# On the Convergence of Credit Risk in Current Consumer Automobile Loans

Jackson P. Lautier[*†] Vladimir Pozdnyakov[‡] Jun Yan[‡]

March 14, 2024

**Abstract**

Risk-based pricing within consumer lending is ubiquitous. It considers both prevailing interest rates and the credit profile of a borrower to determine the cost of borrowing. All else equal, higher default risks pay higher borrowing costs. This cost is the annual percentage rate (APR), and it is set at the loan's origination. A borrower's credit profile is dynamic, however, and the risk of default gradually declines for current loans. In this article, we derive a novel large-sample statistical hypothesis test suitable for loans sampled from asset-backed securities to populate a credit risk transition matrix between consumer credit risk groups. We find that current loans in all risk groups eventually converge to the top credit tier before scheduled termination, a phenomenon we call *credit risk convergence*. We then use these convergence estimates for two empirical economic studies. We first estimate that lender conditional risk-adjusted expected profits significantly increase as high-risk, high-APR borrowers stay active and paying. We then estimate current borrowers are entitled to \$1,153-\$2,327 in potential credit-based savings from their improving risk profiles. Because we study consumer auto loans, a large-scale and essential economic good, we opine on the social implications of these results and suggest areas of further study.

**Keywords**: Adjustable payment loans, competing risks, discrete, prepayment, Reg AB II

[*]Department of Mathematical Sciences, Bentley University

[†]Corresponding to Jackson P. Lautier, Bentley University, Morison Hall, 175 Forest Street, Waltham, MA 02452; e-mail: jlautier@bentley.edu.

[‡]Department of Statistics, University of Connecticut

# 1 Introduction

The consumer auto lending market in the United States operates on a massive scale: consumer auto asset-backed securities (ABS) issuance tops \$200 billion (Securities Industry and Financial Markets Association, 2023) and consumer automobile debt generally exceeds \$1,400 billion (Federal Reserve, 2023). The vast majority of these loans are assigned a price under the standard practice of risk-based pricing (e.g., Edelberg, 2006; Phillips, 2013). Risk-based pricing considers both prevailing interest rates and the risk of default. All else equal, borrowers perceived to be a higher risk of default will be charged a higher borrowing cost to compensate the lender. This higher borrowing cost, which is effectively a higher annual percentage rate (APR), is set at the time the loan is originated. A borrower's instantaneous credit risk is dynamic, however, and it generally declines the longer a borrower stays active and paying. In this article, we utilise novel statistical methods to carefully study the staleness of a consumer's APR against the dynamism of a consumer's default risk.

The statistical methods we derive and employ allow us to utilise large pools of consumer automobile loans sampled from ABS (e.g., Securities and Exchange Commission, 2014). Specifically, such data is subject to incompleteness in the form of left-truncation and right-censoring, and the nature of financial loans requires a competing risks model to differentiate between defaults and prepayments. Similar statistical approaches have appeared in the study of auto loans (e.g., Heitfield and Sabarwal, 2004; Agarwal et al., 2007, 2008), but they do not operate under the assumption of discrete time. The discrete time assumption is more appropriate than continuous time for monthly auto loan data, and we assume it at the onset in deriving our asymptotic results. Furthermore, in studying the default risk hazard rate, we provide estimates for default risk in a current month conditional on survival. This is precisely a dynamic view of borrower credit risk. In organizing the data by credit risk band, we arrive at a formal statistical hypothesis test to determine the exact month a current borrower migrates into a superior credit risk band. We refer to this point as the moment of *credit risk convergence* between risk bands. Such a hypothesis test has not yet appeared to our knowledge, and it allows for us to obtain financial estimates of the theoretical cost to current borrowers that continue to pay an APR that does not dynamically adjust to their improving conditional risk profile.

With the necessary statistical methods in hand, we turn our attention to a formal empirical analysis. This is the central objective of our efforts. We find evidence that conditional credit risk converges between disparate risk bands of 72-73 month auto loans after just 12 months. Even for risk bands with large differences in their initial credit risk assessment, convergence in conditional credit risk occurs well before scheduled termination (e.g., subprime loans behave like super-prime credits after 48 months of payments). For complete risk band transition matrix details, see Section 3 and Table 1. All results withstand a series of robustness checks, the details of which may be found in the Supplemental Material. Notably, because collateralised loans on used autos have rapidly depreciating collateral values (Storchmann, 2004), these results cannot be explained by traditional loan-to-value (LTV) default behaviour expectations (e.g., Deng et al., 1996). The entirety of the methodological treatment and subsequent data analysis to estimate the point two different risk bands converge in a go-forward assessment of credit risk, i.e., the *credit risk convergence* analysis, may be found in Sections 2 and 3, respectively.

Because the APR of consumer automobile loans has a wide range (approximately 0-30%), there are significant financial implications to our empirical credit risk convergence estimates. We thus subsequently present a two-part empirical study of the financial details. The first part studies a lender's expected profitability with an actuarial analysis. That is, we use our hazard rate estimates to solve for an expected risk-adjusted rate of lender profitability conditional on loan survival. We find that lender profits are back-loaded, which is consistent with the insurance-like pricing for pools of risky loans. In other words, the high-risk, high-APR borrowers that don't default gradually become more profitable to the lender. These greater profits help compensate the lender for the high-risk, high-APR borrowers that do default. Socially, this arrangement may be viewed as a wealth transfer from high-risk, low-income borrowers to other high-risk, low-income borrowers. This runs counter to wealth redistribution schemes like progressive taxation, in which those with higher incomes make larger financial contributions to shared goods than those with lower incomes.

In the second part of our empirical financial study, we shift our focus to the consumer. We estimate the potential savings available to consumers, assuming the average borrower in one risk band refinanced at the average rate in a superior risk band, once eligible based on our credit risk convergence point estimates (ceteris paribus). We find that the riskiest borrowers (deep subprime, subprime) can potentially save between \$11-63 dollars in monthly payments or \$193-1,616 in total by refinancing. Our estimates suggest deep subprime and prime borrowers should refinance after about 42 and 50 months, respectively, when they become prime borrowers. We find evidence that these borrowers generally wait too long to refinance. In a surprise, we find that less risky loans (near-prime, prime) leave even more money on the table, with total savings ranging from \$160-2,327 (or \$13-56 in monthly payments). Our estimates suggest that near-prime and prime borrowers should refinance quickly, after about only one year, but they also generally wait too long. Hence, in a result counter to expectations about borrower sophistication, it is the near-prime and prime loans that behave less efficiently.[1] For more details, see Section 4.

We proffer the results of this article offer both potential intellectual and social benefits. Let us first discuss the former, by which we contextualise our research results. To begin, the observation that default risk for collateralised loans declines as a borrower continues to make payments is generally known. Within finance, it is the concept of *loan seasoning*, which is well-documented for residential mortgages (e.g., Adelino et al., 2019) (see also the Supplemental Material for an introduction). Our study differs from this in meaningful ways, however. First, we address loans secured by used automobiles, which is a type of collateral known to rapidly depreciate in value (Storchmann, 2004). Hence, the loan seasoning we document runs counter to traditional loan-to-value (LTV) default behaviour expectations (e.g., Campbell and Cocco, 2015). Second, it is not demonstrating that default risk declines that is our main interest. Rather, we desire to indicate the precise moment a borrower's default risk changes, and it is our novel statistical methods that allow us to do so. Third, our estimated savings to consumers are attributable to a potential credit-based refinance, which differs from the traditional interest rate-based refinance analysis (e.g., Keys et al., 2016; Agarwal et al., 2017; Andersen et al., 2020). Finally, we of course study automobiles, which is not the focus of many related mortgage studies (e.g., Deng et al., 2000; Calhoun and Deng, 2002; Ambrose and Sanders, 2003; Jones and Sirmans, 2019).

What of the potential social benefits of our work? More broadly within consumer automobile research, there is evidence that consumers are subject to various forms of troubling economic behaviour. For example, racial discrimination has been 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 behaviour within the subprime auto market. Namely, they observe sharp increases in demand during tax rebate season and high sensitivity to minimum down payment requirements. Grunewald et al. (2020) find that arrangements between auto dealers and lenders lead to incentives that increase loan prices. They also find consumers are less responsive to finance charges than vehicle charges and that consumers benefit when dealers do not have discretion to price loans. While consumer auto loans and subprime borrowers have attracted the attention of previous researchers, we again do not find consideration of the borrower risk profile over the lifespan of the loan juxtaposed against a stale APR. Within this backdrop, therefore, our results find an additional challenge to consumers with auto loans in that such borrowers struggle to recoup additional

[1] One benefit of greater affluence is the mental freedom that accompanies an ability to overpay with limited consequences. We thank Susan Woodward for this observation.

potential savings available from a credit-based refinance.

These results suggest potential socio-economic interpretations. For example, market frictions may exist that prevent both borrowers and lenders alike from reducing these suspected consumer auto refinance market inefficiencies. In hopes of encouraging related research, we will conclude with a proposal that lenders may consider offering new loan products that reward borrowers for good performance or potential regulatory interventions. Indeed, with significant technological advancements in real-time motorist driving data (e.g., Peiris et al., 2024) and health data (e.g., Sim, 2019), it is likely real-time default risk data is available such that a single point-in-time, stale APR may soon become antiquated risk pricing for consumer loans. Because high-risk, high-APR borrowers are traditionally low income and struggle financially, a consumer lending market that more dynamically prices borrowers as their risk profile changes can lead to meaningful financial savings. From a merit point-of-view, these earned savings to borrowers may be thought of as a reward for good performance. Indeed, after the failures of the global financial crisis, 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"(Obama, 2010). Hence, attempting to reward borrowers based on good performance feels aligned in spirit with an ideal of merit-based economic gains. It is our hope this study offers new meditations on this fundamental idea.

The paper proceeds as follows. We first introduce and detail the statistical methods we derive in Section 2. Section 3 is then a formal empirical analysis with our statistical methods to populate a credit risk convergence matrix between disparate risk bands. Section 4 is a further empirical study designed to estimate the financial implications of our credit risk convergence point estimates for both lenders and borrowers alike. Section 5 then concludes with an overview discussion. The Appendix provides brief additional details on the empirical results of Section 3. For an introduction to the concept of loan seasoning, proofs of major results, complete data details, a thorough robustness analysis, a simulation study, and an additional financial approach, see the Supplemental Material. For reference, all data and replication code is publicly available at the repository: https://github.com/jackson-lautier/credit-risk-convergence/.

# 2 Statistical Methods

This section comprises the methodological novelty of this work: a new financial econometric hypothesis testing technique to estimate the exact age two different risk bands converge in conditional default risk (i.e., the point of *credit risk convergence*). We begin with a review of the relevant statistical results. We will introduce the field of survival analysis along with its subfield of competing risks within the context of loan default modelling. We then present the financial econometric tools we derive in the form of an estimator, its asymptotic properties, and the resulting large sample statistical hypothesis test. The formal statements are located within this section, and we provide complete proofs in the Supplemental Material.

From an economic perspective, not all defaults are equivalent. For example, there is an obvious profitability difference between a loan that defaults shortly after it is originated versus a loan that defaults after a much longer period of time: a loan that makes more payments before defaulting will be more profitable, ceteris paribus. Therefore, we seek a time-to-event distribution estimate, where the general event of interest is the end of a loan's payments. We require this information to adequately address our research question centred around analysing a loan's conditional probability of default given its survival. We are thus in the realm of *survival analysis*, which is dedicated to estimating a random time-to-event distribution. In addition to estimating a time-to-event random variable, we also desire to distinguish between the type of event. Again, from an economic perspective, this is natural: a loan that is repaid (or prepaid) in a given month is more profitable than a loan that defaults in the same month, ceteris paribus. Succinctly, we wish to differentiate between loans ending in default and loans ending in prepayment. To do so, we can define the problem in terms of a *competing risks* framework, which is a specialised branch of survival analysis.

For completeness, our data is sampled from pools of consumer automobile loans found in publicly traded ABS (see the Supplemental Material for details). Thus, we must consider an estimator appropriately calibrated to work in both discrete-time and with incomplete data subject to random left-truncation and random right-censoring. For extended details on these incomplete data challenges with ABS applications, see the discrete time work of Lautier et al. (2023a) for the case of left-truncation and Lautier et al. (2023b) for the discrete time case of both left-truncation and right-censoring. Neither Lautier et al. (2023a) nor Lautier et al. (2023b) allow for competing risks, however. To address this need, we elect to define competing risks in terms of a multistate process,[2] which allows us to make direct estimates. Formally, we will be using a multistate process adjusted for left-truncation and right-censoring in discrete time but over a known, finite time horizon for two competing events. Hence, the major objective of this section is to generalise the discrete time, left-truncation and right-censoring work of Lautier et al. (2023b) to the case of two competing events: default and repayment.

