a sum total of the `recoveredAmount` field for all loans that ended in default. The `recoveredAmount` field includes any additional loan payments made by the borrower after defaulting, legal settlements, and repossession proceeds (Securities and Exchange Commission, 2016). As a next step, we took an average of this total `recoveredAmount` divided by `originalLoanAmount` at each month a default was observed. Next, for convenient use within the profitability analysis of Section III.A, the results were then nonparametrically smoothed using the `loess()` function in R (R Core Team, 2022). Finally, for the purposes of extrapolating beyond the recoverable sample space of loan age from the data, we fitted the estimated `loess` curve to a gamma-kernel via an ordinary minimization of a sum-of-squared differences. The results are in Figure 1.

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, 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 around month 18 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 most likely reflects the depreciating value of the automobile over time.

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 preferred 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
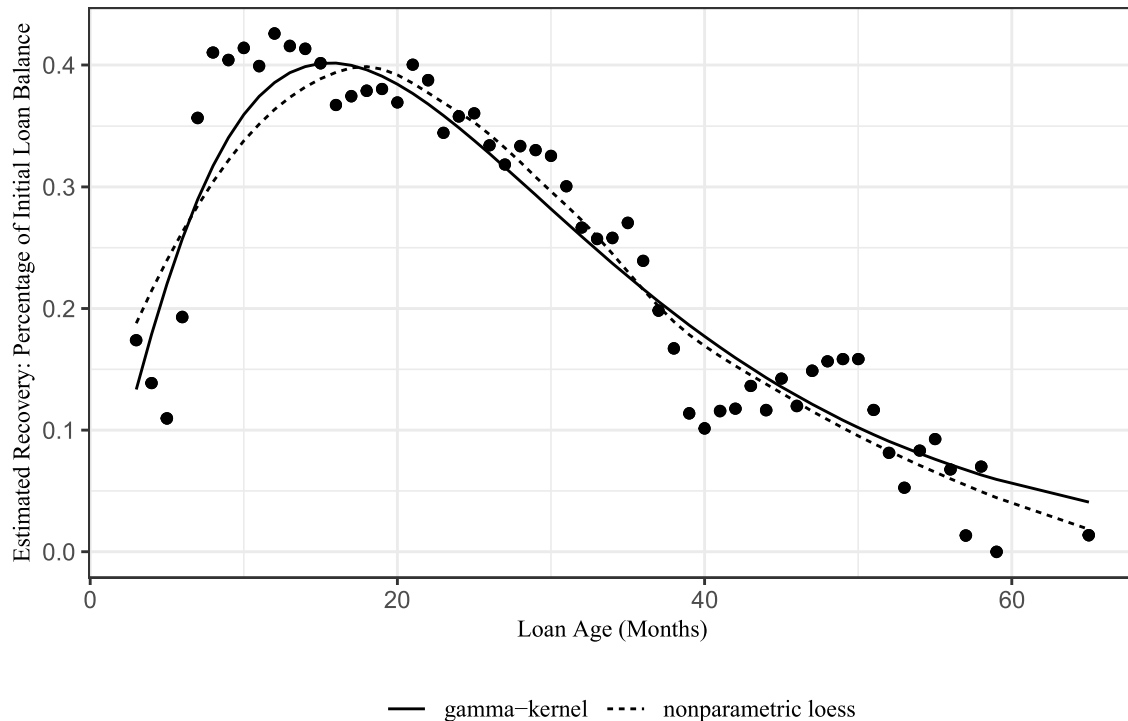
Figure 1: **Estimation of the recovery upon default assumption**. The point estimates are formed using the asset-level data of Securities and Exchange Commission (2016) for the 50,107 filtered loans summarized in Figure 2. Specifically, they are the monthly average of the sum total of the `recoveredAmount` field, which includes any additional loan payments made by the borrower after defaulting, legal settlements, and repossession proceeds (Securities and Exchange Commission, 2016), divided by the `originalLoanAmount` field for each loan that ended in default. The dashed line represents a nonparametric smoothing via the `loess` function in R Core Team (2022). The solid line is a fitted gamma-kernel using an ordinary minimization of a sum-of-squared differences, which may be used to extrapolate beyond the recoverable sample space.

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.

12

I.C. Summary of Selected Loans

After the data cleaning and filtering of Section I.A (again, further detail and replication code is available in the online supplemental material), we have payment performance for 50,107 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. We can see that generally DRIVE is a deep subprime/subprime pool of borrowers, SDART is a subprime/near-prime pool, CARMX is a near-prime/prime pool, and AART is a prime/super-prime pool of borrowers. Naturally, the interest rate percentage is higher for borrowers of lower credit, with DRIVE and SDART generally at or above a 20% annual percentage rate (APR) and declining to around a 10-15% APR for CARMX and under a 10% APR for AART.

The loans are well dispersed geographically among all 50 states and Washington, DC, 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 (12%), Chevrolet (9%), Ford (7%), Toyota (7%), and Hyundai (7%). Notably, none of the parent companies of the four bonds represent financing subsidiaries of auto manufacturers. 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 50,107 loans is weighted towards deep subprime and subprime borrowers, which are 44% and 24% of the total, respectively. Similarly, DRIVE and SDART supply over 80% of the total loans in our sample. The smallest risk band is super-prime, which totals 3,636 loans for 7% of all loans. This suggests a robust sample for estimating a probabilistic model for each risk band.

In terms of loan performance, we can observe some clear trends in Table 1. First, nearly half of all deep subprime risk band loans defaulted, and this percentage declines by risk band until super-prime, in which only 7% of loans defaulted during the observation window. We also see varying performance by bond. For example, while 51% of deep subprime loans defaulted in the DRIVE pool, only 24% of deep subprime loans in the CARMX pool defaulted. Overall, within each risk band, DRIVE loans perform the worst, followed by SDART, AART, and finishing with the best performance of CARMX auto loans. Within
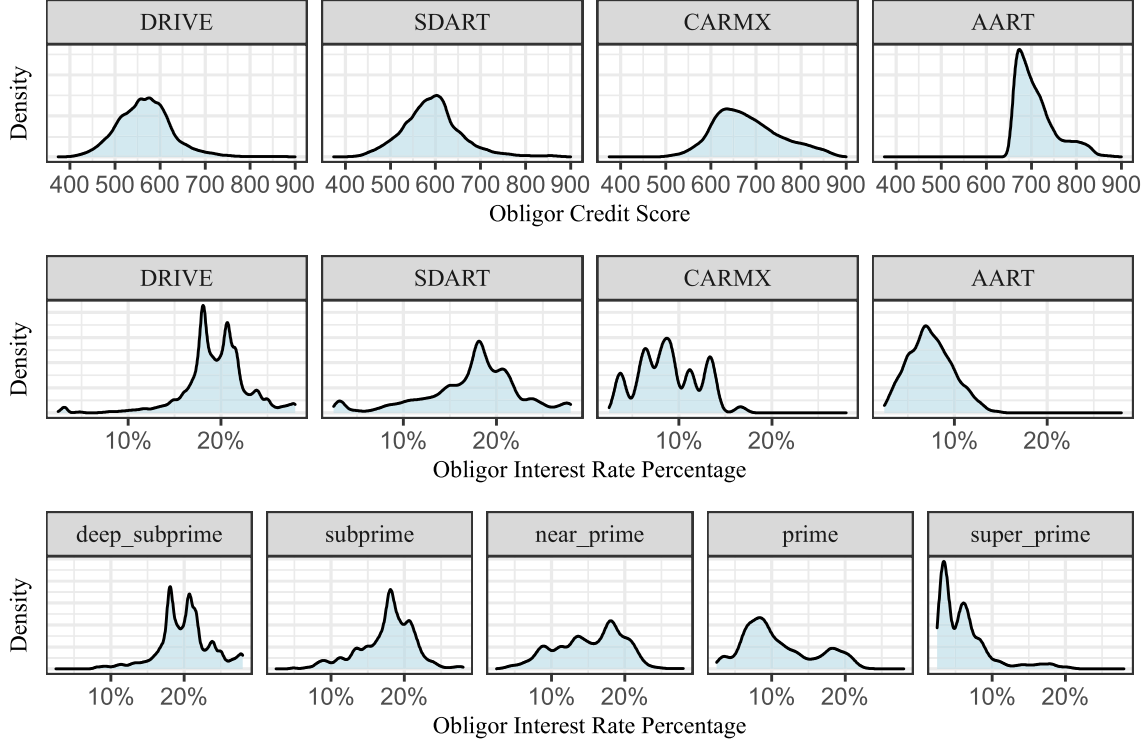
Figure 2: **Borrower credit profile and annual interest rate by bond, risk band**. A summary of the borrower credit profiles (top row) and charged APR (middle row) of the 50,107 filtered consumer automobile loans used in the analysis of Sections II and III by ABS bonds CarMax Auto Owner Trust 2017-2 (CarMax, 2017) (CARMX, 6,834), Ally Auto Receivables Trust 2017-3 (Ally, 2017) (AART, 2,171), Santander Drive Auto Receivables Trust 2017-2 (Santander, 2017b) (SDART, 16,246), and Drive Auto Receivables Trust 2017-1 (Santander, 2017a) (DRIVE, 24,856). The bottom row shows the distribution of interest rates by risk band for the same set of 50,107 loans.

each bond, however, we see that the percentage of defaulted loans declines as the credit quality of the risk band increases. This again suggests that our credit-score based risk band approximation has largely captured the main component of the different parent company's underwriting practices.

Table 1: **Borrower Counts by Risk Band, Bond, and Loan Outcome**. Summary statistics and loan outcomes of the 50,107 filtered consumer automobile loans summarized in Figure 2.

		deep subprime	subprime	near-prime	prime	super-prime	Total
	Total	22,093 (44%)	11,853 (24%)	6,486 (13%)	6,039 (12%)	3,636 (7%)	50,107 (100%)
	DRIVE	14,904 (67%)	6,095 (51%)	2,293 (35%)	1,172 (19%)	392 (11%)	24,856 (50%)
	SDART	6,804 (31%)	4,740 (40%)	2,602 (40%)	1,451 (24%)	649 (18%)	16,246 (32%)
	CARMX	385 (2%)	1,018 (9%)	1,591 (25%)	1,953 (32%)	1,887 (52%)	6,834 (14%)
	AART	0 (0%)	0 (0%)	0 (0%)	1,463 (24%)	708 (19%)	2,171 (4%)
	Total	22,093 (100%)	11,853 (100%)	6,486 (100%)	6,039 (100%)	3,636 (100%)	50,107 (100%)
	Defaulted	10,655 (48%)	4,405 (37%)	1,793 (28%)	940 (16%)	246 (7%)	18,039 (36%)
	Censored	4,352 (20%)	2,731 (23%)	1,777 (27%)	1,952 (32%)	1,435 (39%)	12,247 (24%)
	Repaid	7,086 (32%)	4,717 (40%)	2,916 (45%)	3,147 (52%)	1,955 (54%)	19,821 (40%)
	Total	22,093 (100%)	11,853 (100%)	6,486 (100%)	6,039 (100%)	3,636 (100%)	50,107 (100%)
DRIVE	Defaulted	7,651 (51%)	2,567 (42%)	819 (36%)	307 (26%)	55 (14%)	11,399 (46%)
DRIVE	Censored	2,689 (18%)	1,196 (20%)	445 (19%)	213 (18%)	115 (29%)	4,658 (19%)
DRIVE	Repaid	4,564 (31%)	2,332 (38%)	1,029 (45%)	652 (56%)	222 (57%)	8,779 (35%)
DRIVE	Total	14,904 (100%)	6,095 (100%)	2,293 (100%)	1,172 (100%)	392 (100%)	24,856 (100%)
SDART	Defaulted	2,912 (43%)	1,639 (35%)	710 (27%)	271 (19%)	56 (9%)	5,588 (34%)
SDART	Censored	1,496 (22%)	1,095 (23%)	638 (25%)	361 (25%)	227 (35%)	3,817 (23%)
SDART	Repaid	2,396 (35%)	2,006 (42%)	1,254 (48%)	819 (56%)	366 (56%)	6,841 (42%)
SDART	Total	6,804 (100%)	4,740 (100%)	2,602 (100%)	1,451 (100%)	649 (100%)	16,246 (100%)
CARMX	Defaulted	92 (24%)	199 (20%)	264 (17%)	162 (8%)	76 (4%)	793 (12%)
CARMX	Censored	167 (43%)	440 (43%)	694 (44%)	891 (46%)	833 (44%)	3,025 (44%)
CARMX	Repaid	126 (33%)	379 (37%)	633 (40%)	900 (46%)	978 (52%)	3,016 (44%)
CARMX	Total	385 (100%)	1,018 (100%)	1,591 (100%)	1,953 (100%)	1,887 (100%)	6,834 (100%)
AART	Defaulted	0	0	0	200 (14%)	59 (8%)	259 (12%)
AART	Censored	0	0	0	487 (33%)	260 (37%)	747 (34%)
AART	Repaid	0	0	0	776 (53%)	389 (55%)	1,165 (54%)
AART	Total	0	0	0	1,463 (100%)	708 (100%)	2,171 (100%)

II. The Probabilistic Model

From the perspective of a lender, it is important to distinguish between a loan that defaults shortly after it is originated versus a loan that defaults after a much longer period of time. The reason for this is natural; a loan that makes more payments before defaulting will likely be more profitable. The same may be said for prepayments. Thus, it is essential to derive a time-to-event distribution to effectively model the expected present value of a potential consumer loan. From a statistical perspective, modeling a random time-to-event falls under

the branch of *survival analysis*. As a nuanced but important point of emphasis, our data is from pools of consumer automobile loans found in publicly traded asset-backed securities (see Section I for details). Thus, we must consider a model and subsequent estimator capable of working in discrete-time and with incomplete data subject to random left-truncation and random right-censoring. For details on this point, see the discrete-time work of Lautier et al. (2021) for the case of left-truncation and Lautier et al. (2022) for the case of both left-truncation and right-censoring in the context of modeling a consumer lease asset-backed security. The finite term of a consumer loan contract allows us to anchor our model on a finite time horizon.