We now present the mathematical details of the estimator in the context of an automobile loan ABS. We will follow the notation of Lautier et al. (2023b). Define the random time

[2]See Andersen et al. (1993, Example III.1.5) or Beyersmann et al. (2009) for an introduction.

until a loan contract ends by the random variable $X$. The classical quantity of interest in survival analysis is the *hazard rate*, which in discrete time represents the probability of a loan contract terminating in month $x$, given a loan has survived until at least month $x$. We denote the hazard rate by the traditional, $\lambda$, and so formally,

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

Because we desire to model the probability of loan payments terminating given a loan remains current, it is clear that (1) is the ideal quantity of interest. Additionally, let $F$ represent the cumulative distribution function (cdf) of $X$. If we can reliably estimate (1), we can recover the complete distribution of $X$ by the uniqueness of the cdf because

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

where $x_{\min}$ is the lower bound of the distribution of $X$. In (2), we take the the convention $\prod_{k=x_{\min}+1}^{x_{\min}} \{1 - \lambda(k)\} = 1$.

We now account for incomplete data. To address random left-truncation, let $Y$ represent the left-truncation random variable, which is a shifted random variable derived from the random time a loan is originated and the securitised trust begins making monthly payments. That is, we observe $X$ if and only if $X \geq Y$. We further assume $X$ and $Y$ are independent, an important assumption we now briefly justify within a securitization context. The random variable $Y$ represents the time an ABS first starts making payments. Typically, the decision to issue a securitization is more related to investment market conditions and the financing needs of the parent company than the performance of the underlying assets, in this case automobile loans. In other words, the forming and subsequent issuance of an ABS bond has little to do with the time-to-event distribution of each individual loan, which is represented by $X$.[3] Hence, the assumption that $X$ and $Y$ are independent is generally quite reasonable within the context of the securitization process. To account for right-censoring, define the censoring random variable as $C = Y + \tau$, where $\tau$ is a constant that depends on the last month the securitization is active and making monthly payments. Note that independence between $X$ and $C$ follows trivially from the assumed independence of $X$ and $Y$. We thus observe the exact loan termination time, $x$, if $x \leq C \mid X \geq Y$, and we only know that $x > C$ if $x > C \mid X \geq Y$.

For those familiar with incomplete data from observational studies, we can think of the period of time the ABS is active and paying as the observation window. Hence, random

[3] Indeed, this is the main economic motivation of the securitization process.

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

We now generalise Lautier et al. (2023b) 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 $\{Z_x\}_{\Delta+1 \leq x \leq \xi}$ be a set of random variables with probability distributions that depend on $x$, $\Delta + 1 \leq x \leq \xi$. More specifically, given a loan terminates at time $x$, we assume the loan must be in one of two states, $Z_x \in \{1, 2\}$:[4]

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

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

$$\lambda^{0i}(x) = \Pr(X = x, Z_x = i \mid X \geq x) = \frac{\Pr(X = x, Z_x = i)}{\Pr(X \geq x)}, \tag{3}$$

for $i = 1, 2$. Conveniently, therefore, from the law of total probability, we have

$$\begin{aligned} \lambda(x) = \frac{\Pr(X = x)}{\Pr(X \geq x)} &= \frac{\Pr(X = x, Z_x = 1)}{\Pr(X \geq x)} + \frac{\Pr(X = x, Z_x = 2)}{\Pr(X \geq x)} \\ &= \lambda^{01}(x) + \lambda^{02}(x). \end{aligned}$$

Within a competing risks framework, $\lambda(x)$ may be referred to as the *all-cause hazard*.[5]

Given this framework, it is not difficult to account for securitization data subject to right-censoring and left-truncation along the lines of Lautier et al. (2023b). Formally, as-

[4] It may be of help to see the related Beyersmann et al. (2009, Figure 1).

[5] The Supplemental Material provides a simulation study, which may be a helpful numerical reference of our competing risks model.

sume 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 observable data from a trust, $\{X_j, Y_j, C_j, Z_{X_j}\}_{1 \leq j \leq n}$ differs from the random variables, $\{X, Y, C, Z_X\}$. For example, the random variables $X$ and $(Y, C)$ are independent, whereas $X_j$ and $(Y_j, C_j)$ clearly are not.[6] In what follows, we will use a subscript of $\tau$ where appropriate to remind us that right-censoring is present in the data.

If we assume independence between $Y$ and the random vector $(X, Z_X)$ (reasonable given the securitzation backdrop and our earlier discussion), then we may derive estimators for (3) along the same lines as Lautier et al. (2023b). We demonstrate as follows. Let $\alpha = \Pr(Y \leq X)$ and for $i = 1, 2$, define

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

and

$$U_\tau(x) = \Pr(Y_j \leq x \leq \min(X_j, C_j)) = \frac{\Pr(Y \leq x \leq C) \Pr(X \geq x)}{\alpha}.$$

Thus,

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

In terms of our observable data, for a given loan $j$, $1 \leq j \leq n$, we observe $Y_j$, $\min(X_j, C_j)$, and $\mathbf{1}_{X_i \leq C_i}$, 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

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

[6] Lautier et al. (2023a) expounds on this point thoroughly.

we obtain the estimate for (4)

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

Pleasingly, (5) is equivalent to the related classical work of Huang and Wang (1995), despite our assumption of discrete-time at the problem's onset. It may also be shown that (5) is a *parametric* MLE in the case of left-truncated data without censoring (Lautier et al., 2024).

We are now prepared to introduce our novel financial econometric hypothesis test. For a single sample, we refer to (5) as an estimate. If we instead consider a population of all possible samples, then we may interpret (5) as an *estimator*. Under this interpretation, (5) is now a random variable, and any single estimate is just one possible realization. As such, the random variable estimator version of (5) has a number of attractive asymptotic properties. First, the complete vector of estimators over the recoverable space of $X$, $\hat{\mathbf{\Lambda}}^{0i}_{\tau,n} = (\hat{\lambda}^{0i}_{\tau,n}(\Delta+1), \ldots, \hat{\lambda}^{0i}_{\tau,n}(\xi))^{\top}$, is asymptotically unbiased for the true CSH rates. Further, $\hat{\mathbf{\Lambda}}^{0i}_{\tau,n}$ is asymptotically multivariate normal with a completely specifiable diagonal covariance structure (i.e., two estimators within $\hat{\mathbf{\Lambda}}^{0i}_{\tau,n}$ are asymptotically independent). The formal statement is as follows.

**Theorem 2.1** ($\hat{\mathbf{\Lambda}}^{0i}_{\tau,n}$ *Asymptotic Properties*)**.** *For $i \in \{1,2\}$, define $\hat{\mathbf{\Lambda}}^{0i}_{\tau,n} = (\hat{\lambda}^{0i}_{\tau,n}(\Delta+1), \ldots, \hat{\lambda}^{0i}_{\tau,n}(\xi))^{\top}$, where*

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

*Then,*

*(i)*

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

*(ii)*

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

*where $\mathbf{\Lambda}^{0i}_{\tau} = (\lambda^{0i}_{\tau}(\Delta+1), \ldots, \lambda^{0i}_{\tau}(\xi))^{\top}$ with $\lambda^{0i}_{\tau}(x) = f^{0i}_{*,\tau}(x)/U_{\tau}(x)$ and*

$$\mathbf{\Sigma}^{0i} = \operatorname{diag}\left( \frac{f^{0i}_{*,\tau}(\Delta+1)\{U_{\tau}(\Delta+1) - f^{0i}_{*,\tau}(\Delta+1)\}}{U_{\tau}(\Delta+1)^3}, \ldots, \frac{f^{0i}_{*,\tau}(\xi)\{U_{\tau}(\xi) - f^{0i}_{*,\tau}(\xi)\}}{U_{\tau}(\xi)^3} \right).$$

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

*Proof.* See the Supplemental Material. □

Additionally, we may use Theorem 2.1 to produce asymptotic confidence intervals that are appropriately bounded within $(0, 1)$. The formal statement is as follows.

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

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

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

*Proof.* See the the Supplemental Material. □

Finally, the asymptotic confidence intervals of Lemma 1 and asymptotic independence of Theorem 2.1 may be combined to form a straightforward large sample statistical hypothesis test. Formally, for two risk bands $a$, $a'$, where $a \neq a'$ (i.e., $a, a'$ would represent one of the risk bands deep subprime, subprime, near-prime, prime, or super-prime), we may test

$$H_0 : \lambda^{0i}_{\tau,(a)}(x) = \lambda^{0i}_{\tau,(a')}(x), \quad \text{v.s.} \quad H_1 : \lambda^{0i}_{\tau,(a)}(x) \neq \lambda^{0i}_{\tau,(a')}(x), \tag{7}$$

for each age $x$ by determining if the $(1-\theta)\%$ asymptotic confidence intervals of the estimators $\hat{\lambda}^{0i}_{\tau,n,(a)}(x)$ and $\hat{\lambda}^{0i}_{\tau,n,(a')}(x)$ overlap for $\Delta + 1 \leq x \leq \xi$ and $i = 1, 2$. The decision rules and interpretations are as follows. Fix $x \in \{\Delta + 1, \dots, \xi\}$ and $i = 1$. If the asymptotic confidence intervals of $\hat{\lambda}^{01}_{\tau,n,(a)}(x)$ and $\hat{\lambda}^{01}_{\tau,n,(a')}(x)$ overlap, we fail to reject $H_0$, and we cannot claim $\lambda^{01}_{\tau,(a)}(x) \neq \lambda^{01}_{\tau,(a')}(x)$. That is, conditional default risk given survival to time $x$ has potentially converged. On the other hand, if the asymptotic confidence intervals do not overlap, we reject $H_0$, and we may claim with $(1-\theta)\%$ confidence that $\lambda^{01}_{\tau,(a)}(x) \neq \lambda^{01}_{\tau,(a')}(x)$. That is, conditional default risk given survival to time $x$ has not yet converged.

# 3 Credit Risk Convergence

We now demonstrate the utility of the new financial econometric tools we derive in Section 2 through an empirical study of consumer auto ABS data. Specifically, we provide empirical estimates of the credit risk convergence points for the five standard risk bands with data sampled from ABS bonds.[7] We begin with a brief review of the data, and the section closes

[7]For reference, risk bands are commonly associated with credit score. That is, credit scores below 580 are considered *deep subprime*, credit scores between 580-619 are *subprime*, 620-659 is *near-prime*, 660-719 is

with the estimates. Where appropriate, references to additional detail in the Supplement Material (e.g., data details, robustness analysis, and simulation studies) are noted.

On September 24, 2014, the Securities and Exchange Commission (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 utilised the Electronic Data Gathering, Analysis, and Retrieval system operated by the SEC to compile complete loan-level performance data for the consumer automobile loan ABS bonds CarMax Auto Owner Trust 2017-2 (CarMax, 2017) (CARMX), Ally Auto Receivables Trust 2017-3 (Ally, 2017) (AART), Santander Drive Auto Receivables Trust 2017-2 (Santander, 2017b) (SDART), and Drive Auto Receivables Trust 2017-1 (Santander, 2017a) (DRIVE). By count, the total number of loans for CARMX, AART, SDART, and DRIVE were 55,000, 67,797, 80,636, and 72,515, respectively. The bonds were selected because of the credit profile of the underlying loans, the lack of a direct connection to a specific auto manufacturer, and the observation window of each bond's performance spanning approximately the same macroeconomic environment. We elaborate on each point in turn.

The credit profile of a DRIVE borrower is generally deep subprime to subprime, SDART is subprime to near-prime, CARMX is near-prime to prime, and AART is prime to super-prime. Thus, the collection of all four bonds taken together span the full credit spectrum of individual borrowers. Next, it is common that an auto manufacturer will originate loans using its financial subsidiary (e.g., Ford Credit Auto Owner Trust). The bonds selected do not have a direct connection to a specific auto manufacturer, however, and so we may allay concerns that our convergence point estimates are biased by oversampling loans secured by a specific brand of automobile. For completeness, we acknowledge the business objectives of CarMax, a used auto sales company, differ from the traditional banks of Santander and Ally. We sensitivity test this point in the robustness checks of the Supplemental Material. Lastly, the bonds were selected to span approximately the same months to ensure all underlying loans were subject to the same macroeconomic environment. Specifically, CARMX, AART, SDART, and DRIVE began actively paying in March, April, May, and April of 2017, respectively, and each trust was active for 50, 44, 52, and 52 months, respectively. Finally, the loans were selected to fit into a single selection criteria (i.e., loan term, col-

*prime*, and credit scores of 720 and above are *super-prime* (Consumer Financial Protection Bureau, 2019). We will increase precision by defining risk bands by the market price (i.e., annual percentage rate), but this terminology and risk association will be consistent throughout the manuscript.

lateral type, underwriting standards, etc.), and they were grouped into the five standard risk bands by interest rate per the principles of risk-based pricing (e.g., Edelberg, 2006; Phillips, 2013). Specifically, we assign 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. In total, we analyse a final set 58,118 loans selected from these four bonds. For vastly expanded data details, see the Supplemental Material. For access to the data and replication code, navigate to https://github.com/jackson-lautier/credit-risk-convergence/.