In addition to modeling a time-to-event random variable, we also desire to distinguish between the type of event. Again, from the perspective of a lender, this is natural; a loan that is repaid (or prepaid) in a given month is much more profitable than a loan that defaults in the same month. We therefore desire to differentiate between loans ending in default and loans ending in prepayment. To do so, we can define the problem in terms of a *competing risk* 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. (2022) 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 meets our setting of two competing states in discrete-time over a finite time horizon adjusted for both random left-truncation and right-censoring.

II.A.   *Defining the Cause-Specific Hazard Rate Estimator*

We will use the notation of Lautier et al. (2022) and construct the probabilistic model within the context of an automobile loan ABS. Specifically, define the random time until a loan contract ends by the random variable X. 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 only observe X if

$X \geq Y$.

Since we require that X and Y are independent, it is worthwhile to justify this assumption. In a securitization context, 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.

Further, define the censoring random variable as $C = Y + \tau$, where τ 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. 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 $\Delta \geq 0$, $\Delta \in \mathbb{N}$, where $\mathbb{N}$ denotes the nonnegative integers, and a finite maximum end point, which we denote by $\xi \geq \Delta + \tau$, $\xi \in \mathbb{N}$. We emphasize the word *recoverable*, further discussion of which may be found in Lautier et al. (2021) and Lautier et al. (2022).

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. Formally,

$$\lambda(x) = \Pr(X = x \mid X \geq x) = \frac{\Pr(X = x)}{\Pr(X \geq x)}. \tag{1}$$

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

$$1 - F(x-) = \Pr(X \geq x) = \prod_{\Delta+1 \leq k < x} \{1 - \lambda(k)\}, \tag{2}$$

with the convention $\prod_{k=\Delta+1}^{\Delta}\{1 - \lambda(k)\} = 1$.

We generalize Lautier et al. (2022) to the case of two competing risks as follows. First, consider two competing risks as a multistate process, such as in Section 3 of Beyersmann et al. (2009). Formally, let $(Z_x)_{x \geq 1}$ be a discrete-time competing risk process. Given a loan terminates at time x, we assume the loan must be in one of two states, $Z_x \in \{1, 2\}$:

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 $\lambda^{01}(x)$.

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 $\lambda^{02}(x)$.

It may be of help to see the related Beyersmann et al. (2009, Figure 1). Given this framework, it is not difficult to introduce right-censoring and left-truncation along the lines of Lautier et al. (2022). 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 $i = 1, 2$. The discrete-time cause-specific hazard rate is defined as

$$\lambda_\tau^{0i}(x) = \Pr(X = x, Z_x = i \mid X \geq x) = \frac{\Pr(X = x, Z_x = i)}{\Pr(X \geq x)}, \quad i = 1, 2. \tag{3}$$

We include a τ subscript to remind us that right-censoring is present in the data. Conveniently, therefore, from the law of total probability, we have

$$\lambda_\tau(x) = \frac{\Pr(X = x)}{\Pr(X \geq x)} = \frac{\Pr(X = x, Z_x = 1)}{\Pr(X \geq x)} + \frac{\Pr(X = x, Z_x = 2)}{\Pr(X \geq x)} = \lambda_\tau^{01}(x) + \lambda_\tau^{02}(x).$$

Within a competing risk framework, $\lambda_\tau(x)$ may be referred to as the *all-cause hazard.* For

some readers, it may be illuminating to review Table 5 in the simulation study of Appendix C for a numeric example of our competing risk model. If we assume independence between Y_j and the random vector (X_j, Z_{X_j}) (not at all unreasonable given the securitzation backdrop), then we may similarly derive estimators for (3) that follow the set-up of Lautier et al. (2022). We demonstrate as follows. Let $\alpha = \Pr(Y \leq X)$ and for $i = 1, 2$, define

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

and

$$
\begin{aligned}
C_\tau(x) &= \Pr(Y_j \leq x \leq \min(X_j, C_j)) \\
&= \Pr(Y \leq x \leq \min(X, C) \mid X \geq Y) \\
&= (1/\alpha)\{\Pr(Y \leq x \leq C)\Pr(X \geq x)\}.
\end{aligned}
$$

Thus,

$$
\lambda_\tau^{0i}(x) = \frac{\Pr(X = x, Z_x = i)}{\Pr(X \geq x)} = \frac{f_{*,\tau}^{0i}(x)}{C_\tau(x)}. \tag{4}
$$

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 $\mathbf{1}_{X_i \leq C_i}$, where $\mathbf{1}_Q = 1$ 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

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

and

$$
\hat{C}_{\tau,n}(x) = \frac{1}{n}\sum_{j=1}^{n}\mathbf{1}_{Y_j \leq x \leq \min(X_j, C_j)},
$$

we obtain the estimate for (4)

$$\hat{\lambda}_{\tau,n}^{0i}(x) = \frac{\hat{f}_{*,\tau,n}^{0i}(x)}{\hat{C}_{\tau,n}(x)} = \frac{\sum_{j=1}^{n} \mathbf{1}_{X_j \leq C_j} \mathbf{1}_{Z_{X_j}=i} \mathbf{1}_{\min(X_j,C_j)=x}}{\sum_{j=1}^{n} \mathbf{1}_{Y_j \leq x \leq \min(X_j,C_j)}}. \tag{5}$$

II.B.   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 Appendix A.

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

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

Then,

(i)

$$\hat{\mathbf{\Lambda}}_{\tau,n}^{0i} \xrightarrow{\mathcal{P}} \mathbf{\Lambda}_\tau^{0i}, \ \ as \ n \to \infty;$$

(ii)

$$\sqrt{n}(\hat{\mathbf{\Lambda}}_{\tau,n}^{0i} - \mathbf{\Lambda}_\tau^{0i}) \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{\Sigma}^{0i}), \ \ as \ n \to \infty,$$

where $\mathbf{\Lambda}_\tau^{0i} = (\lambda_\tau^{0i}(\Delta + 1), \ldots, \lambda_\tau^{0i}(\xi))^\top$ *with* $\lambda_{\tau,n}^{0i}(x) = f_{*,\tau}^{0i}(x)/C_\tau(x)$ *and*

$$\mathbf{\Sigma}^{0i} = \mathrm{diag}\left(\frac{f_{*,\tau}^{0i}(\Delta + 1)\{C_\tau(\Delta + 1) - f_{*,\tau}^{0i}(\Delta + 1)\}}{C_\tau(\Delta + 1)^3}, \ldots, \frac{f_{*,\tau}^{0i}(\xi)\{C_\tau(\xi) - f_{*,\tau}^{0i}(\xi)\}}{C_\tau(\xi)^3} \right).$$

That is, the cause-specific hazard rate estimators $\hat{\lambda}_{\tau,n}^{0i}(\Delta + 1), \ldots, \hat{\lambda}_{\tau,n}^{0i}(\xi)$ *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 ($\lambda_\tau^{0i}(x)$ $(1-\theta)\%$ *Confidence Interval*). *The* $(1-\theta)\%$ *asymptotic confidence interval bounded within* $(0,1)$ *for* $\lambda_\tau^{0i}(x)$, $x \in \{\Delta+1,\ldots,\xi\}$, $i = 1,2$ *is*

$$\exp\left\{\ln\hat{\lambda}_{\tau,n}^{0i}(x) \pm \mathcal{Z}_{(1-\theta/2)}\sqrt{\frac{\hat{C}_{\tau,n}(x) - \hat{f}_{*,\tau,n}^{0i}(x)}{n\hat{C}_{\tau,n}(x)\hat{f}_{*,\tau,n}^{0i}(x)}}\right\}, \tag{6}$$

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

II.C.   Cause-Specific Hazard Rate Estimates by Risk Band

We now apply the methods of Sections II.A and II.B to the consumer loan data of Section I. Recall again the cause-specific hazard rate of (3). To make the economic connection between loan default risk over time and the cause-specific hazard for default (cause 01), we will emphasize the probability that the cause-specific hazard rate represents. Suppose the current age of a loan is x months, where $1 \le x \le 72$. Then the quantity $\lambda_\tau^{01}(x)$ 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.

We plot the matrix of cause-specific hazard rate estimates for default in Figure 3. There are a few notable observations. First, we generally see that the monthly conditional default rate declines as the credit quality of the risk band improves, as expected. There is also a large spike in the hazard rate for the deep subprime, subprime, and near prime risk bands. 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). 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 cause-specific hazard

rate has been effective. It is notable 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 from month 4 through month 54 only; thus, $4 \leq X \leq 54$ is the estimable window.

In addition to the cause-specific hazard rate estimators, we also plot the 95% asymptotic confidence intervals using (6). Because of the asymptotic properties of the estimator, we cannot claim that the cause-specific hazard rate estimate of two different risk bands is different on a statistically significant basis whenever the confidence intervals of two different risk bands overlap. In other words, 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 would fail to reject the following null hypothesis, H_0,

$$H_0 : \lambda^{0i}_{\tau,(a)}(x) = \lambda^{0i}_{\tau,(a')}, \tag{7}$$

at the 5% level of significance for an age x loan and $i = 1, 2$. Hence, for every comparison of two risk bands, we can search for the first month that the conditional monthly default risk converges by finding the first month that the cause-specific hazard rate confidence intervals overlap. For example, the deep subprime risk band conditional monthly default rate appears to converge with the subprime, near-prime, and prime risk bands, but it does not appear to converge with the super-prime risk band. To interpret this economically, let's again compare a deep subprime and near-prime borrower. We see the cause-specific hazard rate for default estimate's confidence intervals overlap beginning around three years (loan age 36 months). This would suggest that there is no statistically significant difference between the conditional monthly default rate between these two risk bands for any loans that have remained active and paying for three years. Since the deep subprime borrower is likely paying a higher interest rate, however, it indicates that the lender is enjoying an increased profit for these borrowers and the borrower, by not refinancing, is effectively overpaying and thus not acting as a rational economic agent. In the upcoming Section III, we attempt to measure and quantify these changing profit dynamics over time.

We will refer to the point of overlap of the asymptotic confidence intervals for the cause-specific hazard rate estimate for default of two different risk bands as the point of *credit*

Table 2: **Estimated Month of Credit Risk Convergence**. A summary matrix of the first month after ten months that the asymptotic confidence intervals in Figure 3 overlap for three consecutive months by risk band. This matrix can be interpreted as the estimated month of consumer *credit risk convergence* between disparate risk bands of 72–73 month automobile loans. Estimates are only available for loan ages $4 \leq X \leq 54$.

	deep subprime	subprime	near-prime	prime	super-prime
deep subprime	10	23	34	46	NA
subprime		10	11	42	49
near-prime			10	28	42
prime				10	10
super-prime					10

risk convergence. In Table 2, we present a matrix of the first month after ten months that the asymptotic confidence intervals overlap for three consecutive months by risk band. Under these conditions, we find a point of consumer loan risk convergence prior to the last observation month (loan age 54 months) for all risk bands except for the deep subprime and super-prime risk bands. The overall implication of Figure 3 and Table 2 is fairly clear: regardless of the initial risk category of a borrower, the more payments made on-time, the less likely it is to observe a default. Since the contracted interest rate is invariant to these credit changes, more mature loans will likely be more profitable for the lender on a risk-adjusted return basis. More detail on estimating the changing profit dynamics may be found in Section III.