We now apply the financial econometric tools of Section 2 to this consumer auto loan data. Specifically, we both plot estimates of the CSH rates for default by loan age and risk band, $\hat{\lambda}^{01}_{\tau,n}$, and perform the hypothesis test described by (7) to the filtered loan population. For convenience of exposition, we will initially focus our discussion on two risk bands: subprime and prime borrowers. A plot of $\hat{\lambda}_{\tau,n}$ by loan age may be found in Figure 1 for the 21,332 subprime loans (solid line) and 6,300 prime loans (dashed line) of the total data analysis sample of 58,118. As an initial observation, we can see that the estimated default CSH rates for subprime loans are initially higher than the default CSH rates for prime loans. This is expected given our expectations about credit risk, risk-based pricing, and the difference in APRs between the two risk bands. This pattern does not maintain for the full lifetime of the loan, however. As the subprime loans continue to stay current (i.e., given survival), the CSH rate declines. This is an alternative visualization of loan seasoning with enhanced precision. Interestingly, the CSH rates for prime loans in this sample appear to increase slightly, though they remain generally stable even as loan age increases. Due to left-truncation and right-censoring, we are unable to fully recover the complete loan term for all risk bands. Nonetheless, we have reliable estimates from approximately $5 \leq X \leq 60$ for all risk bands, though we report $10 \leq X \leq 55$ for conservatism. In the instance of no observed defaults at a particular loan age within the recoverable window, we interpolate with a constant hazard rate.

This brings us to the major methodological result of this paper, which is the lower right corner of Figure 1. In addition to plotting the point estimates, we also provide the asymptotic confidence intervals (shaded regions surrounding each line). Eventually, as the two lines slowly approach each other, the confidence intervals begin to overlap. The first evidence of this is around loan age 42, and it is consistent by approximately loan age 50 for these 72-73 month consumer auto loans. With the statistical test outlined in (7), therefore, for any age in which we observe overlapping confidence intervals, we cannot claim the true CSH rates for default are different between the subprime and prime risk bands within this sample. It is this point at which two CSH rates for default between two different risk bands

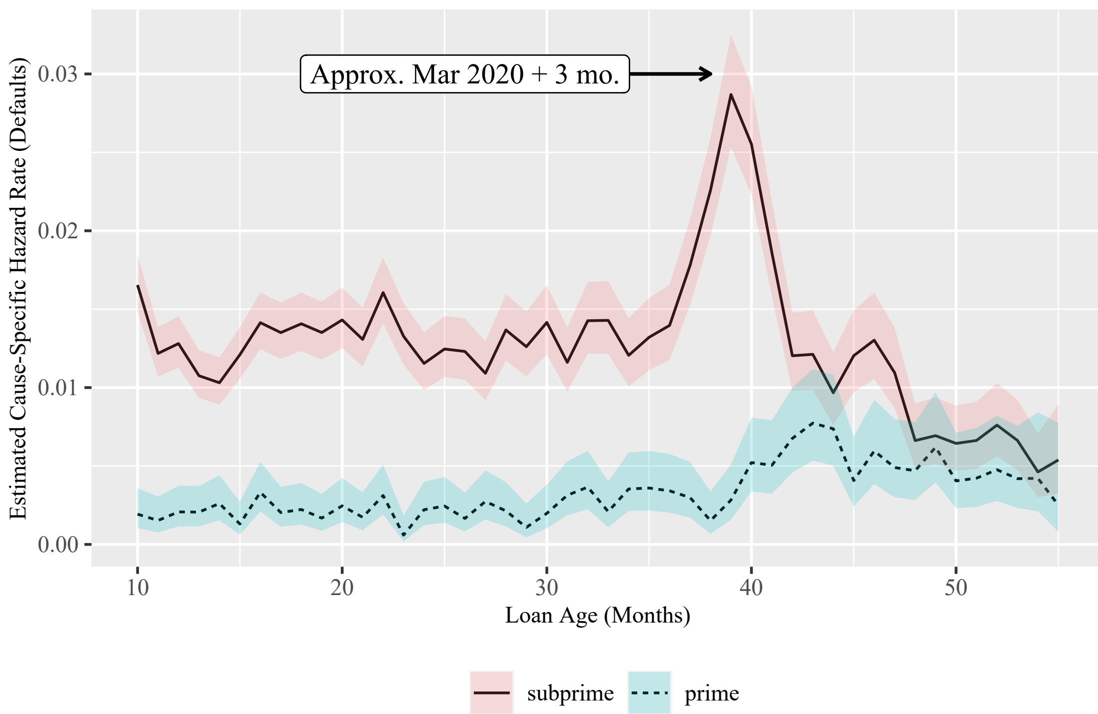


Figure 1: **Credit Risk Convergence: Subprime and Prime Loans.**
A plot of $\hat{\lambda}^{01}_{\tau,n}$ (defaults) defined in (5) by loan age for the subprime and prime risk bands within the sample of 58,188 loans, plus 95% confidence intervals using Lemma 1. We may use the hypothesis test described in (7) by searching for the minimum age that the confidence intervals overlap. In this case, we see the first evidence of credit risk convergence by approximately loan age 42 months. The upward spike in $\hat{\lambda}^{01}_{\tau,n}$ for the subprime risk band by loan age 40 is related to the economic impact of COVID-19 (for robustness analysis, see the Supplemental Material).

become statistically indistinguishable that we estimate as the point of *credit risk convergence*.

From the perspective of risk-based pricing, we find that measuring default risk conditional on survival gleans additional insight in comparison to a binary default analysis, such as that performed in the Supplemental Material. For example, 40% of all subprime loans in the sample of 58,118 defaulted at some point, versus only 10% of prime loans. Given just this analysis, it is not surprising subprime borrowers receive a higher APR than prime borrowers. What we show in Figure 1 is that the default rates conditional on survival are not constant, however, and it implies that subprime borrowers that do not refinance are eventually overpaying based on an updated assessment of their risk profile, all else equal. We come back to this point much more extensively in Section 4.2.

Table 1 (top) provides a transition matrix of the estimated month of credit risk convergence among the five risk bands considered for the sample of 58,118 filtered loans. For conservatism, we define the point of credit risk convergence as the minimum of (1) two consecutive months of confidence interval overlap after a loan age of 10 months or (2) the

Table 1: **Credit Risk Convergence: Transition Matrix.**
This table reports a summary matrix of the estimated month of credit risk convergence for the sample of 58,118 72-73 month consumer automobile loans. For conservatism, the month of credit risk convergence is defined as the earlier of (1) the first of two consecutive months after ten months that the asymptotic confidence intervals for $\hat{\lambda}^{01}_{\tau,n}$ overlap or (2) once $\hat{\lambda}^{01}_{\tau,n}$ is consistently zero for both risk bands. Visually, it is helpful to compare Figure 1 with the subprime-prime cell below. Full comparisons may be made with Figure 6 in the Appendix.

| | deep subprime | subprime | near-prime | prime | super-prime |
|---|---|---|---|---|---|
| deep subprime | 10 | 36 | 50 | 50 | 52 |
| subprime | | 10 | 23 | 42 | 48 |
| near-prime | | | 10 | 13 | 34 |
| prime | | | | 10 | 10 |
| super-prime | | | | | 10 |

minimum shared age that the hazard estimates are consistently zero. Based on these results, we would say that a deep subprime loan eventually converges in risk to a subprime loan after three years, and it converges to a prime risk after 50 months and a super-prime risk after 52 months. Similarly, subprime loans converge in risk to prime loans after 42 months, and they become super-prime risks after four years. Near-prime loans become prime risks quite quickly, just after one year, and then become super-prime risks after 34 months. For completeness, we plot the full five-by-five matrix of CSH rate and confidence interval comparisons along the lines of Figure 1 in Figure 6 found in the Appendix. For financially inclined readers, it may be of interest to recall consumer auto loans are collateralised with rapidly depreciating assets in the form of used cars (see Figure 4). In other words, these results cannot be explained by traditional LTV optionality arguments found in mortgages (e.g., Campbell and Cocco, 2015).

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

# 4 Financial Analysis

We now apply the methods of Section 2 to offer new financial perspectives on consumer auto loans. The present section proceeds in two parts. In Section 4.1, we demonstrate how the CSH estimates may be used to visualise the back-loading of a lender's expected profits. In Section 4.2, we then focus our analysis on the individual consumer. By presenting a counterfactual of a perfectly efficient borrower in terms of credit-based refinancing behaviour, we find that borrowers in all non-super-prime risk bands delay prepayment inefficiently, all else equal. In a surprise based on expectations of borrower sophistication, we find that borrowers in lower risk bands, near-prime and prime, operate less efficiently than borrowers in higher risk bands, deep subprime and subprime. Details may be found in Table 2. We also evaluate borrower conditional prepayment behaviour using the sibling estimator (5) for prepayment. In a visual analysis, we find that borrower's prepayment decisions appear to be driven by economic stimulus payments and unusual used auto markets rather than by financial sophistication. Additional details for estimating a recovery upon default assumption and broader methodological extensions may be found in the Supplemental Material.

## 4.1 Lender Profitability

A common term to describe the profit of a high-risk, high-interest-rate loan that remains current is *back-loaded*.[8] Quite simply, a high-risk, high-interest-rate loan gradually becomes more profitable as it continues paying, and it is these increased profits later in the loan's life that offset the losses taken on other similar loans that have defaulted. To provide some formality to this idea, we will utilise an actuarial approach to calculate an implied, expected risk-adjusted return for each month a loan stays current. Specifically, we will examine a rolling monthly expected annualised rate of return assuming an investor purchases a risky fixed-income asset at a price of the outstanding balance of the consumer loan at age $x$ for risk band $a$, $B_{a|x}$, with a one-month term. This hypothetical risky asset pays either (1) the outstanding balance at loan age $x+1$ for risk band $a$, $B_{a|x+1}$, plus the next month's payment

[8] We thank Jonathan A. Parker for this concise descriptive term.

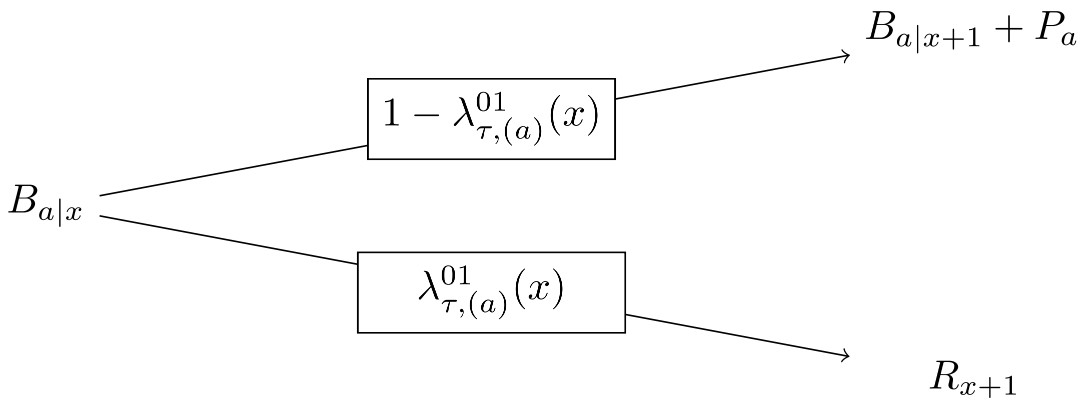


Figure 2: **Hypothetical Risky Fixed-Income Asset and Path Probabilities.**
The risky asset, $B_{a|x}$, pays either (1) the outstanding balance at loan age $x+1$ for risk band $a$, $B_{a|x+1}$, plus the next month's payment due, $P_a$, with probability $1 - \lambda^{01}_{\tau,(a)}(x)$ or (2) the recovery amount at time $x+1$ in the event of default, $R_{x+1}$, with probability $\lambda^{01}_{\tau,(a)}(x)$. The subscript $a$ denotes one of the five standard risk bands: deep subprime, subprime, near-prime, prime, or super-prime. The CSH rates for default are adjusted for prepayments by the competing risks methodology.

due, $P_a$, with probability $1 - \lambda^{01}_{\tau,(a)}(x)$ or (2) the recovery amount at time $x+1$ in the event of default, $R_{x+1}$, with probability $\lambda^{01}_{\tau,(a)}(x)$. Because we utilise a competing risks framework, the CSH rates are adjusted for prepayment probabilities. The subscript $a$ denotes one of the five standard risk bands: deep subprime, subprime, near-prime, prime, and super prime. We illustrate this hypothetical asset in Figure 2.