Figure 4 plots the cause-specific hazard rate for prepayments (cause 02) by loan age. It is interesting to observe that prepayment behavior is generally consistent across risk bands. Further, we see that the conditional probability of prepayment increases as a loan matures. This suggests that some borrowers may realize that the offered interest rate is no longer an accurate reflection of their credit risk and self-correct by refinancing. On the other hand, borrowers may simply having an increasing desire to update their vehicle as the loan matures.

III. Financial Implications

Risk-based pricing is the standard in pricing consumer credit (Edelberg, 2006), and it has been attributed to lowering the cost of credit for a majority of borrowers and expanding
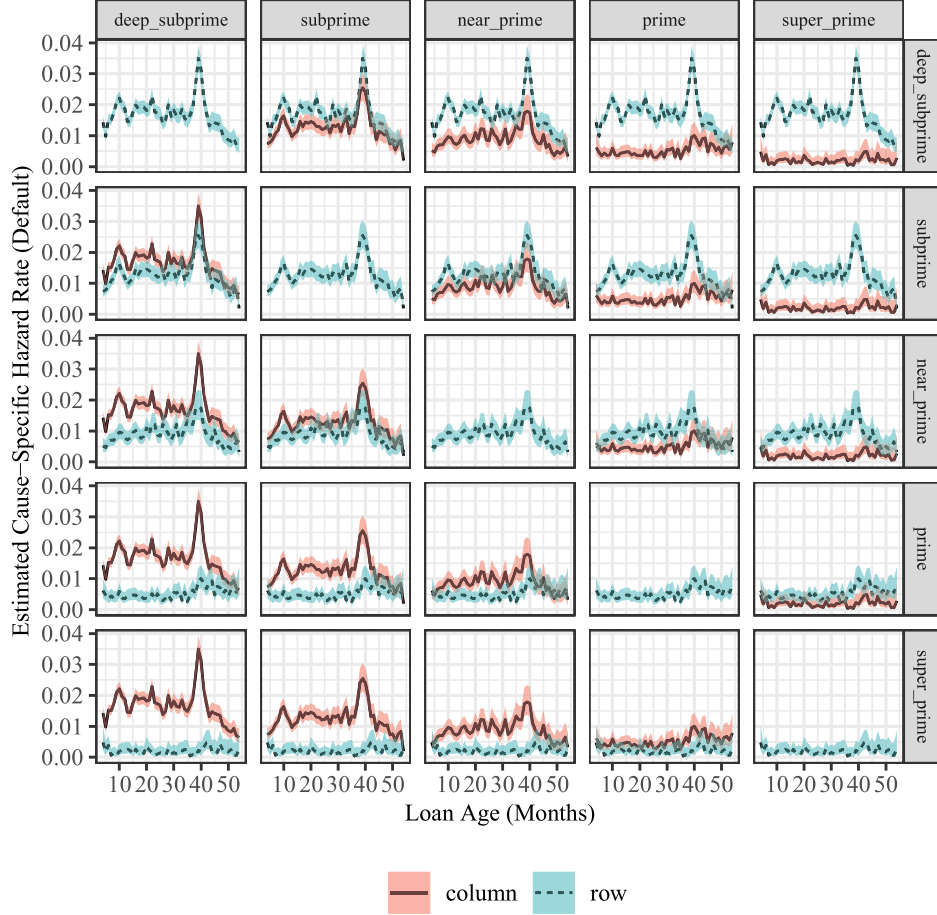
Figure 3: **Estimated cause-specific hazard rate for default risk-band comparison**. A plot of $\hat{\lambda}^{01}_{\tau,n}$ (defaults) by loan age and risk band plus 95% confidence intervals using (6). If the confidence intervals overlap for two disparate risk bands, we fail to reject the null hypothesis of equivalence between the cause-specific hazard rate for default between the risk bands stated in (7).

credit availability to higher risk borrowers (Staten, 2015). Further, models have developed in sophistication such that elements of an auto loan contract can be used to filter prospective borrowers (Einav et al., 2012). For readers seeking additional detail, Livshits (2015) presents a useful survey.

Generally, the process entails using the application details of a prospective borrower to assign said borrower into a risk band. Following Phillips (2013), the interest rate for risk
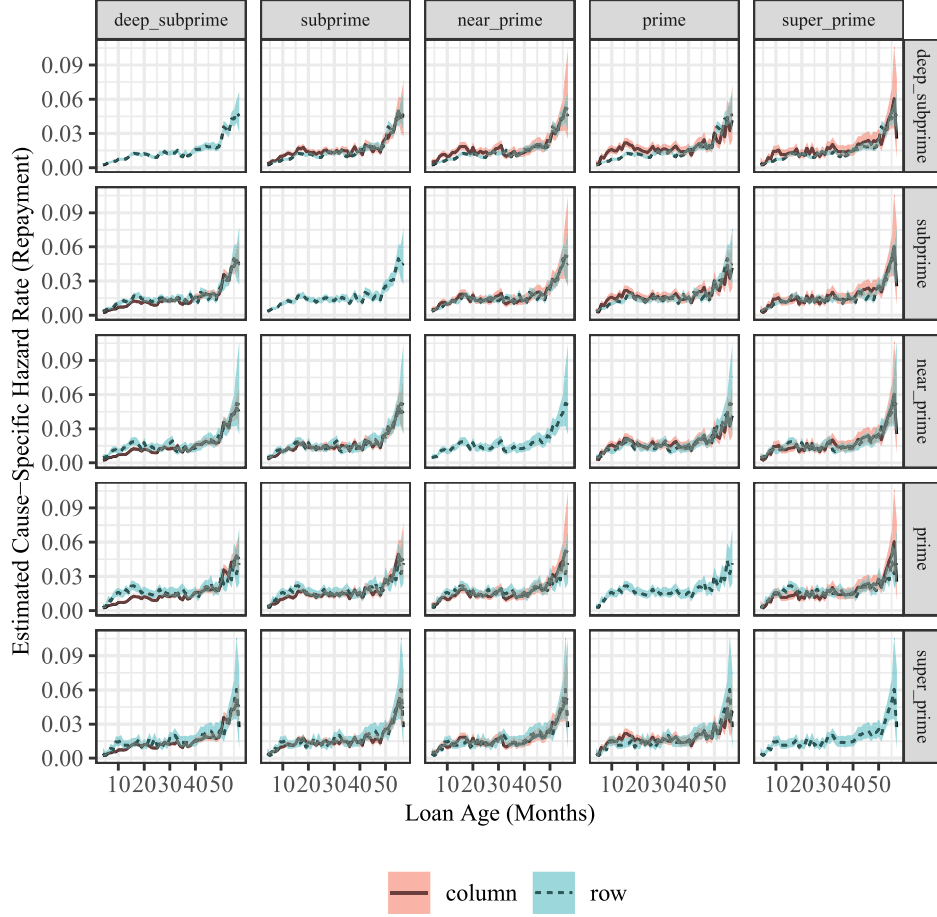
Figure 4: **Estimated cause-specific hazard rate for prepayment risk-band comparison.** A plot of $\hat{\lambda}_{\tau,n}^{02}$ (prepayments) by loan age and risk band plus 95% confidence intervals using (6). If the confidence intervals overlap for two disparate risk bands, we fail to reject the null hypothesis of equivalence between the cause-specific hazard rate for prepayment between the risk bands stated in (7).

band a, r_a, can be divided into three components

$$r_a = r_c + m + l_a, \tag{8}$$

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. Typically, l_a is used to compensate for the expected loss rate and associated risk for borrowers in band a. In other words, we can view l_a as the all-in credit spread lenders

demand to lend to a prospective borrower in risk band a. To see this, consider the combined term $r_c + m$ in (8) as the rate assigned to a borrower in risk band a, assuming no uncertainty in the future loan payments. Hence, by the Law of One Price, $r_c + m$ is identical for all risk bands.

Using an actuarial approach, however, we can breakdown (8) further. There is a component of l_a that is attributable to the expected present value of a loan in risk band a and a second component attributable to the riskiness (i.e., randomness) of the loan payment cash flows. Let e_a denote the component of l_a representing the expected present value of future payments of a loan in risk band a, and π_a denote the component of l_a to compensate for the uncertainty of future loan payments. Then,

$$r_a = r_c + m + e_a + \pi_a. \tag{9}$$

If we consider the initial loan balance as the price of the loan payment cash flows, then we may interpret (9) as follows. The quantity $r_c + m + e_a$ represents the "fair" expected or actuarial return of the future loan payments in that the expected future payments discounted at the rate $r_c + m + e_a$ equals the initial loan balance. In other words, an investor that is indifferent to cash flow uncertainty would accept the rate $r_c + m + e_a$ as a fair price. Most investors are not indifferent to cash flow uncertainty, however, and so any remaining positive difference between r_a and $r_c + m + e_a$ is π_a, which may be interpreted as the market-implied credit spread demanded by lenders for borrowers in risk band a. Thus, if $r_c + m + e_a$ is a small portion of r_a, then borrowers in risk band a are expected to produce limited future cash flows and the market-implied credit spread is large to reflect this high default probability. On the other hand, if $r_c + m + e_a$ is a large component of r_a, then borrowers in risk band a are expected to produce sizable future cash flows, and so the market-implied credit spread is low to reflect this lower default probability. Our next objective, therefore, is to estimate $r_c + m + e_a$.

III.A. Estimating the Conditional Expected Rate of Return

Define the expected rate of return for a loan in risk band a as

$$\rho_a = r_c + m + e_a, \tag{10}$$

and so (9) becomes $r_a = \rho_a + \pi_a$. Given reliable estimates of borrower default and prepayment probabilities, we may estimate (10) for a given loan in risk band a. In particular, we may estimate ρ_a for each month a loan is still active and paying to find a *conditional expected rate of return*. We remark here there is a pleasing property of ρ_a in the event the future loan payments will proceed as scheduled with no uncertainty, which we state formally in Proposition 2. For proof, see Appendix B.

PROPOSITION 2 (*The Conditional Expected 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 ξ months. Let $\rho_{a|x}$ denote the expected 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 $x \in \{1, \ldots, \xi\}$.*

We now formalize the estimation of $\rho_{a|x}$. For convenience of notation, we will drop a to denote the arbitrary risk band and assume the proceeding calculations will be performed entirely within one risk band. Using our loan schedule notation from Section II.A, the first possible month of an event is $\Delta + 1$, and the last possible month of an event is ξ. Denote the current age of a loan by x, $\Delta + 1 \leq x \leq \xi$. Let the cause-specific hazard rate for default at time x be denoted by $\lambda^{01}(x)$ and the cause-specific hazard rate for repayment at time x be denoted by $\lambda^{02}(x)$. Assuming no other causes for a loan termination, the all-cause hazard rate is then $\lambda(x) = \lambda^{01}(x) + \lambda^{02}(x)$. Further, recall (2) and observe for $i = 1, 2$, $x \leq j \leq \xi$,

$$
\begin{aligned}
\Pr(X = j, Z_x = i) &= \frac{\Pr(X = j, Z_x = i)}{\Pr(X \geq x)} \Pr(X \geq x) \\
&= \Pr(X = j, Z_x = i \mid X \geq x) \Pr(X \geq x) \\
&= \lambda^{0i}(j) \prod_{k=x}^{j-1} \{1 - \lambda(k)\},
\end{aligned}
$$

again with the convention $\prod_{k=x}^{x-1} \{1 - \lambda(k)\} = 1$. For convenience, denote $p_j^{0i} = \Pr(X = j, Z_x = i)$ for $i = 1, 2$, $x \leq j \leq \xi$. By definition,

$$
\sum_{j=x}^{\xi} \sum_{i=1}^{2} \hat{p}_j^{0i} = 1.
$$

Again, it may be of help to review the numeric example of Table 5 in Appendix C.

We may derive ρ_x as follows. Let the scheduled amortization loan balance of a consumer auto loan at month x, $\Delta + 1 \leq x \leq \xi$ be denoted by B_x, where $B_\xi = 0$. Denote the scheduled monthly payment by P. If we denote the recovery of a defaulted consumer auto loan at month x, $\Delta + 1 \leq x \leq \xi$ by R_x, then the default matrix at loan age $x \leq \xi - 1$ for the possible future default paths is

$$
\mathbf{D}_{(\xi-x+1)\times(\xi-x+1)} = \begin{bmatrix} R_x & 0 & 0 & \ldots & 0 & 0 \\ P & R_{x+1} & 0 & \ldots & 0 & 0 \\ P & P & R_{x+2} & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ P & P & P & \ldots & R_{\xi-1} & 0 \\ P & P & P & \ldots & P & R_\xi \end{bmatrix}.
$$

Note that row 1 of $\mathbf{D}$ would be the cash flows assuming a default at loan age x, which occurs with probability p_x^{01}. Similarly, row 2 of $\mathbf{D}$ would be the cash flows assuming a default at loan age $x + 1$, which occurs with estimated probability p_{x+1}^{01}, and so on and so forth. In the same way, we can define the repayment matrix at loan age $x \leq \xi - 1$ as

$$
\mathbf{R}_{(\xi-x+1)\times(\xi-x+1)} = \begin{bmatrix} B_x + P & 0 & 0 & \ldots & 0 & 0 \\ P & B_{x+1} + P & 0 & \ldots & 0 & 0 \\ P & P & B_{x+2} + P & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ P & P & P & \ldots & B_{\xi-1} + P & 0 \\ P & P & P & \ldots & P & P \end{bmatrix}.
$$

As with defaults, row 1 of $\mathbf{R}$ would be the cash flows assuming a prepayment at loan age x, which occurs with estimated probability p_x^{02}. Similarly, row 2 of $\mathbf{R}$ would be the cash flows assuming a prepayment at loan age $x + 1$, which occurs with estimated probability p_{x+1}^{02}, and so on and so forth. Therefore, if we denote the discount vector assuming the unknown monthly rate of ρ_x as

$$
(\boldsymbol{\nu})_{\xi-x+1}^\top = \left((1+\rho_x)^{-1} \quad (1+\rho_x)^{-2} \quad \ldots \quad (1+\rho_x)^{-(\xi-x+1)} \right)^\top,
$$

and the cause-specific probability vector as

$$\left(\boldsymbol{p}^{0i}\right)^{\top}_{\xi-x+1} = \left(\begin{matrix} p^{0i}_x & p^{0i}_{x+1} & \cdots & p^{0i}_{\xi-x+1} \end{matrix}\right)^{\top},$$

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

$$\text{EPV}_x = \left(\boldsymbol{p}^{01}\right)^{\top}_{\xi-x+1}\left[\mathbf{D}_{(\xi-x+1)\times(\xi-x+1)}(\boldsymbol{\nu})_{\xi-x+1}\right] + \left(\boldsymbol{p}^{02}\right)^{\top}_{\xi-x+1}\left[\mathbf{R}_{(\xi-x+1)\times(\xi-x+1)}(\boldsymbol{\nu})_{\xi-x+1}\right].$$

Therefore, ρ_x is the interest rate such that $B_x = \text{EPV}_x$; that is,

$$\{\rho_x : B_x = \text{EPV}_x\}. \tag{11}$$

In words, ρ_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. Note that $\rho_x \leq r$ for a given contract, with equality only in the circumstances of Proposition 2.

Finally, we of course do note know the true distribution of X. We do have the estimators in (5), however, and Proposition 1. Thus, we may estimate ρ_x by replacing the cause-specific hazard rates λ^{0i} with the estimate in (5). We close this section with the following lemma.

LEMMA 2 ($\hat{\rho}_{n,x}$ *Asymptotic Properties*). *Replace the cause-specific hazard rates in* (11) *with the estimators from* (5). *Define the estimated expected rate of return given a loan has survived to month x as $\hat{\rho}_{n,x}$. Then,*

$$\hat{\rho}_{n,x} \xrightarrow{\mathcal{P}} \rho_x, \ as \ n \to \infty.$$

III.B.   Interpreting the Market-Implied Credit Spread Over Time

Recall again (9). In light of the estimator from Lemma 2 and the constant r_a for a given risk band, we can assess the portion of r_a that is attributable to the actuarial "fair" discounted expected cash flows (i.e., $\hat{\rho}_{n,a|x}$) and the portion attributable to the market-implied credit spread (i.e., $\hat{\pi}_{n,a|x}$), as a function of loan age, x. Further, in an efficient market, r_c will be shared by all lenders and borrowers alike, and it may be approximated by the prevailing duration-matched market rate cost of capital at the time the loan was underwritten. More succinctly, r_c is also locked-in at contract pricing, as the lender could in theory borrow-to-

lend. Therefore, the growth or withering of $\hat{\rho}_{n,a|x}$ in x gives a useful approximation of the expected total margin at time x, $\tilde{m}_{a|x} = [m + e_a]_x$, by (10) for a given risk band. In other words, we can directly estimate the changing profitability dynamics of a consumer loan over time.

We close by considering three cases for illustrative purposes. Suppose first that the portions of r_a, as represented by $\hat{\rho}_{n,a|x}$ and $\hat{\pi}_{n,a|x}$ are relatively stable for $x \in \{1, \ldots, \xi - 1\}$. In this case, the risk dynamics are stable throughout the lifetime of a loan, which implies a consistent $\tilde{m}_{a|x}$ for a loan in risk band a. That is, a lender would be largely indifferent between lending in the early or later portion of the loan, on a relative risk-adjusted return basis. On the other hand, if $\hat{\pi}_{n,a|x}$ declines over the life of the loan, this would imply that $\tilde{m}_{a|x}$ is increasing in loan age, x. In this case, such a loan would become relatively more profitable to the lender over time. Hence, a borrower should seek to refinance the loan, as the static rate in risk band a does not actively reflect the declining cash flow uncertainty. In the last case we illustrate, suppose $\hat{\pi}_{n,a|x}$ increases over the life of the loan. This would imply that $\tilde{m}_{a|x}$ is decreasing in loan age, x. In this case, such a loan would become relatively less profitable to the lender over time, as the static rate in risk band a does not actively reflect the increasing uncertainty.

III.C. An Empirical Estimate of Market Inefficiency

We apply the methods of Sections III.A to the data of Section I. Figure 5 plots the estimated market-implied credit spread, $\hat{\pi}_{a|x}$, for a given risk band comparison and the 95% empirical confidence intervals. That is, a would represent one of the risk bands deep subprime, subprime, near-prime, prime, or super-prime. We also remark here that for loan ages beyond the estimable window (i.e., loan ages beyond month 54), we were required to make an assumption for the cause-specific hazard rates for default and prepayment because of the censoring described in Section II. It is standard practice in survival analysis to assume a constant hazard rate (i.e., an exponential tail in the continuous case or a geometric tail in the discrete case) for the remaining lifetime ages beyond the the final estimable hazard rate that equals the final estimable rate (Klugman et al., 2012, Section 12.1). Given the consumer credit risk convergence results observed in Figure 3 and Table 2, we assumed that the hazard rate of the two risk bands extended beyond loan age 54 would be equal to a simple average
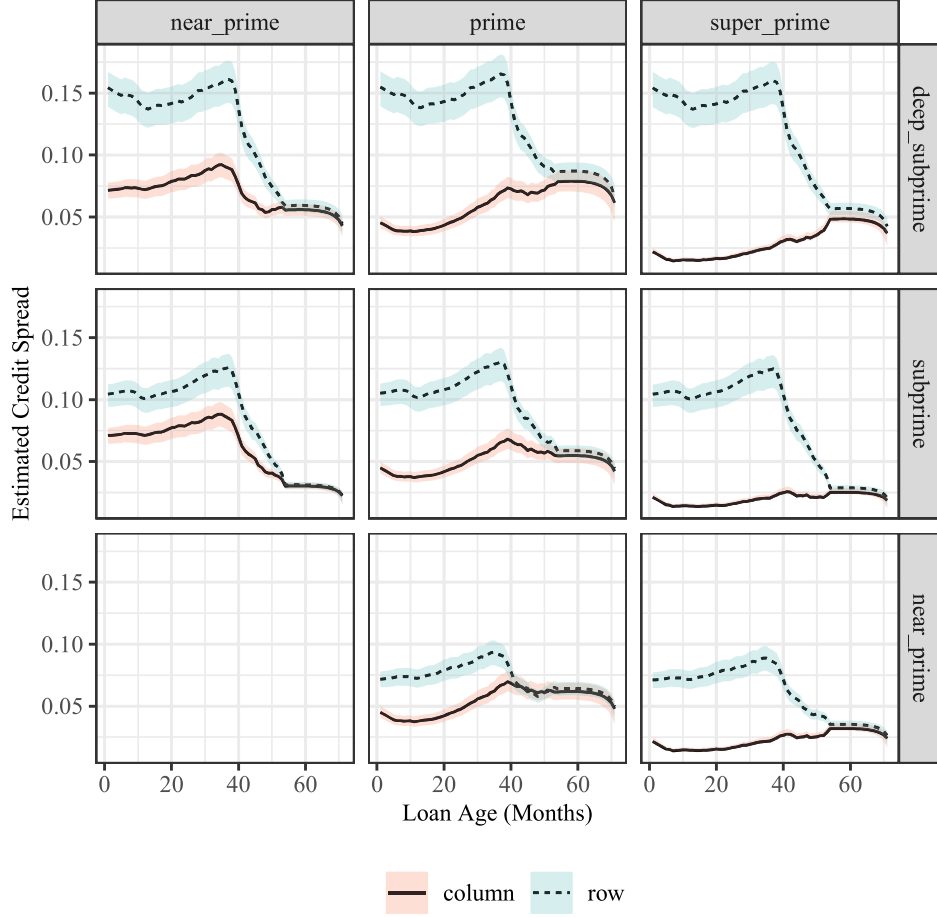
Figure 5: **Estimated market-implied credit spread given survival**. A plot of $\hat{\pi}_{n,a|x}$ by loan age and risk band plus 95% empirical confidence intervals. If credit risk is (increasing) constant ((decreasing)) conditional on survival over the life of the loan for risk band a, $\hat{\pi}_{n,a|x}$ will (increase) remain flat ((decrease)) over loan age, x.

of $\hat{\lambda}^{0i}_{\tau,n(a)}(54)$ and $\hat{\lambda}^{0i}_{\tau,n(a')}(54)$, where a represents the higher risk band and a' represents the lower risk band, for $i = 1, 2$. We also assume that all ongoing loans end in repayment at month 72. We applied the same process for loan ages prior to the beginning of the estimable window (i.e., loan ages prior to month 4).

As expected, we can see that $\hat{\pi}_{n,a|x}$ increases as the credit risk of the risk band a increases. As of pricing ($x = 1$), in order of increasing credit risk (i.e., deep subprime to super-prime), the market-implied credit spreads are approximately 250, 500, 750, 1,000, and 1,500 basis

points, respectively. Further, the narrowness of the confidence intervals demonstrates the competitiveness of the consumer auto lending market, especially for borrowers in higher credit quality risk bands.

The respective shape of $\hat{\pi}_{n,a|x}$ over time by risk band is of particular interest. Consider first the deep subprime risk band (dashed line, row one of Figure 5). The market-implied credit spread initially declines before peaking just before loan age 40 and then dropping precipitously. As discussed in Section III.B, this implies a rapid gain in relative profitability for these loans as they remain active and paying. We also observe a similar shape for the subprime risk band. The near-prime risk band slowly increases to a peak around loan age 40 months and then also declines. Thus, for the three lowest quality credit risk bands, the relative profitability increases as the loans remain active and paying. Given the lowest credit quality risk bands typically have the highest interest rates and the observed credit risk convergence observed in Section III.C, this is not surprising.

Loans in the prime risk band have a declining market-implied credit spread until about loan age ten months, before increasing to a peak by loan age 40 and remaining approximately level. This suggests that relatively profitability for prime loans generally decreased as the loans remained active and paying. Finally, the overall pattern for super-prime loans mirrors that of prime loans, but the absolute change is relatively minimal. In other words, the profitability dynamics of super-prime consumer auto loans was fairly stable for the life of the loan. Again, see Section III.B as needed.