To calculate the annualised risk-adjusted return by month, we first define the expected present value (EPV) of a $B_{a|x}$ risky one-month loan depicted in Figure 2 as

$$\mathrm{EPV}^1_{a|x} = \lambda^{01}_{\tau,(a)}(x)\left[\frac{R_{x+1}}{1+\tilde{r}_{a|x}}\right] + (1 - \lambda^{01}_{\tau,(a)}(x))\left[\frac{B_{a|x+1} + P_a}{1+\tilde{r}_{a|x}}\right], \tag{8}$$

where $\tilde{r}_{a|x}$ is some unknown one-month effective rate of interest. To calculate the annualised risk-adjusted return, we can interpret the outstanding balance of an age $x$ loan in risk band $a$, $B_{a|x}$, as the market-implied price of a risky zero coupon bond following the payment pattern of Figure 2. Therefore, we can use (8) to solve for $\tilde{r}_{a|x}$ such that $\mathrm{EPV}^1_{a|x} = B_{a|x}$. This rate, $\tilde{r}_{a|x}$, is then the expected monthly effective risk-adjusted return, which can then be annualised.[9] The calculation in (8) also requires an estimate for the recovery upon default, $R_{x+1}$, for each age $x$. A recovery in the event of default was estimated from the sample

[9]An implicit assumption in this analysis is that the remaining payments beyond month $x+1$ are a tradable asset with no friction (i.e., the risky asset may be traded at time $x+1$ for $B_{a|x+1}$). We can instead perform an expected risk-adjusted return calculation over the entire remaining lifetime of the loan (i.e., assuming uncertainty for each future payment following the estimates in Section 3). For details, see the Supplemental Material.

of 58,118 loans of Section 3 (see the Supplemental Material for details). The probabilities, $\lambda^{01}_{\tau,(a)}(x)$, for each age, $x$, and risk band, $a$, may be estimated using the methods of Section 2.

For ease of interpretation, we consider a single loan of \$100 for 72 months with a payment and amortization schedule determined by the average APR of each risk band: deep subprime (22.65%), subprime (17.97%), near-prime (12.74%), prime (7.82%), and super-prime (3.59%). The estimated results may be found in Figure 3. In the initial stages of a loan's lifetime, the deep subprime, subprime, near-prime, and prime risk bands generally group together around 7.5%. This demonstrates that the risk-adjusted pricing is generally accurate by risk band, as the higher APRs help offset the higher default risk. It also reveals that the overall consumer auto lending market is quite efficient across risk bands at origination. The super-prime risk-band consistently hovers around a 2.5% annualised expected risk-adjusted return, which may suggest the lender secures other economic gains from these loans (e.g., reduced risk capital requirements). We then see the negative impact of COVID-19 around loan age 40, which is consistent with the discussion in Sections 3 and the Supplemental Material. It is notable that the impact on the expected risk-adjusted return for the super-prime risk band due to COVID-19 is minimal. As the loans mature, however, and credit risk convergence begins, we see the expected risk-adjusted returns for the higher APR loans begin to accelerate.

## 4.2 Consumer Perspectives

If a borrower's default risk conditional on survival declines as a loan stays current, but the loan's original APR is a single point-in-time estimate of risk at origination, then it is possible a gradual credit-based economic inefficiency from the perspective of the consumer may develop. The purpose of the present section is an attempt to quantify this inefficiency, which may be done using the techniques of Section 2. We first estimate the potential savings available to consumers by way of a comparison with the counterfactual of a perfectly efficient borrower in terms of credit-based refinancing behaviour (ceteris paribus). Next, we perform a visual analysis to observe borrower conditional prepayment behaviour over the observation period, which may also be done using the techniques of Section 2.

Before doing so, some contextualization to financial theory is helpful. Specifically, it may be tempting to trivialise the results of Table 1 as a simple artefact of collateralised loans. This line of thinking supposes that a 72-month auto loan with less than two years remaining will almost certainly be "in-the-money", and so the declining conditional default risk would naturally follow. Such reasoning ignores the rapidly depreciating value of the collateral of used automobiles, however. As a reference point, Storchmann (2004) estimates an average annual depreciation of 31% in Organization for Economic Co-operation and Development

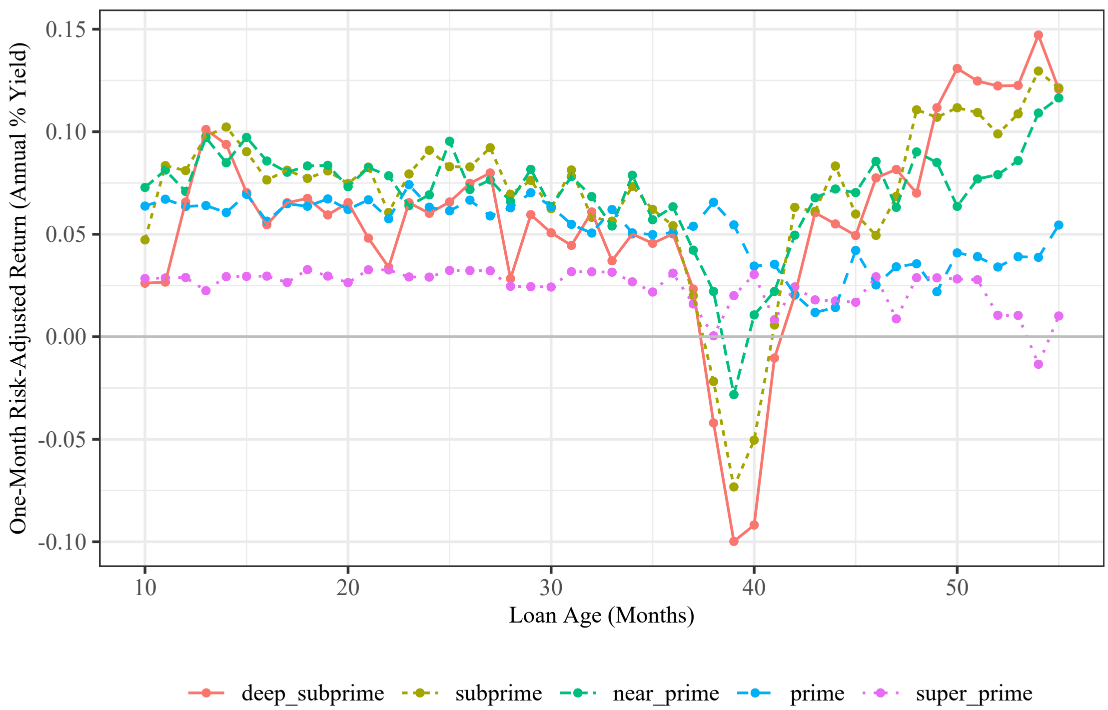


Figure 3: **Estimated Expected Rolling Risk-Adjusted Return by Age, Issuance.** A plot of the annualised, expected risk-adjusted one-month return, $\tilde{r}_{a|x}$, by loan age, risk band, and issuance year for the filtered loan sample of Section 3. The calculations utilise (8) and the two-path risky zero coupon bond formulation from Figure 2. The probabilities of each path stem from (3), and they may be estimated with (5). The one-month expected annualised risk-adjusted rate of return is roughly equal to 7.5% for the deep subprime, subprime, near-prime, and prime risk bands until the point of credit risk convergence (approximately age 40), after which the higher APR risk bands show accelerating returns. The clear negative impact of COVID-19 is also apparent near loan age 40. The CSH rates for default are adjusted for prepayments by the competing risks methodology.

(OECD) countries. Further, it is not uncommon to see deep subprime loans with APRs north of 20%,[10] hindering a borrower's ability to pay down principal. This is illustrated in Figure 4, which presents an estimated LTV by loan age for current loans in our filtered sample of 51,118 loans. It is not until loan age 60 that super prime loans finally get under an LTV of 100%, and the riskier bands possess LTVs largely north of 150-200% well beyond the convergence point estimates of Table 1. Given these estimates, it is of interest that we find conditional credit risk behaviour that cannot be explained by the standard in-the-moneyness analysis of mortgages (e.g., Deng et al., 1996), a perhaps unique economic feature of consumer auto loans. In a robustness check, we halve the 31% depreciation rate of Storchmann (2004) to 16% annually, and the deep subprime and subprime risk bands keep LTVs north of 100% beyond loan age 52, the latest convergence point in Table 1.

[10] See the Supplement Material for data details.

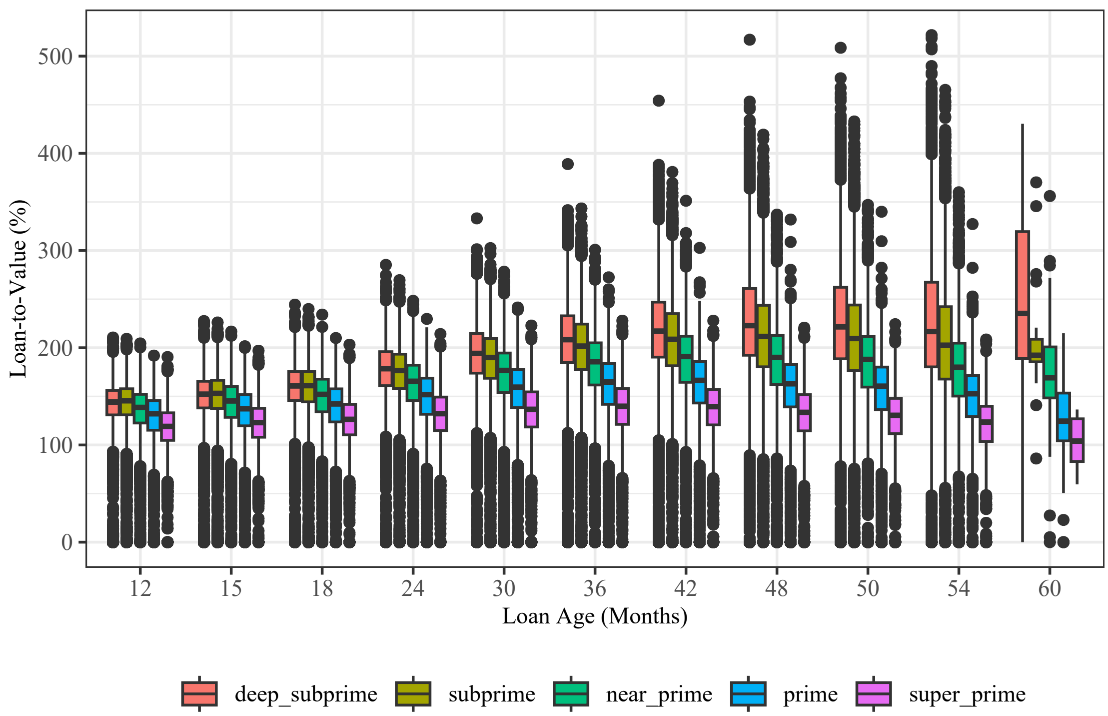


Figure 4: **Outstanding Loan-to-Value by Loan Age, Risk Band.**
A standard box plot of the outstanding LTV by loan age and risk band for current loans in the filtered sample of 51,118 loans from the ABS bonds CARMX, ALLY, SDART, and DRIVE of Section 3. Estimated LTVs remain well-above 100% past the point of credit risk convergence, which suggests that improved credit performance is not attributable to borrowers building equity in the collateral.

We estimate such potential credit-based refinance savings in Table 2, all else equal. Moving left-to-right along the column headings, we first report a count of the current loans by loan age. Next, of the active loans, we present an average outstanding balance, average payment, and average APR. The "Pmts (#)" column calculates the remaining payments needed to pay-off the average loan balance given the average payment. The next four columns represent the potential savings in monthly payment if a borrower refinances at the average APR of the superior risk band, after the estimated point of credit risk convergence. If two risk bands have not yet converged in credit risk (i.e., Table 1), the numbers are not provided in the table. The calculations do not assume any upfront refinancing charge.