Finally, we attempt to estimate how much a borrower that fails to refinance a loan is overpaying by risk band after the point of credit risk convergence to a higher credit quality risk band by using an assumption of an efficient market. That is, the difference in interest rates between two contracts with an equivalent level of credit risk is a useful estimate of the amount one borrower is overpaying. Following this approach, the estimated merit-based interest rate reductions, based on the average interest rate of each risk band, may be found on the left-hand side of Table 3. We may compare Table 3 with Figure 5 and Table 2. As an example, consider the case of a deep subprime borrower. After making approximately three years of on-time payments, we estimate this borrower becomes a near-prime credit risk. By not refinancing, however, such a borrower is effectively overpaying by an average annual percentage rate (APR) of 5.61%. The same deep subprime borrower becomes a prime credit risk after making 46 months of on-time payments. By not refinancing, however, such

33

Table 3: **Estimated Overpayment by APR**. The average difference in rates for borrowers in disparate risk bands, which suggests the amount borrowers in the lower credit risk band are overpaying versus the higher credit risk band in light of the going risk-adjusted market rate after the point of credit risk convergence between risk bands. The same comparison is given for loans that survived at least 40 months (right) as a robustness check that the average rate differences are stable (i.e., lenders are not identifying these strong borrowers at onset).

	Avg. APR Diff.			Avg. APR Diff. ($X \geq 40$)		
	near-prime	prime	super-prime	near-prime	prime	super-prime
deep subprime	561	905	1,391	601	956	1,366
subprime	285	630	1,115	316	670	1,080
near-prime		344	830		354	765

a borrower is effectively overpaying by an APR of 9.05%. The level a subprime borrower might be overcharged may be even higher. We estimate subprime borrowers become prime credit risks after 42 months, and they become super-prime credit risks after 49 months. This suggests that subprime borrowers who do not refinance after 42 months are overpaying by an average APY of 6.30%. This becomes a much higher 11.15% APR over-payment after 49 months. We may repeat such estimates for near-prime borrowers to find an average APR overcharge of 3.44% after 28 months of on-time payments and an average APR overcharge of 8.30% after 42 months of on-time payments.

The figures on the left-hand side of Table 3 represent average APRs for all borrowers within each risk band regardless of the time the loan ended. It is possible that our risk band definitions are less accurate than the risk-based pricing of the lenders, and a concern is that a borrower in a lower credit quality risk band that survives was originally assigned a lower interest rate (i.e., the lender "got it right"). Figure 6 shows a density plot of the distribution of interest rates within each risk band by survival time. It is interesting to observe little change in the interest rate distribution for the deep subprime and subprime risk bands as loan age increases. On the other hand, the prime risk band appears bimodal for loans that default early but then peaks just under a 10% APR for loans that survive at least until month 40. Thus, we repeat the analysis of the left-hand side of Table 3 but this time only consider loans that have survived until at least month 40 on the right-hand side of Table 3. In general, the results are comparable across both sides of Table 3. Further robustness checks may be found in Section IV.
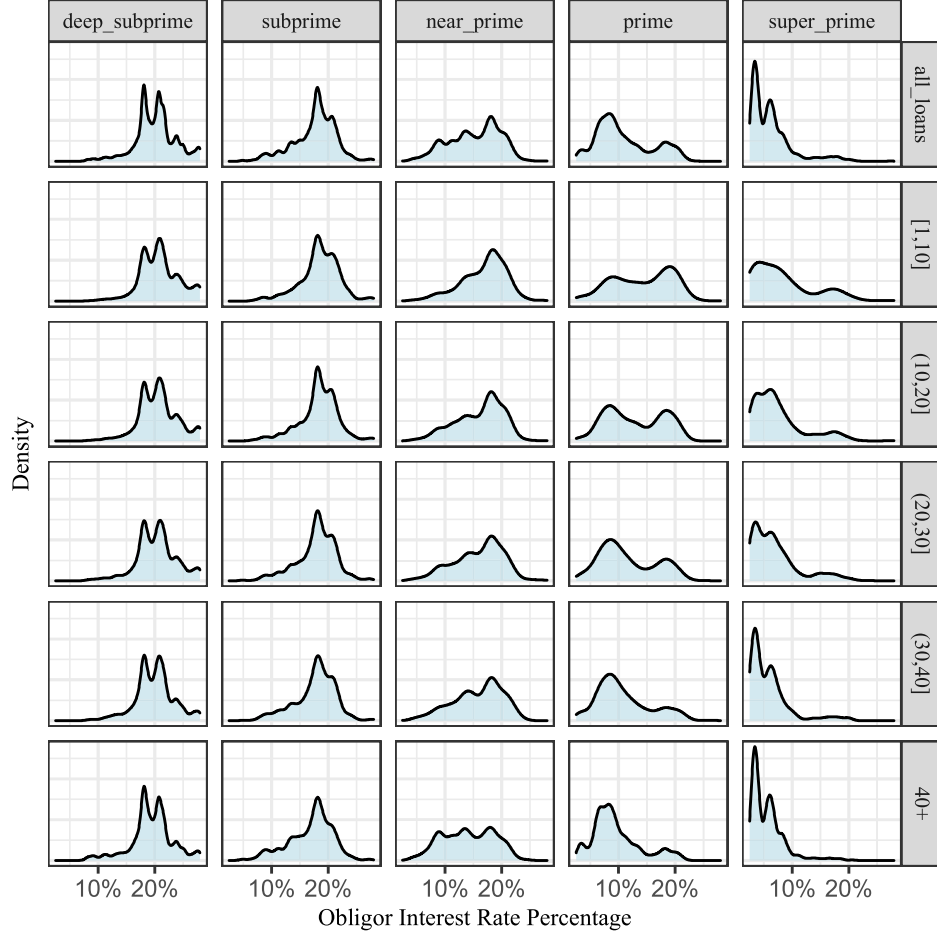
Figure 6: **Interest rate distribution by survival**. A distribution plot of the interest rates within credit risk band by length of survival. The consistency of the distributions within each risk band by survival time suggests that the lenders are not assigning loans that remain current a lower rate than loans that default early. This serves as a robustness check that comparing the average rate in each risk band is a reasonable proxy for the overpayment estimations in Table 3.

IV.   Sensitivity Analysis

The results in Sections II.C and III.C may have some potential limitations that would prevent widespread generalizations. In this section, we will discuss four potential limitations related to the possible influence of the Coronavirus pandemic, our definition of the risk bands, assumptions made for the cause-specific hazard rates beyond the estimable window, and our

focus on consumer automobile loans. We also provide additional analysis and arguments to support why our larger conclusion of credit risk convergence is robust to these possible limitations.

IV.A.   *Influence of COVID-19*

As alluded to in Section II.C, we have attributed the large spike around loan age 40 for the default cause-specific hazard rate estimate observable in Figure 3 to the Spring 2020 economic shutdown in response to initial rapid spread of the Coronavirus disease. Since 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 risk convergence occurs by loan age 46), this may suggest 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 some inherent property of consumer risk behavior. In other words, only the strongest credits could survive such a shock, and consumer loan 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, deep subprime and near-prime, or subprime and near 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, 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. Thus, 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 (the final 2019 sample consists of 56,683 loans). Figure 7 presents the estimated cause-specific hazard rates for default rates plus confidence intervals for the 2019 sample. 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

36

to the Spring 2020 economic shutdown. We also repeat the credit risk convergence matrix of Table 2 in the top portion of Table 4. First, 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 35, which was the last observable age in the 2019 sample as of May 2022. In the 2017 issuance, the subprime risk band converged with the super-prime risk band by loan age 49. This suggests that loan age also plays a role. Similarly, while the 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 46, which is only 6 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.

IV.B.   Influence of Risk Band Definition

In Section I.A and for the subsequent analysis of Sections II.C and III.C, we assigned borrowers to a risk band based on credit score, as we did not have access to each lender's credit scoring model. This is at best an approximation because, while credit score is a major component of a prospective borrower's loan application, it is typically not the only factor. For example, a borrower's payment-to-income ratio is likely also an important component. Indeed, a close examination of the third row of Figure 2 reveals a wide range of interest rates within each risk band. Therefore, it is possible that the observed credit risk convergence between risk bands is just a product of crudely defined risk bands.
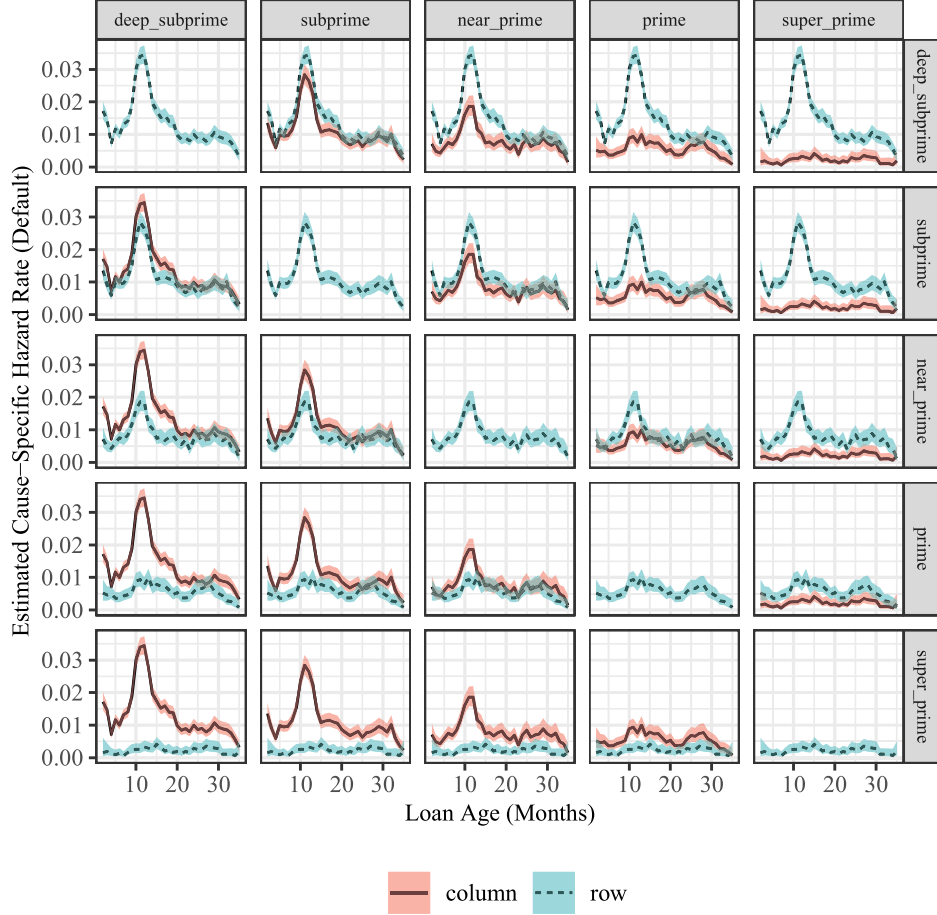
Figure 7: **Estimated cause-specific hazard rate for default risk-band comparison (2019 issuance)**. A repeat of Figure 3 for the same set of four bonds and loan filtering of Section I but for the 2019 issuance as a sensitivity check that the economic shock of COVID-19 is not the sole reason for credit risk convergence between consumers in 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.

We believe this is not the case, however. As justification, we redefined the risk bands by assigned interest rate, as the offered rate is another indicative representation of the lender's perception of a borrower's risk profile. Specifically, we assigned borrower's with an APR of 0-5% to the super-prime risk band, 5-10% to the 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. A plot of the cause-specific hazard rates for default plus 95% asymptotic confidence intervals

Table 4: **Estimated Month of Credit Risk Convergence Sensitivity Analysis**. A summary matrix of the first month after ten months that the asymptotic confidence intervals in Figure 7 overlap for three consecutive months by risk band (top). Estimates are only available for loan ages $2 \leq X \leq 35$. This matrix can be compared with Table 2 as a sensitivity analysis for the impact of COVID-19. A summary matrix of the first month after ten months that the asymptotic confidence intervals in Figure 8 overlap for three consecutive months by risk band (bottom). This matrix can be compared with Table 2 as a sensitivity analysis for the definition of risk bands by credit score.

COVID-19 Sensitivity Analysis					
	deep subprime	subprime	near-prime	prime	super-prime
deep subprime	10	18	20	25	NA
subprime		10	14	15	NA
near-prime			10	13	26
prime				10	10
super-prime					10

Risk Band Definition Sensitivity Analysis					
	deep subprime	subprime	near-prime	prime	super-prime
deep subprime	11	36	50	50	51
subprime		11	41	48	48
near-prime			11	13	43
prime				11	11
super-prime					11

for risk bands defined by APR rather than credit score for the 2017 issuance may be found in Figure 8. A comparison with Figure 3 reveals a very similar trend of loan risk convergence. In addition, the lower portion of Table 4 also compares favorably with Table 2.