We find that borrowers in all four non-super-prime risk bands, deep subprime, subprime, near-prime, and prime, appear to leave money on the table. On a monthly payment basis, deep subprime borrowers begin to overpay between \$11-63 per month around loan age 36, for a total potential savings between \$193-1,153. Based on our estimates, deep subprime borrowers would benefit the most in terms of total savings by waiting until approximately loan age 50, when they converge in risk with prime borrowers. In terms of monthly payment savings, deep subprime borrowers should wait to refinance until they converge in credit risk with super-prime borrowers. Encouragingly, we see that most deep subprime borrowers have prepaid or refinanced by loan age 60, which suggests some self-correction, albeit slower than our calculations would recommend. The situation for subprime borrowers is similar; they benefit the most in total savings by refinancing by loan age 42, when they converge in credit risk with prime borrowers. Overall, the potential total savings over the life of the loan for subprime borrowers ranges between \$299-1,616. In terms of monthly payment, subprime borrowers benefit the most by waiting until loan age 48, when they converge in credit risk to super-prime borrowers. In total, the potential monthly payment savings for subprime borrowers ranges between \$22-61. As with deep subprime borrowers, it seems most have refinanced by loan age 60. While this is slower than our calculations would suggest, it still indicates borrowers may be attempting to self-correct. These results would exacerbate any consumer refinance inefficiency attributable to changes in interest rates (e.g., Keys et al., 2016; Agarwal et al., 2017; Andersen et al., 2020).

In moving to discuss borrowers in superior risk bands, we find slightly different results. As with deep subprime and subprime borrowers, we also find evidence that near-prime and prime borrowers operate inefficiently with respect to a credit-based refinance, ceteris paribus. We estimate that near-prime borrowers are eligible for a potential monthly payment savings of \$13-56 with a potential total savings of \$160-2,206. The figures for prime borrowers are similar; a potential \$18-39 in monthly savings with a potential total savings of \$261-2,327. On the other hand, we find that both near-prime and prime borrowers should refinance as soon as possible, after 15 months for near-prime borrowers when they converge in credit risk with prime borrowers and after 12 months for prime borrowers when they converge in credit risk with super-prime borrowers. We find that both near-prime and prime borrowers do not start refinancing in earnest until approximately loan age 60, similar to borrowers in the higher risk bands. This suggests that near-prime and prime borrowers manage their loans less efficiently than deep subprime and subprime borrowers, a result that is surprising given typical expectations about borrower sophistication and credit score.[11]

It is of further interest to examine loan prepayment behaviour, which is also possible with the techniques of Section 2. Specifically, the CSH rate estimator defined in (5), $\hat{\lambda}^{02}_{\tau,n}$, is a direct estimator for prepayment behaviour, also conditional on survival. Hence, we can report similar figures to Section 3 but instead focus on borrower prepayment behaviour conditioning on the set of current loans. From this, we can attempt to explain consumer

[11]It may be that the greater affluence of near-prime and prime borrowers allows a non-optimal efficiency to persist out of the perceived inconvenience of going through a refinancing versus the potential savings. We thank Susan Woodward for this observation. It also of interest to compare this finding with the profitability analysis of FHA-insured mortgages in Deng and Gabriel (2006).

Table 2: **Estimated Savings by Risk Band, Loan Age.**
Estimated savings assuming the counterfactual of a perfectly efficient borrower who refinances at the average interest rate of a superior risk band immediately after the point of credit risk convergence in Table 1. Abbreviations: S = subprime, NP = near-prime, P = prime, and SP = super-prime.

| | | | Averages | | | | Mo Pmt Savings ($) | | | | Total Savings ($) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Age | Cnt | Bal ($) | Pmt ($) | APR (%) | Pmts (#) | S | NP | P | SP | S | NP | P | SP |
| deep subprime | 12 | 17,558 | 14,245 | 365 | 22.58 | 65 | | | | | | | | |
| | 15 | 16,125 | 13,844 | 364 | 22.56 | 62 | | | | | | | | |
| | 18 | 14,375 | 13,520 | 363 | 22.54 | 60 | | | | | | | | |
| | 24 | 11,628 | 12,836 | 361 | 22.50 | 56 | | | | | | | | |
| | 30 | 9,492 | 11,973 | 361 | 22.46 | 50 | | | | | | | | |
| | 36 | 7,746 | 10,985 | 359 | 22.46 | 44 | 16 | | | | 586 | | | |
| | 42 | 6,050 | 9,833 | 357 | 22.46 | 38 | 16 | | | | 490 | | | |
| | 48 | 4,899 | 8,799 | 358 | 22.43 | 33 | 18 | | | | 438 | | | |
| | 50 | 4,622 | 8,312 | 358 | 22.44 | 30 | 12 | 33 | 52 | | 267 | 729 | 1,153 | |
| | 54 | 3,568 | 7,485 | 360 | 22.37 | 26 | 11 | 30 | 47 | 61 | 193 | 531 | 845 | 1,093 |
| | 60 | 12 | 6,923 | 377 | 22.00 | 23 | 21 | 39 | 54 | 63 | 251 | 466 | 643 | 759 |
| subprime | 12 | 18,261 | 16,693 | 395 | 17.97 | 64 | | | | | | | | |
| | 15 | 17,021 | 16,126 | 394 | 17.96 | 61 | | | | | | | | |
| | 18 | 15,487 | 15,619 | 393 | 17.95 | 59 | | | | | | | | |
| | 24 | 12,997 | 14,621 | 389 | 17.94 | 54 | | 32 | | | | 1,557 | | |
| | 30 | 11,021 | 13,420 | 388 | 17.94 | 48 | | 30 | | | | 1,275 | | |
| | 36 | 9,309 | 12,194 | 386 | 17.94 | 42 | | 25 | | | | 904 | | |
| | 42 | 7,481 | 10,835 | 384 | 17.93 | 37 | | 29 | 54 | | | 857 | 1,616 | |
| | 48 | 6,192 | 9,506 | 383 | 17.92 | 31 | | 22 | 44 | 61 | | 526 | 1,055 | 1,473 |
| | 50 | 5,901 | 8,953 | 383 | 17.93 | 29 | | 23 | 44 | 60 | | 508 | 963 | 1,325 |
| | 54 | 4,542 | 7,975 | 386 | 17.94 | 25 | | 22 | 40 | 55 | | 389 | 723 | 988 |
| | 60 | 22 | 7,021 | 414 | 17.47 | 20 | | 25 | 40 | 50 | | 299 | 477 | 596 |
| near-prime | 12 | 5,807 | 19,111 | 411 | 12.79 | 64 | | | | | | | | |
| | 15 | 5,587 | 18,245 | 407 | 12.76 | 60 | | | 39 | | | | 2,206 | |
| | 18 | 5,315 | 17,617 | 405 | 12.74 | 58 | | | 40 | | | | 2,158 | |
| | 24 | 4,692 | 16,204 | 402 | 12.72 | 52 | | | 35 | | | | 1,657 | |
| | 30 | 4,146 | 14,694 | 400 | 12.71 | 47 | | | 37 | | | | 1,546 | |
| | 36 | 3,592 | 13,187 | 398 | 12.71 | 41 | | | 31 | 56 | | | 1,116 | 2,000 |
| | 42 | 3,041 | 11,446 | 394 | 12.67 | 35 | | | 28 | 49 | | | 847 | 1,481 |
| | 48 | 2,622 | 9,862 | 394 | 12.68 | 29 | | | 21 | 39 | | | 494 | 928 |
| | 50 | 2,455 | 9,283 | 395 | 12.69 | 27 | | | 20 | 37 | | | 436 | 811 |
| | 54 | 1,663 | 8,218 | 400 | 12.69 | 24 | | | 29 | 44 | | | 526 | 798 |
| | 60 | 63 | 6,435 | 413 | 11.98 | 17 | | | 13 | 22 | | | 160 | 269 |
| prime | 12 | 5,173 | 18,582 | 358 | 7.83 | 64 | | | | 39 | | | | 2,327 |
| | 15 | 5,283 | 17,611 | 354 | 7.81 | 60 | | | | 33 | | | | 1,880 |
| | 18 | 5,315 | 16,706 | 350 | 7.78 | 57 | | | | 30 | | | | 1,627 |
| | 24 | 4,971 | 15,097 | 346 | 7.76 | 52 | | | | 32 | | | | 1,535 |
| | 30 | 4,538 | 13,503 | 345 | 7.74 | 46 | | | | 30 | | | | 1,245 |
| | 36 | 4,096 | 11,866 | 344 | 7.73 | 39 | | | | 21 | | | | 755 |
| | 42 | 3,697 | 10,274 | 342 | 7.72 | 34 | | | | 23 | | | | 703 |
| | 48 | 3,191 | 8,615 | 343 | 7.71 | 28 | | | | 21 | | | | 513 |
| | 50 | 2,963 | 8,101 | 345 | 7.71 | 26 | | | | 21 | | | | 460 |
| | 54 | 1,898 | 7,075 | 351 | 7.66 | 22 | | | | 18 | | | | 324 |
| | 60 | 92 | 4,756 | 328 | 7.38 | 16 | | | | 22 | | | | 261 |

behaviour and assess if borrowers are acting on the potential savings reported in Table 2. For context, we also overlay two additional economic variables. The first is the seasonally adjusted Manheim Used Auto Price Index (Manheim, 2023), which is a common industry assessment of the prevailing value of used automobiles. Given the unusual observations in the used auto market during the COVID-19 pandemic (Rosenbaum, 2020), it is possible that

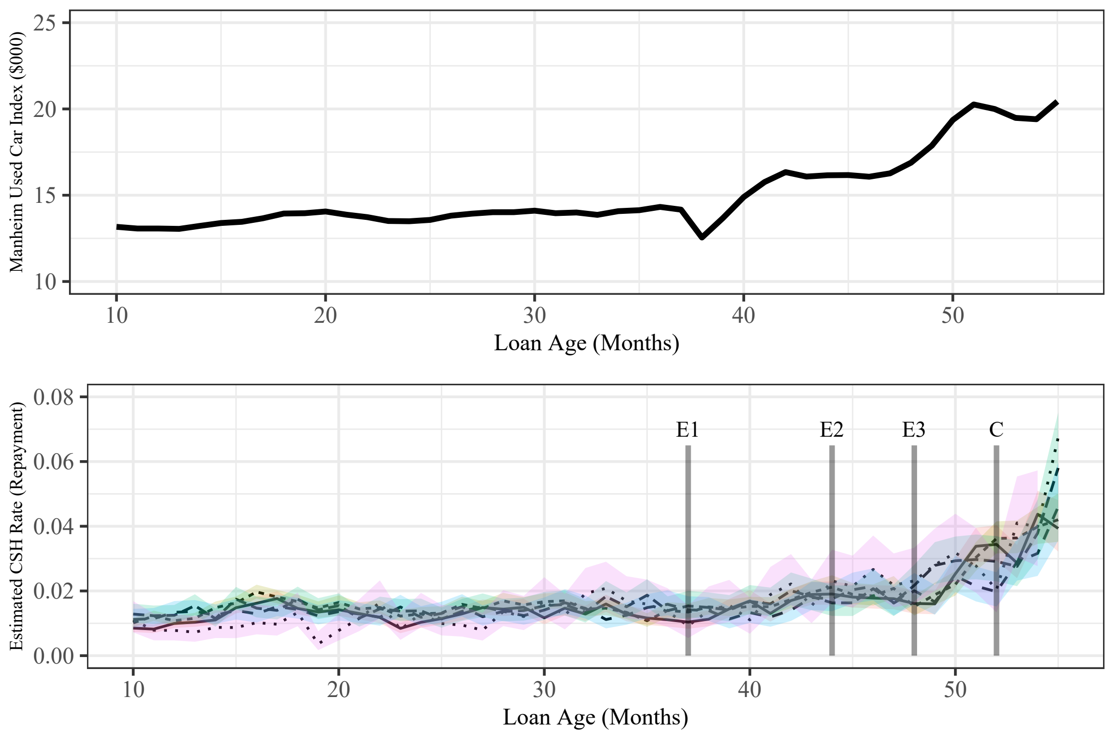


Figure 5: **Consumer Prepayment behaviour, Used Autos, Economic Stimulus.** (top) A plot of the Manheim Used Auto Index (price) (Manheim, 2023) by approximate loan age for the sample of 58,118 filtered loans of Section 3. (bottom) A plot of $\hat{\lambda}^{02}_{\tau,n}$ (prepayments) defined in (5) by loan age for all risk bands within the sample of 58,118 filtered loans of Section 3, plus 95% confidence intervals using Lemma 1. By the hypothesis test defined in (7), there is very little difference in prepayment behaviour conditional on survival by risk band. The labels E1, E2, E3, and C indicate the timing of the Economic Impact Payments and Childcare Tax Credit expansion (U.S. Government Accountability Office, 2022).

higher-than-expected trade-in values motivated consumers to prepay their loans. Additionally, the United States federal government provided individuals with three direct payments known as Economic Impact Payments (EIPs) and expanded the Childcare Tax Credit (CTC) during the observation period of our sample (U.S. Government Accountability Office, 2022). It is thus possible that borrowers, upon receiving these cash payments, made the decision to purchase a different vehicle and thus prepay. The results are presented in Figure 5.

There appears to be very little difference in prepayment behaviour by risk band throughout the life of the loan, which differs significantly from default rates (compare Figure 5 with Figures 1 and 6). Further, there does appear to be a meaningful connection between prevailing used auto prices and borrower prepayment behaviour. That is, as the value of used autos rose, borrowers of current loans appear to increase prepayment frequency. Furthermore, the timing of economic stimulus payments plotted against prepayment behaviour is also telling. The prepayment rates increase shortly after individuals would have received the first direct EIP from the U.S. federal government. Because of the potential savings we observe in Table 2, it is possible that the EIPs may have also provided individuals with further implicit economic gains, if they used the EIPs to refinance at a lower, credit-based interest rate. The results of Figure 5 in connection with Table 2 taken together suggest that individual borrowers may not consider their updated risk profile in deciding to prepay. Instead, the borrowers may be more motivated by economic indicators that are more tangible, such as direct cash payments or higher trade-in values.

# 5 Discussion

Conventional profitability wisdom of risk-based pricing from the perspective of a lender is that the high-returns of high-risk loans that don't default help offset the losses from the high-risk loans that do default. In other words, there is an implied insurance arrangement in which the cost of the losses are dispersed among the individual borrowers. Furthermore, it can be argued that through its precision, risk-based pricing has been attributed to lowering the cost of credit for a majority of borrowers and expanding credit availability to higher risk borrowers (Staten, 2015).[12] These are positive economic outcomes, and we do not attempt to argue against the overall practice of risk-based pricing at loan origination. Within the scope of our analysis, two comments are warranted. First, all loans considered herein have been sampled from pools of securitised bonds. Hence, the risk of default has already been transferred off the lender's balance sheet after the point of sale. In other words, the lender no longer has a direct financial interest in any of the loans we study. Second, we believe it is reasonable to put forth a nuanced argument in light of our results and the current practices of risk-based pricing. That is, the consumer auto loan market is capable of operating more efficiently with respect to a dynamic view of default risk. Because we study consumer auto loans with a scale north of $1,400 billion (Federal Reserve, 2023) collateralised against an essential economic asset, we find the social implications of a better attuned, dynamic risk-based pricing system within consumer auto lending to be potentially meaningful.

This article puts forth this social argument built upon a three-part story. The first part is statistical. Specifically, we estimate the point of credit risk convergence between disparate risk bands using a novel financial econometric hypothesis test we derive via large-sample asymptotic statistics from the field of survival analysis. The second part is an empirical analysis of lender risk-adjusted expected profitability. Given our statistical estimates, we find that high-risk, high-APR loans that do not default eventually become quite profitable to the lender on an updated, risk-adjusted basis. The third part is an empirical analysis of potential savings available to a consumer, assuming the consumer refinanced at the superior credit risk

[12] See also Livshits (2015) for a more thorough introduction to risk-based pricing.

band rate immediately after the point of credit risk convergence. We find borrowers are slow to recoup these savings across all risk bands. The average income in each risk band of the sample we study is super-prime ($98,162), prime ($65,559), near-prime ($56,746), subprime ($46,064), and deep subprime ($41,093). Taken together, we conjecture that current risk-based pricing practices, in which a borrower receives an APR after a single point-in-time risk assessment, creates a social wealth redistribution system in which high-risk, high-APR borrowers that don't default offset losses from high-risk, high-APR borrowers that do default. Given the financial difficulties of such borrowers, these insurance-like risk-based pricing pool arrangements appear to run counter to more progressive wealth redistribution schemes.

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

It may not be fair to blame this perceived borrower inefficiency solely on the borrowers, however. A borrower's main tool to assess creditworthiness is their credit score. While consumers have obtained better access to credit scores, they may update too slowly within the context of a 72-73 month consumer auto loan to motivate a borrower to seek out a lower rate. Additionally, such borrowers may face friction in attempting to refinance mature auto loans, either through limited options, refinancing fees, or perceived hassle. Indeed, encouraging borrowers to self-correct has proven to be less effective in practice (e.g., Keys et al., 2016; Agarwal et al., 2017). From this point of view, we see an opportunity for outside lenders to target these mature loans from borrowers in higher risk bands. Because a borrower that stays current eventually outperforms their initial risk profile and loan APRs are constant throughout the life of the loan, there likely exists a lower rate that would both lower this borrower's financing cost and be profitable to a second lender. There are examples of speciality finance companies in the student loan space that attempt to refinance borrowers into lower interest rates (e.g., SoFi). The size (and potential profitability) of such loans may be larger than auto loans, however. In addition, given all students loans are originally subject to the same underwriting standards and the wide disparity of ultimate educational outcomes, the level of risk mispricing is likely more egregious and thus easier to exploit than for auto loans. On the other hand, lenders themselves may face similar market frictions, such as an inability to identify these borrowers or unattractive returns after accounting for the full scope of origination costs.[13] We are optimistic that continued increases in financial technology may lower these possible hurdles for both borrowers and lenders.

To spur future research, we suggest two potential solutions. The first is that we see a market ripe for financial innovation. Specifically, we propose that lenders offer a loan structure with a reducing payment based on good performance, an *adjustable payment loan.* It is likely lenders already possess the data needed to provide pricing structures capable of adjusting for a borrower's updated risk profile. This is especially true given real-time data advancements in other industries (e.g., Sim, 2019; Peiris et al., 2024). 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). Such an approach may be of interest to speciality finance companies connected to *responsible investing* (i.e., environmental, social, and governance (ESG), socially responsible investing (SRI), or impact investing).

Second, there is always the regulatory angle, which has been successful in other consumer lending spaces (e.g., Stango and Zinman, 2011; Agarwal et al., 2014). For example, there is potentially minimal additional cost to lenders to require ongoing loans to be underwritten again after a set period of good performance, say 36 months, especially given the lender will already have most of the borrower's information. Ideally, this update would not count as a formal inquiry against the borrower's credit report. Encouragingly, sending reminder notices about refinancing has had some success (e.g., Byrne et al., 2023). Further, given Figure 5, an initial cash payment incentive to borrowers may provide sufficient motivation to get borrowers to refinance. The overall economic impact of such a program may be mixed, given the ambivalent results for the "cash for clunkers" program (Mian and Sufi, 2012). Alternatively, competing lenders themselves may offer cash to borrowers in exchange for refinancing. On the other hand, regulatory intervention to increase the cost of lending may lead to these extra costs being pushed back to the borrowers.

In closing, our theoretical and empirical findings complement each other to establish a new framing of lending practices within consumer automobile loans. Specifically, we believe borrowers may be better served, especially those that are traditionally low-income and finan-

[13]We thank Chellappan Ramasamy for drawing our attention to the nuances of refinancing auto loans.

cially at-risk. Perhaps most notably, we argue on behalf of the borrowers that stay current. In other words, we contend the consumer auto lending market is capable of better serving borrowers that have "earned it". To this end, we close with a restatement of the words of former United States President Barack Obama at the signing of the Dodd-Frank Wall Street Reform and Consumer Protection Act (Obama, 2010),

> "We all win when folks are rewarded based on how well they perform."

# Appendix: Section 3 Supplement

We plot the full five-by-five matrix of CSH rate estimates for default in Figure 6 for the sample of 58,118 loans of Section 3. It is a complete extension of the subprime versus prime plot in Figure 1. That is, Figure 1 is a zoomed-in view of the subprime-prime cell (row 4, column 2) in Figure 6. By comparing the asymptotic confidence intervals within each risk band comparison by loan age, we may estimate the point of credit risk convergence. Figure 6 is a visualization of the point estimates summarised in Table 1.

# Acknowledgements

We thank conference participants at the 2023 Joint Statistical Meetings (Toronto), Fifth Biennial Auto Lending Conference (Federal Reserve Bank of Philadelphia's SURF, CFI), the 2023 New England Statistics Symposium (Boston University), the 2023 Boulder Summer Conference on Consumer Financial Decision Making (University of Colorado Boulder Leeds School of Business) and seminar participants at Bentley University, the University of Connecticut (School of Business, Department of Finance), the University of Connecticut (Department of Economics), Université Concordia University (John Molson School of Business, Department of Finance), and helpful comments from Joseph Golec, Brian Melzer, Jonathan A. Parker, and Jonathan Zinman.

# Funding

Jackson P. Lautier's work was supported by a National Science Foundation Graduate Research Fellowship under Grant No. DHE 1747453 and the W. Michael Hoffman Center for Business Ethics at Bentley University.

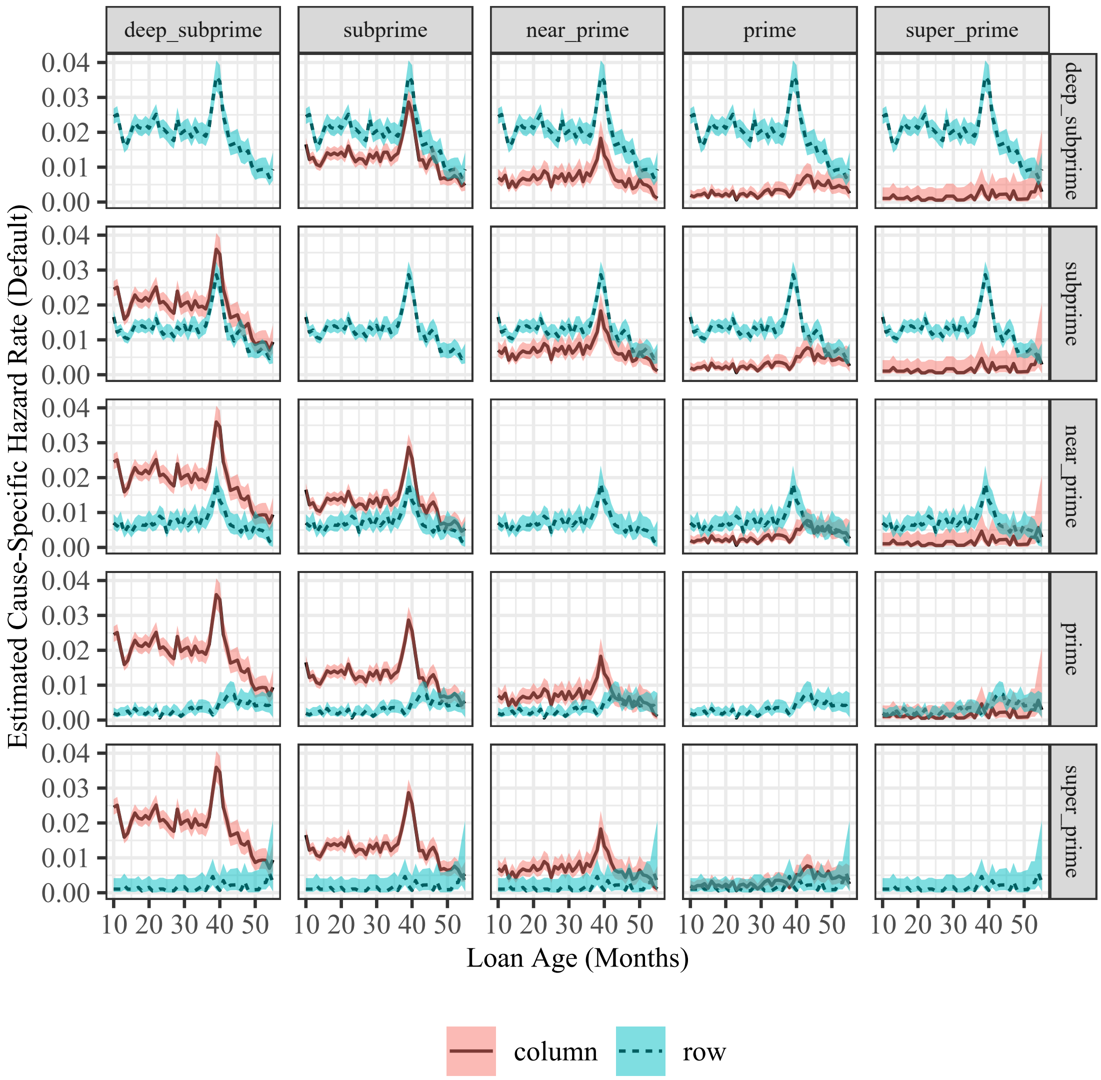


Figure 6: **Credit Risk Convergence: All Risk Bands (2017).**
A plot of $\hat{\lambda}^{01}_{\tau,n}$ (defaults) defined in (5) by loan age for all five risk bands within the sample of 58,118 loans of Section 3, plus 95% confidence intervals using Lemma 1 to compare with Table 1 via (7).

# Data Availability

All data and replication code is publicly available at the repository: https://github.com/jackson-lautier/credit-risk-convergence/.