IV.C.   Influence of Assumptions

The market-implied credit spread analysis of Section III.C relied on two important assumptions because the ABS data was right-censored (i.e., the oldest estimable loan age probability was for 54 months). The first assumption was that the cause-specific hazard rates were constant (geometric) for all loan ages beyond the estimable window. While this assumption is common in survival analysis (Klugman et al., 2012, Section 12.1) and seems reasonable
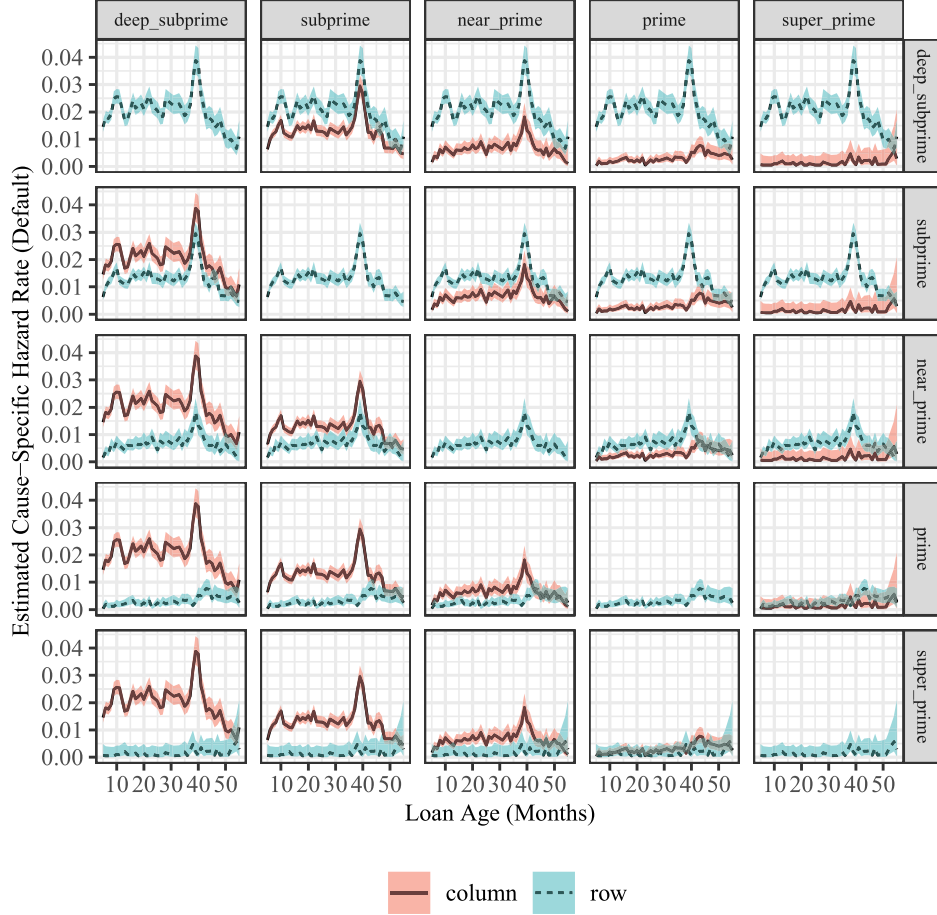
Figure 8: **Estimated cause-specific hazard rate for default risk-band comparison (risk bands defined by interest rate)**. A repeat of Figure 3 but redefining the risk bands by interest rate rather than credit score as a sensitivity analysis against the idea that lenders' pricing models are materially more precise than a categorization by credit score. The consistency between this figure and Figure 3 suggests the conclusions of this paper are robust to an alternative definition of consumer credit risk band by assigned interest rate.

based on the hazard rate estimates of the very right tail of the estimable window in Figure 3, it is a strong assumption about the shape of the cause-specific hazard rates that may prove to be inaccurate with additional data. The second assumption was that the constant hazard rate beyond the estimable window was the same for two risk bands after the credit risk convergence point of those two risk bands. While this assumption is reasonable once the confidence intervals of two cause-specific hazard rate estimators began overlapping, it
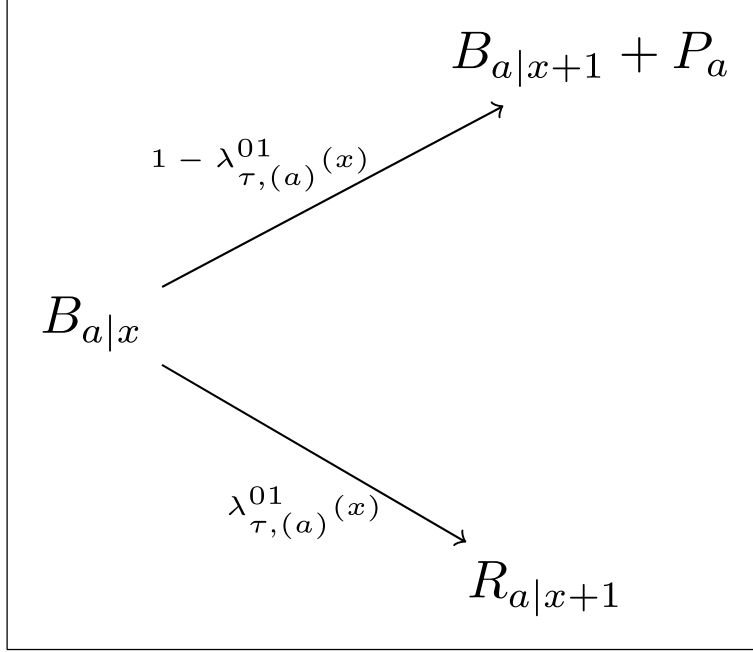
Figure 9: **Hypothetical risky fixed-income asset and path probabilities**. This hypothetical risky asset pays either (1) the outstanding balance at loan age $x+1$, $B_{a|x+1}$, plus the next month's payment due, P_a, with probability $1 - \lambda_{\tau,(a)}^{01}(x)$ or (2) the recovery amount at time $x + 1$, $R_{a|x+1}$, based on the estimates in Figure 1, with probability $\lambda_{\tau,(a)}^{01}(x)$, where a denotes one of the five standard risk bands (deep subprime, subprime, near-prime, prime, or super-prime).

is possible the behavior may be different outside the estimable window. Concisely, the assumption beyond month 54 of *ceteris paribus* simply may not hold; we just do not have the data to know for sure.

As a robustness check, we propose another point of analysis that does not require re-assessing the remaining full life of the loan. Specifically, we will examine a rolling monthly expected annualized rate of return assuming an investor purchased a risky fixed-income asset at a price of the outstanding balance of the consumer loan at age x, $B_{a|x}$, with a one-month term. This hypothetical risky asset pays either (1) the outstanding balance at loan age $x+1$, $B_{a|x+1}$, plus the next month's payment due, P_a, with probability $1 - \lambda_{\tau,(a)}^{01}(x)$ or (2) the recovery amount at time $x + 1$, $R_{a|x+1}$, based on the estimates in Figure 1, with probability $\lambda_{\tau,(a)}^{01}(x)$, where a denotes one of the five standard risk bands. We illustrate this hypothetical asset in Figure 9.

Formally, if we let x be in the estimable window, $x \in \{4, 5, \ldots, 54\}$, and define the

expected value of a $\$B$ loan to a borrower in risk band a for month x as

$$\text{EPV}^1_{a|x} = \lambda^{01}_a(x)(R_{a|x+1}) + (1 - \lambda^{01}_a(x))(B_{a|x+1} + P_a), \tag{12}$$

then expected monthly rate of return is the rate $\tilde{r}_{a|x}$ such that $\text{EPV}^1_{a|x} = B_x$. Note that we assume the remaining payments are certain (i.e., the risky asset may be traded at time $x+1$ for $B_{a|x}$). It's clear that the behavior of the loan outside of the estimable window does not play into the calculation of (12) or $\tilde{r}_{x|a}$. For ease of interpretation, we elected to consider a single loan of \$100 for 72 months with a payment determined by the average interest rate of each risk band. The results for both the 2017 and 2019 issuance may be found in Figure 10, where we replace the cause-specific hazard rate in (12) with its estimate from (5). We can see that, as the loans mature, the rolling monthly return for the lower credit quality risk bands — deep subprime, subprime, near-prime — outpaces the higher credit quality risk bands — prime, super-prime. This output is consistent with the results of Section III.C. Namely, as the default risk declines with loan age, the bonds in lower credit risk bands become much more profitable.

IV.D.   Influence of Loan Type

Our analysis has only focused on consumer automobile loans. We elected to use consumer auto loans because they generally represent a high priority of payment to the consumer and conveniences of data availability and familiarity with the ABS auto loan asset-class (see additional discussion in Section I). Thus, it is possible credit risk convergence may not occur in other types of consumer loans, particularly those not secured by collateral. Along the same lines, a borrower is slowly building an equity position as the loan matures, and so the declining default risk for lower credit risk bands may also reflect a borrower's desire not to lose an asset they are gradually getting closer to owning 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. Thus, we suggest caution before attempting to generalize these results outside of consumer auto loans, and we leave this open as an area of further study.
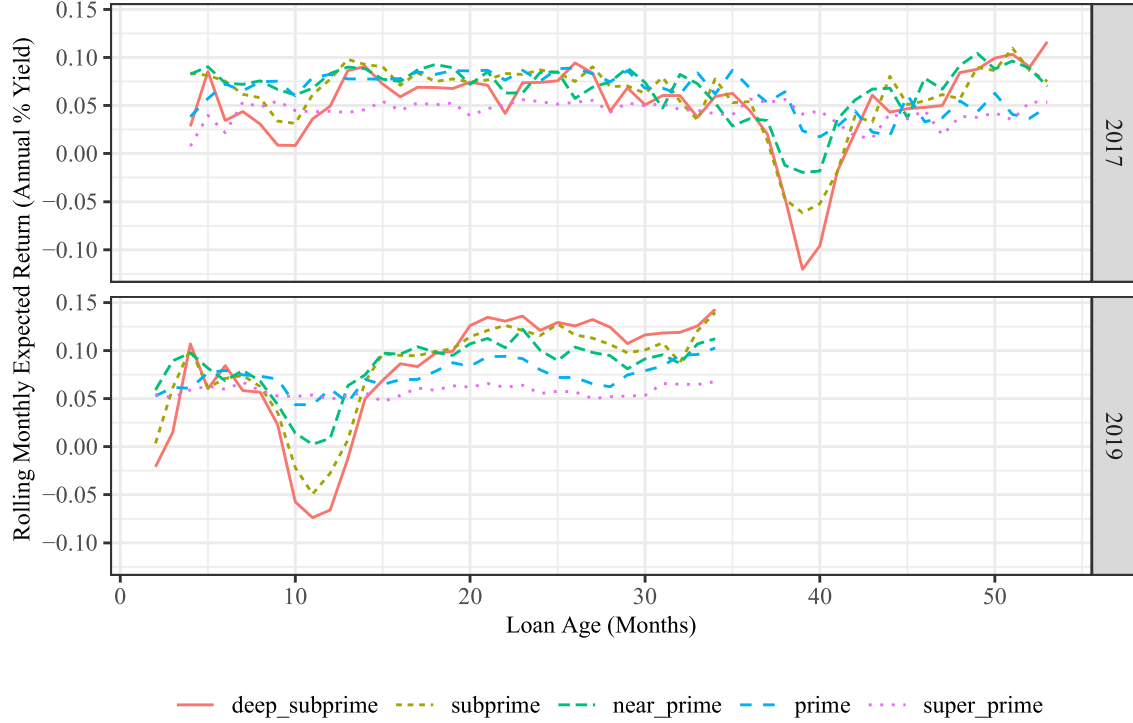
Figure 10: **The expected monthly rate of return given survival**. A plot of $\tilde{r}_{a|x}$ by loan age, risk band, and issuance as calculated using the formulations in (12) and Figure 9. This calculation is a rolling one-month expected return, which serves as a sensitivity test against the assumptions made beyond the estimable window for the calculations in Figure 5. We can see that the one-month expected rate of return is roughly equal prior to the point of credit risk convergence, at which point the borrowers in the lowest credit quality risk bands that remain current gradually become more and more profitable.

V.   Conclusion

In a consumer lending market driven by risk-based pricing, lower credit borrowers incur a higher cost of borrowing at contract signing to compensate for a higher risk of default. Using survival analysis techniques on data from securitization pools of 72–73 month consumer automobile loans spanning the full spectrum of consumer credit quality, we find that borrowers originally categorized in high-risk credit bands (deep subprime, subprime, near-prime) who remain active and paying converge in conditional monthly default risk to borrowers originally categorized in low-risk credit bands (near-prime, prime, super prime) after approximately

28 to 49 months. This is a statistical demonstration of *credit risk convergence.* Credit risk convergence is an intuitive result that may be rephrased in simple terms: regardless of a borrower's credit profile at contract signing, the longer a borrower remains active and paying, the lower the risk of default. Our first major contribution is thus an empirical study of *when* the borrower's risk profile changes over the lifetime of the loan. We further find credit risk convergence cannot be fully explained by the economic shock of the Coronavirus and is robust to definitions of risk bands by credit score or interest rate.