# Supplemental Material

Please see the Supplemental Material for a brief introduction to loan seasoning, proofs of major results, additional data details, a robustness analysis, a simulation study, and an alternative method to estimate lender risk-adjusted return.

# References

W. Adams, L. Einav and J. Levin (2009). "Liquidity constraints and imperfect information in subprime lending." *American Economic Review* **99**, 49–84.

M. Adelino, K. Gerardi and B. Hartman-Glaser (2019). "Are lemons sold first? Dynamic signaling in the mortgage market." *Journal of Financial Economics* **132**, 1–25.

S. Agarwal, B. W. Ambrose and S. Chomsisengphet (2007). *Asymmetric Information and the Automobile Loan Market*, pp. 93–116. New York: Palgrave Macmillan US.

S. Agarwal, B. W. Ambrose and S. Chomsisengphet (2008). "Determinants of automobile loan default and prepayment." *Economic Perspectives* **32**, 17–28.

S. Agarwal, I. Ben-David and V. Yao (2017). "Systematic mistakes in the mortgage market and lack of financial sophistication." *Journal of Financial Economics* **123**, 42–58.

S. Agarwal, S. Chomsisengphet, N. Mahoney and J. Stroebel (2014). "Regulating consumer financial products: Evidence from credit cards." *The Quarterly Journal of Economics* **130**, 111–164.

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

B. W. Ambrose and A. B. Sanders (2003). "Commercial mortgage-backed securities: Prepayment and default." *The Journal of Real Estate Finance and Economics* **26**, 179–196.

P. K. Andersen, Ø. Borgan, R. D. Gill and N. Keiding (1993). *Statistical Models Based on Counting Processes*. Springer.

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

I. Ayres and P. Siegelman (1995). "Race and gender discrimination in bargaining for a new car." *The American Economic Review* **85**, 304–321.

J. Beyersmann, A. Latouche, A. Buchholz and M. Schumacher (2009). "Simulating competing risks data in survival analysis." *Statistics in Medicine* **28**, 956–971.

A. W. Butler, E. J. Mayer and J. P. Weston (2022). "Racial disparities in the auto loan market." *The Review of Financial Studies* **36**, 1–41.

S. Byrne, K. Devine, M. King, Y. McCarthy and C. Palmer (2023). "The last mile of monetary policy: Inattention, reminders, and the refinancing channel." Working Paper 31043, National Bureau of Economic Research.

C. A. Calhoun and Y. Deng (2002). "A dynamic analysis of fixed- and adjustable-rate mortgage terminations." *The Journal of Real Estate Finance and Economics* **24**, 9–33.

J. Y. Campbell (2016). "Restoring rational choice: The challenge of consumer financial regulation." *American Economic Review* **106**, 1–30.

J. Y. Campbell and J. F. Cocco (2015). "A model of mortgage default." *The Journal of Finance* **70**, 1495–1554.

CarMax (2017). "CarMax Auto Owner Trust." Prospectus 2017-2, CarMax Business Services LLC.

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).

Y. Deng and S. Gabriel (2006). "Risk-based pricing and the enhancement of mortgage credit availability among underserved and higher credit-risk populations." *Journal of Money, Credit and Banking* **38**, 1431–1460.

Y. Deng, J. M. Quigley and R. Van Order (2000). "Mortgage terminations, heterogeneity and the exercise of mortgage options." *Econometrica* **68**, 275–307.

Y. Deng, J. M. Quigley, R. Van Order and F. Mac (1996). "Mortgage default and low downpayment loans: The costs of public subsidy." *Regional Science and Urban Economics* **26**, 263–285. Proceedings of the Conference "Public Policy and the Housing Market".

W. Dobbie, A. Liberman, D. Paravisini and V. Pathania (2021). "Measuring bias in consumer lending." *The Review of Economic Studies* **88**, 2799–2832.

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

W. Edelberg (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.).

Federal Reserve (2023). "Statistical Release, Consumer Credit (G.19)." Url: https://www.federalreserve.gov/releases/g19/current/ (Accessed: 2023-07-14).

D. B. Gross and N. 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.

A. Grunewald, J. A. Lanning, D. C. Low and T. Salz (2020). "Auto dealer loan intermediation: Consumer behavior and competitive effects." Working Paper 28136, National Bureau of Economic Research.

P. Heidhues and B. Kőszegi (2016). "Naïeté-based discrimination." *The Quarterly Journal of Economics* **132**, 1019–1054.

E. Heitfield and T. Sabarwal (2004). "What drives default and prepayment on subprime auto loans?" *The Journal of Real Estate Finance and Economics* **29**, 457–477.

Y. Huang and M.-C. Wang (1995). "Estimating the occurrence rate for prevalent survival data in competing risks models." *Journal of the American Statistical Association* **90**, 1406–1415.

T. Jones and G. S. Sirmans (2019). "Understanding subprime mortgage default." *Journal of Real Estate Literature* **27**, 27–52.

H. J. Karger (2003). "No deals on wheels: How and why the poor pay more for basic transportation." *Journal of Poverty* **7**, 93–112.

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

J. P. Lautier, V. Pozdnyakov and J. Yan (2023a). "Estimating a discrete distribution subject to random left-truncation with an application to structured finance." *Econometrics and Statistics* Forthcoming.

J. P. Lautier, V. Pozdnyakov and J. Yan (2023b). "Pricing time-to-event contingent cash flows: A discrete-time survival analysis approach." *Insurance: Mathematics and Economics* **110**, 53–71.

J. P. Lautier, V. Pozdnyakov and J. Yan (2024). "On the maximum likelihood estimation of a discrete, finite support distribution under left-truncation and competing risks." *Statistics & Probability Letters* **207**, 109973.

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

A. Lusardi and C. de Bassa Scheresberg (2013). "Financial literacy and high-cost borrowing in the United States." Working Paper 18969, National Bureau of Economic Research.

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

A. Mian and A. Sufi (2012). "The effects of fiscal stimulus: Evidence from the 2009 Cash for Clunkers program." *The Quarterly Journal of Economics* **127**, 1107–1142.

B. Obama (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.

H. Peiris, H. Jeong, J.-K. Kim and H. Lee (2024). "Integration of traditional and telematics data for efficient insurance claims prediction." *ASTIN Bulletin* p. 1–17.

R. Phillips (2013). "Optimizing prices for consumer credit." *Journal of Revenue & Pricing Management* **12**.

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

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.

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 S 229.1125 (Item 1125) Schedule AL - Asset-Level Information."

Securities Industry and Financial Markets Association (2023). "US ABS securities: Issuance, trading volume, outstanding." Url: https://www.sifma.org/resources/research/us-asset-backed-securities-statistics/ (Accessed: 2023-06-02).

I. Sim (2019). "Mobile devices and health." *New England Journal of Medicine* **381**, 956–968. PMID: 31483966.

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

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

K. Storchmann (2004). "On the depreciation of automobiles: An international comparison." *Transportation* **31**, 371–408.

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

# Credit Risk Convergence: Supplemental Material

The following is intended as an online companion supplement to the manuscript, *On the convergence of credit risk in current consumer automobile loans*. Please attribute any citations to the original manuscript. This companion includes a brief introduction to loan seasoning, proofs of major results, additional data details, a robustness analysis, a simulation study, and an alternative method to estimate lender risk-adjusted return. All data and replication code is publicly available at the repository: https://github.com/jackson-lautier/credit-risk-convergence/.

## A Loan Seasoning

In chronicling cumulative loss curves for securitization pools of individual consumer automobile loans, there is a familiar pattern every junior credit analyst can sketch from memory: an initial rise in the early months of the securitization followed by a sustained flattening in the curve once the pool eventually settles into its long-term steady state. In higher risk or *subprime* pools of borrowers, the eventual cumulative loss percentage might be many multiples higher than lower risk or *prime* pools of borrowers, but the overall shape follows the familiar natural log-like pattern. Junior credit analysts are trained to look for any sudden upward deviations in the historical pattern, or *peel back*, which may indicate a rapid deterioration in the performance of the loans. We illustrate three such securitization loss curves in Figure A1. It is peculiar that the loss curves all eventually flatten to a similar degree. This suggests an eventual equivalence in the instantaneous default rate conditional on survival, despite the notable cumulative differences between the loss curves. This is the concept of *loan seasoning*, which is documented for residential mortgages (e.g., Adelino et al., 2019).

## B Proofs: Section 2

*Proof of Theorem 2.1.* Statement *(i)* follows from *(ii)*, so it is enough to show *(ii)*. Denote $v_1 \wedge v_2 \equiv \min(v_1, v_2)$ for any $v_1, v_2 \in \mathbb{R}$ and let $\Delta + 1 \leq k \leq \xi$. Observe

$$\begin{aligned}
\hat{\lambda}^{0i}_{\tau,n}(k) - \lambda^{0i}_{\tau}(k) &= \frac{\frac{1}{n}\sum_{j=1}^{n} \mathbf{1}_{X_j \leq C_j} \mathbf{1}_{Z_{X_j}=i} \mathbf{1}_{X_j \wedge C_j = k}}{\hat{U}_{\tau,n}(k)} - \frac{f^{0i}_{*,\tau}(k)}{U_\tau(k)} \\
&= \frac{\{\sum_{j=1}^{n} \mathbf{1}_{X_j \leq C_j} \mathbf{1}_{Z_{X_j}=i} \mathbf{1}_{X_j \wedge C_j = k}\} U_\tau(k) - f^{0i}_{*,\tau}(k)\hat{U}_{\tau,n}(k)}{\hat{U}_{\tau,n}(k) U_\tau(k)} \\
&= \frac{\sum_{j=1}^{n} \{\mathbf{1}_{X_j \leq C_j} \mathbf{1}_{Z_{X_j}=i} \mathbf{1}_{X_j \wedge C_j = k} U_\tau(k) - f^{0i}_{*,\tau}(k) \mathbf{1}_{Y_j \leq k \leq X_j \wedge C_j}\}}{n\hat{U}_{\tau,n}(k) U_\tau(k)}.
\end{aligned}$$

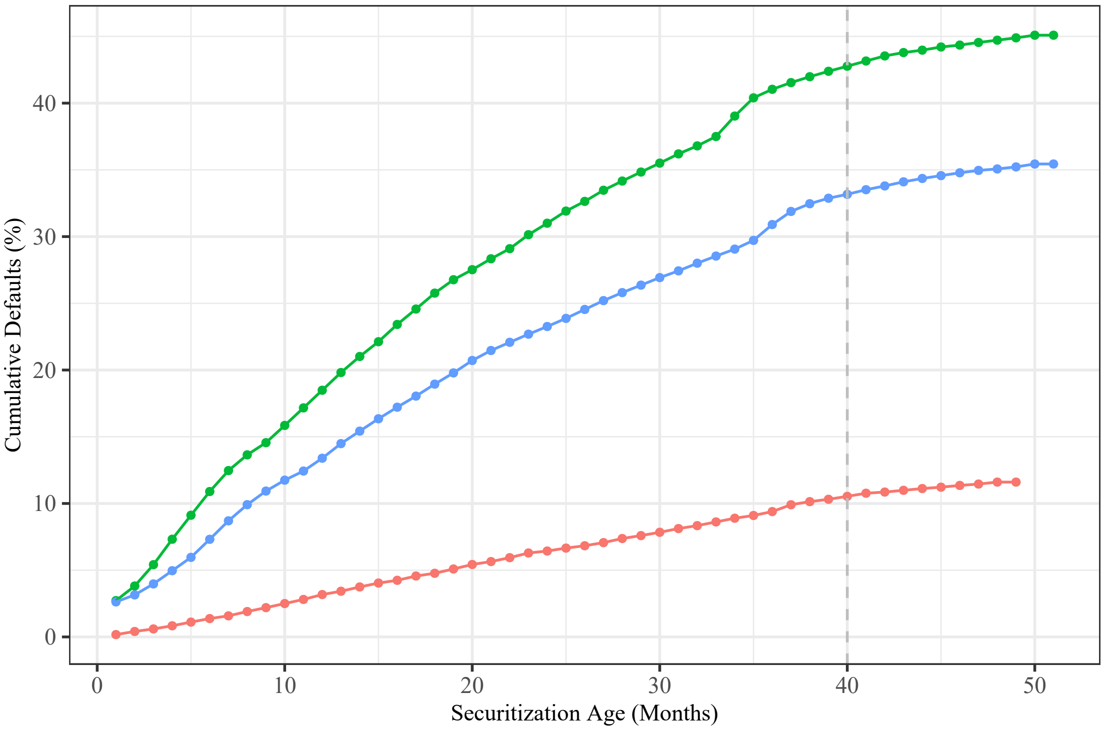


Figure A1: **Classical Consumer Automobile Securitization Loss Curves.**
A plot of three securitization loss curves: the cumulative count (%) of defaults against securitization age (months). The higher two loss curves correspond to riskier (i.e., *subprime*) pools of loans in terms of traditional credit metrics, and the lower is a less risky, *prime* pool. It is a curiosity that all curves eventually flatten (e.g., after the vertical dashed line at 40 months), despite the large differences in underlying borrower credit quality and rapid deterioration of collateral value.