The consumer automobile lending space is not dynamically priced based on a borrower's changing credit profile, however, and so these high risk borrowers that eventually become low risk borrowers slowly become more and more profitable from the perspective of the lender on a conditional expected rate of return basis. Because we identify other borrowers with equivalent credit risk that are paying lower interest rates, we conjecture that such borrowers unjustifiably overpay by APRs ranging from 285 to 1,391 basis points. We find our estimates are robust to assumptions made outside the estimable window and for changing borrower demographics based on loan survival. All of our results are also adjusted for prepayments.

Given our focus on secured automobile loans from within a core economic lending space, such large market inefficiencies are troubling. Though the typical consumer generally has a poor reputation in making financial decisions (e.g. Gross and Souleles, 2002; Stango and Zinman, 2011; Lusardi and de Bassa Scheresberg, 2013; Campbell, 2016; Heidhues and Kőszegi, 2016; Dobbie et al., 2021), we are encouraged that prepayments accelerate as the auto loans mature. This suggests that on some level borrowers self-correct for the market's inability to account for a changing risk profile. On the other hand, prepayments accelerate for borrowers in all risk bands, and so the decision to refinance a loan may instead by driven by a desire to upgrade the underlying automobile rather than financially motivated. So, what can be done?

Aside from encouraging borrowers to refinance these overpriced loans, which may prove less effective in practice than in theory (e.g., Keys et al., 2016; Agarwal et al., 2017), we see three potential solutions from market participants other than the individual consumer. 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.* Certainly, lenders have the data 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, we suggest that competing lenders seek out these mature auto loans for refinancing. If a deep subprime borrower eventually performs like a prime borrower but is paying an APR north of 20%, then there is a lower rate that would both lower this borrower's financing cost and be profitable to a second lender. Third, there is the regulatory angle, which has been successful in other consumer lending spaces (e.g., Stango and Zinman, 2011; Agarwal et al., 2014). For example, we see 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.

We conclude this article by suggesting similar research be explored in consumer lending for residential mortgages and other consumer finance products. If the point of credit risk convergence remains sometime between 28 to 49 months, as we found in 72–73 month automobile loan contracts, for 30-year residential mortgages, for example, then the potential savings to consumers could be even more substantial than the findings herein.

Appendix

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

$$
\begin{aligned}
&\hat{\lambda}_{\tau,n}^{0i}(k) - \lambda_{\tau}^{0i}(k) \\
&= \frac{\frac{1}{n}\sum_{j=1}^{n} \mathbf{1}_{X_j \leq C_j}\mathbf{1}_{Z_{X_j}=i}\mathbf{1}_{\min(X_j,C_j)=k}}{\hat{C}_{\tau,n}(k)} - \frac{f_{*,\tau}^{0i}(k)}{C_{\tau}(k)}
\end{aligned}
$$

$$= \frac{\{\sum_{j=1}^{n} \mathbf{1}_{X_j \leq C_j} \mathbf{1}_{Z_{X_j}=i} \mathbf{1}_{\min(X_j, C_j)=k}\} C_\tau(k) - f_{*,\tau}^{0i}(k) \hat{C}_{\tau,n}(k)}{\hat{C}_{\tau,n}(k) C_\tau(k)}$$

$$= \left[\frac{1}{\hat{C}_{\tau,n}(k) C_\tau(k)} \right] \frac{1}{n} \sum_{j=1}^{n} \{ \mathbf{1}_{X_j \leq C_j} \mathbf{1}_{Z_{X_j}=i} \mathbf{1}_{\min(X_j, C_j)=k} C_\tau(k) - f_{*,\tau}^{0i}(k) \mathbf{1}_{Y_j \leq k \leq \min(X_j, C_j)} \}.$$

Define

$$H_{\tau,k(j)}^{0i} = \mathbf{1}_{X_j \leq C_j} \mathbf{1}_{Z_{X_j}=i} \mathbf{1}_{\min(X_j, C_j)=k} C_\tau(k) - f_{*,\tau}^{0i}(k) \mathbf{1}_{Y_j \leq k \leq \min(X_j, C_j)},$$

for $1 \leq j \leq n$ and

$$\mathbf{A}_{\tau,n} = \mathrm{diag}([\hat{C}_{\tau,n}(\Delta + 1) C_\tau(\Delta + 1)]^{-1}, \ldots, [\hat{C}_{\tau,n}(\xi) C_\tau(\xi)]^{-1}).$$

Then,

$$\hat{\boldsymbol{\Lambda}}_{\tau,n}^{0i} - \boldsymbol{\Lambda}_\tau^{0i} = \mathbf{A}_{\tau,n} \frac{1}{n} \sum_{j=1}^{n} \begin{bmatrix} H_{\tau,\Delta+1(j)}^{0i} \\ \vdots \\ H_{\tau,\xi(j)}^{0i} \end{bmatrix},$$

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

$$\hat{\boldsymbol{\Lambda}}_{\tau,n}^{0i} - \boldsymbol{\Lambda}_\tau^{0i} = \mathbf{A}_{\tau,n} \frac{1}{n} \sum_{j=1}^{n} \mathbf{H}_{\tau,(j)}^{0i}.$$

It is noteworthy the components of $\mathbf{H}_{\tau,(j)}^{0i}$ are uncorrelated. More specifically,

$$\mathrm{Cov}[H_{\tau,k(j)}^{0i}, H_{\tau,k'(j)}^{0i}] = \begin{cases} C_\tau(k) f_{*,\tau}^{0i}(k) [C_\tau(k) - f_{*,\tau}^{0i}(k)], & k = k' \\ 0, & k \neq k'. \end{cases} \tag{13}$$

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

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

$$= f_{*,\tau}^{0i}(k) C_\tau(k) - f_{*,\tau}^{0i}(k) C_\tau(k)$$

$$= 0.$$

Therefore,

$$\operatorname{Cov}[H^{0i}_{\tau,k(j)}, H^{0i}_{\tau,k'(j)}] = \mathbf{E}H^{0i}_{\tau,k(j)}H^{0i}_{\tau,k'(j)}$$
$$= \mathbf{E}\{\mathbf{1}_{X_j \leq C_j}\mathbf{1}_{Z_{X_j}=i}\mathbf{1}_{\min(X_j,C_j)=k}C_\tau(k) - f^{0i}_{*,\tau}(k)\mathbf{1}_{Y_j \leq k \leq \min(X_j,C_j)}\}$$
$$\times \{\mathbf{1}_{X_j \leq C_j}\mathbf{1}_{Z_{X_j}=i}\mathbf{1}_{\min(X_j,C_j)=k'}C_\tau(k') - f^{0i}_{*,\tau}(k')\mathbf{1}_{Y_j \leq k' \leq \min(X_j,C_j)}\}$$
$$= C_\tau(k)C_\tau(k')\mathbf{E}\mathbf{1}_{X_j \leq C_j}\mathbf{1}_{Z_{X_j}=i}\mathbf{1}_{\min(X_j,C_j)=k}\mathbf{1}_{X_j \leq C_j}\mathbf{1}_{Z_{X_j}=i}\mathbf{1}_{\min(X_j,C_j)=k'}$$
$$- C_\tau(k)f^{0i}_{*,\tau}(k')\mathbf{E}\mathbf{1}_{X_j \leq C_j}\mathbf{1}_{Z_{X_j}=i}\mathbf{1}_{\min(X_j,C_j)=k}\mathbf{1}_{Y_j \leq k' \leq \min(X_j,C_j)}$$
$$- C_\tau(k')f^{0i}_{*,\tau}(k)\mathbf{E}\mathbf{1}_{X_j \leq C_j}\mathbf{1}_{Z_{X_j}=i}\mathbf{1}_{\min(X_j,C_j)=k'}\mathbf{1}_{Y_j \leq k \leq \min(X_j,C_j)}$$
$$+ f^{0i}_{*,\tau}(k)f^{0i}_{*,\tau}(k')\mathbf{E}\mathbf{1}_{Y_j \leq k \leq \min(X_j,C_j)}\mathbf{1}_{Y_j \leq k' \leq \min(X_j,C_j)}.$$

We proceed by cases.

Case 1: $k = k'$.

Working through each expectation in $\operatorname{Cov}[H^{0i}_{\tau,k(j)}, H^{0i}_{\tau,k'(j)}]$, we have

$$\mathbf{E}\mathbf{1}_{X_j \leq C_j}\mathbf{1}_{Z_{X_j}=i}\mathbf{1}_{\min(X_j,C_j)=k}\mathbf{1}_{X_j \leq C_j}\mathbf{1}_{Z_{X_j}=i}\mathbf{1}_{\min(X_j,C_j)=k'} = \mathbf{E}\mathbf{1}_{X_j \leq C_j}\mathbf{1}_{Z_{X_j}=i}\mathbf{1}_{\min(X_j,C_j)=k}$$
$$= f^{0i}_{*,\tau}(k),$$

$$\mathbf{E}\mathbf{1}_{X_j \leq C_j}\mathbf{1}_{Z_{X_j}=i}\mathbf{1}_{\min(X_j,C_j)=k}\mathbf{1}_{Y_j \leq k' \leq \min(X_j,C_j)} = \mathbf{E}\mathbf{1}_{X_j \leq C_j}\mathbf{1}_{Z_{X_j}=i}\mathbf{1}_{\min(X_j,C_j)=k'}\mathbf{1}_{Y_j \leq k \leq \min(X_j,C_j)}$$
$$= \mathbf{E}\mathbf{1}_{X_j \leq C_j}\mathbf{1}_{Z_{X_j}=i}\mathbf{1}_{\min(X_j,C_j)=k}\mathbf{1}_{Y_j \leq k \leq \min(X_j,C_j)}$$
$$= \mathbf{E}\mathbf{1}_{X_j \leq C_j}\mathbf{1}_{Z_{X_j}=i}\mathbf{1}_{\min(X_j,C_j)=k}$$
$$= f^{0i}_{*,\tau}(k),$$

and

$$\mathbf{E}\mathbf{1}_{Y_j \leq k \leq \min(X_j,C_j)}\mathbf{1}_{Y_j \leq k' \leq \min(X_j,C_j)} = \mathbf{E}\mathbf{1}_{Y_j \leq k \leq \min(X_j,C_j)} = C_\tau(k).$$

Thus,

$$\operatorname{Cov}[H^{0i}_{\tau,k(j)}, H^{0i}_{\tau,k'(j)}] = C_\tau(k)f^{0i}_{*,\tau}(k)[C_\tau(k) - f^{0i}_{*,\tau}(k)].$$

Case 2: $k \neq k'$.

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

$$\mathbf{E}\mathbf{1}_{X_j \le C_j}\mathbf{1}_{Z_{X_j}=i}\mathbf{1}_{\min(X_j,C_j)=k}\mathbf{1}_{X_j \le C_j}\mathbf{1}_{Z_{X_j}=i}\mathbf{1}_{\min(X_j,C_j)=k'} = 0,$$

$$\mathbf{E}\mathbf{1}_{X_j \le C_j}\mathbf{1}_{Z_{X_j}=i}\mathbf{1}_{\min(X_j,C_j)=k}\mathbf{1}_{Y_j \le k' \le \min(X_j,C_j)}$$
$$= \begin{cases} \Pr(X_j \le C_j, Z_{X_j} = i, \min(X_j,C_j) = k, Y_j \le k'), & k > k' \\ 0, & k < k', \end{cases}$$

$$\mathbf{E}\mathbf{1}_{X_j \le C_j}\mathbf{1}_{Z_{X_j}=i}\mathbf{1}_{\min(X_j,C_j)=k'}\mathbf{1}_{Y_j \le k \le \min(X_j,C_j)}$$
$$= \begin{cases} 0, & k > k' \\ \Pr(X_j \le C_j, Z_{X_j} = i, \min(X_j,C_j) = k', Y_j \le k), & k < k', \end{cases}$$

and

$$\mathbf{E}\mathbf{1}_{Y_j \le k \le \min(X_j,C_j)}\mathbf{1}_{Y_j \le k' \le \min(X_j,C_j)} = \Pr(Y_j \le k \le \min(X_j,C_j), Y_j \le k' \le \min(X_j,C_j)).$$

Thus,

$$\mathrm{Cov}[H_{\tau,k(j)}^{0i}, H_{\tau,k'(j)}^{0i}] = f_{*,\tau}^{0i}(\min(k,k'))\bigg\{$$
$$- C_\tau(\max(k,k'))\Pr(X_j \le C_j, Z_{X_j} = i, \min(X_j,C_j) = \max(k,k'), Y_j \le \min(k,k'))$$
$$+ f_{*,\tau}^{0i}(\max(k,k'))\Pr(Y_j \le k \le \min(X_j,C_j), Y_j \le k' \le \min(X_j,C_j))\bigg\}.$$

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

$$C_\tau(\max(k,k')) = \Pr(Y_j \le \max(k,k') \le \min(X_j,C_j))$$
$$= \Pr(Y \le \max(k,k'), X \ge \max(k,k'), C \ge \max(k,k') \mid Y \le X)$$
$$= \{\Pr(Y \le \max(k,k') \le C)\Pr(X \ge \max(k,k'))\}/\alpha,$$

$$\Pr(X_j \le C_j, Z_{X_j} = i, \min(X_j, C_j) = \max(k, k'), Y_j \le \min(k, k'))$$
$$= \Pr(C \ge \max(k, k'), Z_X = i, X = \max(k, k'), Y \le \min(k, k') \mid Y \le X)$$
$$= \{\Pr(X = \max(k, k'), Z_X = i) \Pr(Y \le \min(k, k'), C \ge \max(k, k'))\}/\alpha,$$

$$f_{*,\tau}^{0i}(\max(k, k')) = \Pr(X = \max(k, k'), C \ge \max(k, k'), Z_x = i \mid Y \le X)$$
$$= \{\Pr(X = \max(k, k'), Z_X = i) \Pr(Y \le \max(k, k') \le C)\}/\alpha,$$

and

$$\Pr(Y_j \le k \le \min(X_j, C_j), Y_j \le k' \le \min(X_j, C_j))$$
$$= \Pr(Y \le \min(k, k'), C \ge \max(k, k'), X \ge \max(k, k') \mid Y \le X)$$
$$= \{\Pr(Y \le \min(k, k'), C \ge \max(k, k')) \Pr(X \ge \max(k, k'))\}/\alpha$$

Therefore,

$$C_\tau(\max(k, k')) \Pr(X_j \le C_j, Z_{X_j} = i, \min(X_j, C_j) = \max(k, k'), Y_j \le \min(k, k'))$$
$$= f_{*,\tau}^{0i}(\max(k, k')) \Pr(Y_j \le k \le \min(X_j, C_j), Y_j \le k' \le \min(X_j, C_j)),$$

and so $\mathrm{Cov}[H_{\tau,k(j)}^{0i}, H_{\tau,k'(j)}^{0i}] = 0$ when $k \ne k'$. This confirms (13). Now define

$$\mathbf{D}_\tau^{0i} = \mathrm{diag}\big(C_\tau(\Delta+1)f_{*,\tau}^{0i}(\Delta+1)[C_\tau(\Delta+1) - f_{*,\tau}^{0i}(\Delta+1)], \ldots, C_\tau(\xi)f_{*,\tau}^{0i}(\xi)[C_\tau(\xi) - f_{*,\tau}^{0i}(\xi)]\big),$$

and

$$\bar{\mathbf{H}}_{\tau,n}^{0i} = \frac{1}{n}\sum_{j=1}^{n} \mathbf{H}_{\tau,(j)}^{0i}.$$

By the multivariate Central Limit Theorem (Lehmann and Casella, 1998, Theorem 8.21, pg. 61), therefore,

$$\sqrt{n}(\bar{\mathbf{H}}_{\tau,n}^{0i} - \mathbf{0}) \overset{\mathcal{L}}{\longrightarrow} N(\mathbf{0}, \mathbf{D}_\tau^{0i}), \text{ as } n \to \infty.$$

Next, define $\mathbf{V}_\tau = \mathrm{diag}(C_\tau(\Delta+1)^{-2}, \ldots, C_\tau(\xi)^{-2})$. By Lemma 1 (Lautier et al., 2022), $\mathbf{A}_{\tau,n} \overset{\mathcal{P}}{\longrightarrow} \mathbf{V}_\tau$, as $n \to \infty$. Thus, by multivariate Slutsky's Theorem (Lehmann and Casella,

1998, Theorem 5.1.6, pg. 283),

$$\sqrt{n}(\mathbf{A}_{\tau,n}\bar{\mathbf{H}}_{\tau,n}^{0i}) \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{V}_\tau \mathbf{D}_\tau^{0i} \mathbf{V}_\tau^\top), \text{ as } n \to \infty.$$

We may complete the proof by observing $\mathbf{V}_\tau \mathbf{D}_\tau^{0i} \mathbf{V}_\tau^\top = \mathbf{\Sigma}^{0i}$ and $\mathbf{A}_{\tau,n}\bar{\mathbf{H}}_{\tau,n}^{0i} = \hat{\mathbf{\Lambda}}_{\tau,n}^{0i} - \mathbf{\Lambda}_\tau^{0i}$. $\qquad \square$

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 $i = 1, 2$,

$$\sqrt{n}\big(\ln\hat{\lambda}_{\tau,n}^{0i}(x) - \ln\lambda_\tau^{0i}(x)\big) \xrightarrow{\mathcal{L}} N\left(0, \frac{f_{*,\tau}^{0i}(x)\{C_\tau(x) - f_{*,\tau}^{0i}(x)\}}{C_\tau(x)^3}\frac{1}{\lambda_\tau^{0i}(x)^2}\right).$$

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. $\qquad \square$

B. Proofs: Section III

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

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

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

$$\begin{aligned}
B_x &= B(1+r_a)^x - P\left[\frac{(1+r_a)^x - 1}{r_a}\right] \\
&= B(1+r_a)^x - B\left[\frac{1 - (1+r_a)^{-\xi}}{r_a}\right]^{-1}\left[\frac{(1+r_a)^x - 1}{r_a}\right].
\end{aligned} \qquad (14)$$

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

$$B_x = P \left[\frac{1}{(1 + \rho_{a|x})} + \cdots + \frac{1}{(1 + \rho_{a|x})^{\xi - x}} \right]$$
$$= B \left[\frac{1 - (1 + r_a)^{-\xi}}{r_a} \right]^{-1} \left[\frac{1 - (1 + \rho_{a|x})^{-(\xi - x)}}{\rho_{a|x}} \right].$$

Use (14) and solve for $\rho_{a|x}$ to complete the proof. $\qquad\square$

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). $\qquad\square$

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 5. 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 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.

Table 5: **Simulation Study Lifetime of Interest Probabilities**. The true probabilities of the lifetime random variable, X, for the simulation study results of Figure 11. The probabilities $p(x)$ and $\Pr(X = x)$ for $x \in \{1, \dots, 10\}$ 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 $\{1, \dots, 5\}$.

$p(x)$	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

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 $j = 1, \dots, n$). 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 $\hat{f}^{0i}_{*,\tau,n}(x)$, $\hat{C}_{\tau,n}(x)$, and $\hat{\lambda}^{0i}_{\tau,n}(x)$ for $i = 1, 2$ and $x \in \{1, \dots, 10\}$ over $r = 1{,}000$ replicates.

To validate the asymptotic results of Theorem 1, we compare the empirical covariance matrix against the derived asymptotic covariance matrix, $\boldsymbol{\Sigma}^{0i}$, by examining estimates of the confidence intervals using Lemma 1. Figure 11 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.
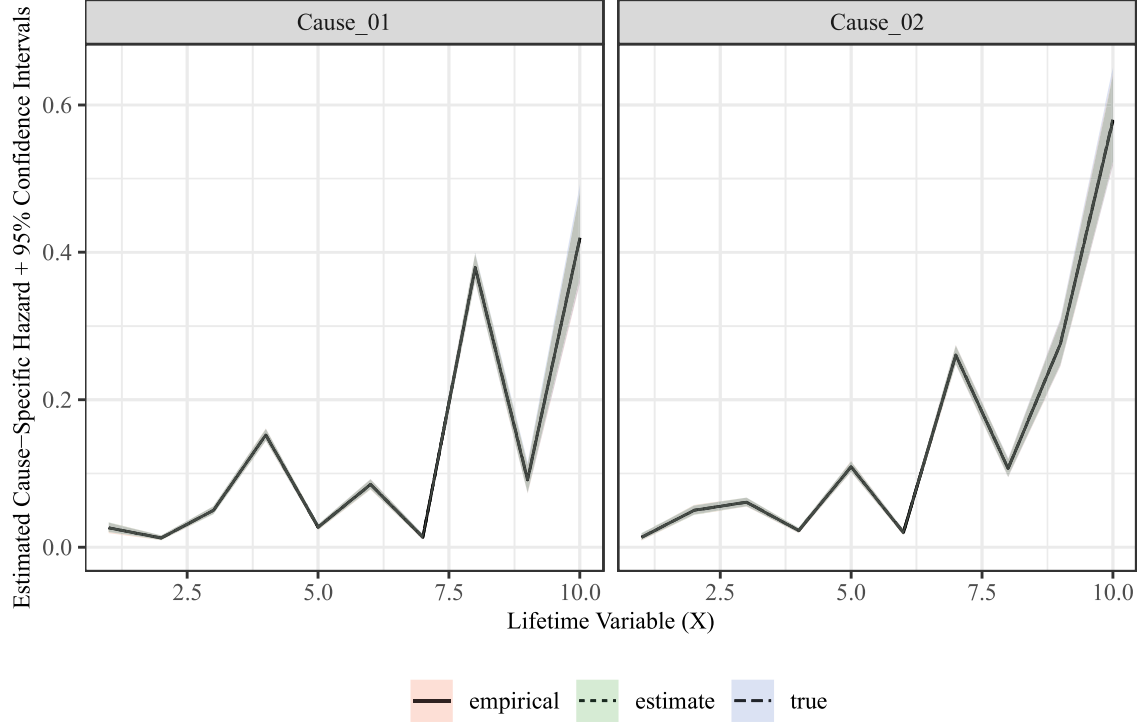
Figure 11: **Simulation study results.** A comparison of true $\lambda_\tau^{0i}(x)$ and estimated $\hat{\lambda}_{\tau,n}^{0i}(x)$, including confidence intervals, for the distribution in Table 5 and $i = 1, 2$. The "true" values are from Proposition 1 and Lemma 1. The "estimate" values use the formulas 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.

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.

Friedman, Milton, 2002, *Capitalism and Freedom, Fortieth Anniversary Edition* (University of Chicago Press).

Frydman, Halina, and Anna Matuszyk, 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.

Hughes, Jonathan, and Louis P. Cain, 2011, *American Economic History, Eight Edition* (Pearson Education, Inc.).

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, Vladimir Pozdnyakov, and Jun Yan, 2022, Modeling time-to-event contingent cash flows: A discrete-time survival analysis approach, *arXiv* 1, 1–41.

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.

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

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.

Roosevelt, Franklin D., 1932, Address accepting the Presidential Nomination at the Democratic National Convention in Chicago, Online by Gerhard Peters and John T. Woolley, The American Presidency Project.

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).

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.

Internet Appendix

A. *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). 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 12.

```
 1: $B \leftarrow$ bond_data                                      ▷ bond_data is a row of the loan performance data
 2: bal_vec ← each month's sequential outstanding principal balance
 3: pmt_vec ← each month's sequential actual payment
 4: prc_vec ← each month's sequential payment applied to principal
 5: init_bal ← current balance as of the first trust month
 6: paid_princ ← sum(prc_vec)                                     ▷ plus $10 pad to avoid odd tie behavior
 7: if paid_princ >= init_bal then
 8:     $D = 0$
 9:     $R = 1$
10:     $C = 0$
11:     $X \leftarrow$ location of first zero in bal_vec                              ▷ loan repaid
12: else
13:     $z \leftarrow$ starting time of three consecutive zero payments in pmt_vec
14:     if $z$ empty then
15:         $D = 0$
16:         $R = 0$
17:         $C = 1$
18:         $X \leftarrow$ length of pmt_vec                                          ▷ loan censored
19:     else
20:         $D = 1$
21:         $R = 0$
22:         $C = 0$
23:         $X \leftarrow z$                                                         ▷ loan defaults
24:     end if
25: end if
```

Figure 12: **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 (bal_vec) represented the ordered monthly balance, the second (pmt_vec) was the ordered monthly payments, and the third (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.