Define, for $1 \leq j \leq n$, $H^{0i}_{\tau,k(j)} = \mathbf{1}_{X_j \leq C_j} \mathbf{1}_{Z_{X_j}=i} \mathbf{1}_{X_j \wedge C_j = k} U_\tau(k) - f^{0i}_{*,\tau}(k) \mathbf{1}_{Y_j \leq k \leq X_j \wedge C_j}$, and $\mathbf{A}_{\tau,n} = \operatorname{diag}([\hat{U}_{\tau,n}(\Delta+1)U_\tau(\Delta+1)]^{-1}, \ldots, [\hat{U}_{\tau,n}(\xi)U_\tau(\xi)]^{-1})$. Then,

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

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

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

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

$$\mathrm{Cov}[H^{0i}_{\tau,k(j)}, H^{0i}_{\tau,k'(j)}] = \begin{cases} U_\tau(k) f^{0i}_{*,\tau}(k)[U_\tau(k) - f^{0i}_{*,\tau}(k)], & k = k' \\ 0, & k \neq k'. \end{cases} \tag{1}$$

To see this, observe $\mathbf{1}_{X_j \le C_j} \mathbf{1}_{Z_{X_j}=i} \mathbf{1}_{X_j \wedge C_j = k}$ and $\mathbf{1}_{Y_j \le k \le X_j \wedge C_j}$ are Bernoulli random variables with probability parameters $f^{0i}_{*,\tau}(k)$ and $U_\tau(k)$, respectively. Hence,

$$\begin{aligned} \mathbf{E} H^{0i}_{\tau,k(j)} &= \mathbf{E}\mathbf{1}_{X_j \le C_j} \mathbf{1}_{Z_{X_j}=i} \mathbf{1}_{X_j \wedge C_j = k} U_\tau(k) - f^{0i}_{*,\tau}(k) \mathbf{E}\mathbf{1}_{Y_j \le k \le X_j \wedge C_j} \\ &= f^{0i}_{*,\tau}(k) U_\tau(k) - f^{0i}_{*,\tau}(k) U_\tau(k) = 0. \end{aligned}$$

Therefore,

$$\begin{aligned} \mathrm{Cov}[H^{0i}_{\tau,k(j)}, H^{0i}_{\tau,k'(j)}] &= \mathbf{E} H^{0i}_{\tau,k(j)} H^{0i}_{\tau,k'(j)} \\ &= \mathbf{E}\{\mathbf{1}_{X_j \le C_j} \mathbf{1}_{Z_{X_j}=i} \mathbf{1}_{X_j \wedge C_j = k} U_\tau(k) - f^{0i}_{*,\tau}(k) \mathbf{1}_{Y_j \le k \le X_j \wedge C_j}\} \\ &\quad \times \{\mathbf{1}_{X_j \le C_j} \mathbf{1}_{Z_{X_j}=i} \mathbf{1}_{X_j \wedge C_j = k'} U_\tau(k') - f^{0i}_{*,\tau}(k') \mathbf{1}_{Y_j \le k' \le X_j \wedge C_j}\} \\ &= U_\tau(k) U_\tau(k') \mathbf{E}\mathbf{1}_{X_j \le C_j} \mathbf{1}_{Z_{X_j}=i} \mathbf{1}_{X_j \wedge C_j = k} \mathbf{1}_{X_j \le C_j} \mathbf{1}_{Z_{X_j}=i} \mathbf{1}_{X_j \wedge C_j = k'} \\ &\quad - U_\tau(k) f^{0i}_{*,\tau}(k') \mathbf{E}\mathbf{1}_{X_j \le C_j} \mathbf{1}_{Z_{X_j}=i} \mathbf{1}_{X_j \wedge C_j = k} \mathbf{1}_{Y_j \le k' \le X_j \wedge C_j} \\ &\quad - U_\tau(k') f^{0i}_{*,\tau}(k) \mathbf{E}\mathbf{1}_{X_j \le C_j} \mathbf{1}_{Z_{X_j}=i} \mathbf{1}_{X_j \wedge C_j = k'} \mathbf{1}_{Y_j \le k \le X_j \wedge C_j} \\ &\quad + f^{0i}_{*,\tau}(k) f^{0i}_{*,\tau}(k') \mathbf{E}\mathbf{1}_{Y_j \le k \le X_j \wedge C_j} \mathbf{1}_{Y_j \le k' \le X_j \wedge C_j}. \end{aligned}$$

We shall calculate $\mathrm{Cov}[H^{0i}_{\tau,k(j)}, H^{0i}_{\tau,k'(j)}]$ by cases, $k = k'$ and $k \neq k'$. Suppose first $k = k'$ and observe

$$\begin{aligned} \mathbf{E}\mathbf{1}_{X_j \le C_j} \mathbf{1}_{Z_{X_j}=i} \mathbf{1}_{X_j \wedge C_j = k} \mathbf{1}_{X_j \le C_j} \mathbf{1}_{Z_{X_j}=i} \mathbf{1}_{X_j \wedge C_j = k'} &= \mathbf{E}\mathbf{1}_{X_j \le C_j} \mathbf{1}_{Z_{X_j}=i} \mathbf{1}_{X_j \wedge C_j = k} \\ &= f^{0i}_{*,\tau}(k), \end{aligned}$$

$$\begin{aligned} \mathbf{E}\mathbf{1}_{X_j \le C_j} \mathbf{1}_{Z_{X_j}=i} \mathbf{1}_{X_j \wedge C_j = k} \mathbf{1}_{Y_j \le k' \le X_j \wedge C_j} &= \mathbf{E}\mathbf{1}_{X_j \le C_j} \mathbf{1}_{Z_{X_j}=i} \mathbf{1}_{X_j \wedge C_j = k'} \mathbf{1}_{Y_j \le k \le X_j \wedge C_j} \\ &= \mathbf{E}\mathbf{1}_{X_j \le C_j} \mathbf{1}_{Z_{X_j}=i} \mathbf{1}_{X_j \wedge C_j = k} \mathbf{1}_{Y_j \le k \le X_j \wedge C_j} \\ &= \mathbf{E}\mathbf{1}_{X_j \le C_j} \mathbf{1}_{Z_{X_j}=i} \mathbf{1}_{X_j \wedge C_j = k} \\ &= f^{0i}_{*,\tau}(k), \end{aligned}$$

and $\mathbf{E}\mathbf{1}_{Y_j\le k\le X_j\wedge C_j}\mathbf{1}_{Y_j\le k'\le X_j\wedge C_j}=\mathbf{E}\mathbf{1}_{Y_j\le k\le X_j\wedge C_j}=U_\tau(k)$. Thus,

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

For the second case, $k\neq k'$, we have instead

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

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

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

and $\mathbf{E}\mathbf{1}_{Y_j\le k\le X_j\wedge C_j}\mathbf{1}_{Y_j\le k'\le X_j\wedge C_j}=\Pr(Y_j\le k\le X_j\wedge C_j,Y_j\le k'\le X_j\wedge C_j)$. Thus, denoting $v_1\vee v_2=\max(v_1,v_2)$ for any $v_1,v_2\in\mathbb{R}$, we have

$$\begin{aligned}\mathrm{Cov}[H^{0i}_{\tau,k(j)},H^{0i}_{\tau,k'(j)}]=f^{0i}_{*,\tau}(k\wedge k')\times&\\ \Big\{-U_\tau(k\vee k')&\Pr(X_j\le C_j,Z_{X_j}=i,X_j\wedge C_j=k\vee k',Y_j\le k\wedge k')\\ &+f^{0i}_{*,\tau}(k\vee k')\Pr(Y_j\le k\le X_j\wedge C_j,Y_j\le k'\le X_j\wedge C_j)\Big\}.\end{aligned}$$

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

$$U_\tau(k\vee k')=\Pr(Y_j\le k\vee k'\le X_j\wedge C_j)=\frac{\Pr(Y\le k\vee k'\le C)\Pr(X\ge k\vee k')}{\alpha},$$

$$\begin{aligned}&\Pr(X_j\le C_j,Z_{X_j}=i,X_j\wedge C_j=k\vee k',Y_j\le k\wedge k')\\&\quad=\frac{\Pr(X=k\vee k',Z_{X_j}=i)\Pr(Y\le k\wedge k',C\ge k\vee k')}{\alpha},\end{aligned}$$

$$f_{*,\tau}^{0i}(k \vee k') = \frac{\Pr(X = k \vee k', Z_{X_j} = i) \Pr(Y \leq k \vee k' \leq C)}{\alpha},$$

and

$$\begin{aligned}&\Pr(Y_j \leq k \leq X_j \wedge C_j, Y_j \leq k' \leq X_j \wedge C_j)\\ &\quad = \frac{\Pr(Y \leq k \wedge k', C \geq k \vee k') \Pr(X \geq k \vee k')}{\alpha}.\end{aligned}$$

Therefore,

$$\begin{aligned}&U_\tau(k \vee k') \Pr(X_j \leq C_j, Z_{X_j} = i, X_j \wedge C_j = k \vee k', Y_j \leq k \wedge k')\\ &\quad = f_{*,\tau}^{0i}(k \vee k') \Pr(Y_j \leq k \leq X_j \wedge C_j, Y_j \leq k' \leq X_j \wedge C_j),\end{aligned}$$

and so $\mathrm{Cov}[H_{\tau,k(j)}^{0i}, H_{\tau,k'(j)}^{0i}] = 0$ when $k \neq k'$. This confirms (1). Now define

$$\mathbf{D}_\tau^{0i} = \mathrm{diag}\begin{bmatrix} U_\tau(\Delta+1) f_{*,\tau}^{0i}(\Delta+1)[U_\tau(\Delta+1) - f_{*,\tau}^{0i}(\Delta+1)] \\ \vdots \\ U_\tau(\xi) f_{*,\tau}^{0i}(\xi)[U_\tau(\xi) - f_{*,\tau}^{0i}(\xi)] \end{bmatrix}$$

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,

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

Next, define $\mathbf{V}_\tau = \mathrm{diag}(U_\tau(\Delta+1)^{-2}, \ldots, U_\tau(\xi)^{-2})$. By Lautier et al. (2023, Lemma 1), $\mathbf{A}_{\tau,n} \xrightarrow{\mathcal{P}} \mathbf{V}_\tau$, as $n \to \infty$. Thus, by the multivariate version of Slutsky's Theorem (Lehmann and Casella, 1998, Theorem 5.1.6, pg. 283),

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

Observe $\mathbf{V}_\tau \mathbf{D}_\tau^{0i} \mathbf{V}_\tau^\top = \mathbf{\Sigma}^{0i}$ and $\mathbf{A}_{\tau,n}\bar{\mathbf{H}}_{\tau,n}^{0i} = \hat{\mathbf{\Lambda}}_{\tau,n}^{0i} - \mathbf{\Lambda}_\tau^{0i}$ to complete the proof. □

*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 $i = 1, 2$,

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

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. □

# C Data Details

We shall first review how the loans were selected and how the risk bands were defined. Next, we summarise the selected loans. We then include details on the definition of a default. Finally, we provide details on how we estimated the recovery given default.

## C.1 Loan Selection and Defining Risk Bands

All data derives from the consumer automobile loan ABS bonds CarMax Auto Owner Trust 2017-2 (CarMax, 2017) (CARMX), Ally Auto Receivables Trust 2017-3 (Ally, 2017) (AART), Santander Drive Auto Receivables Trust 2017-2 (Santander, 2017b) (SDART), and Drive Auto Receivables Trust 2017-1 (Santander, 2017a) (DRIVE).

To ensure the underlying loans in our analysis are as comparable as possible, we employ a number of filtering mechanisms. First, we remove any loan contracts that include a co-borrower. Second, we require each loan to have been underwritten to the level of "stated not verified" (`obligorIncomeVerificationLevelCode`), which is a prescribed description of the amount of verification done to a borrower's stated income level on an initial loan application (Securities and Exchange Commission, 2016). Third, we remove all loans originated with any form of subvention (i.e., additional financial incentives, such as added trade-in compensation or price reductions on the final sale price). We then require all loans to correspond to the sale of a used vehicle. This was mainly to keep the loans from CARMX, of which used cars predominate. We sensitivity test this requirement in the robustness checks of Section D.2. We further drop any loan with a current status of "repossessed" as of the first available reporting month of the corresponding ABS. Further, to minimise the chance of inadvertently including a loan that has been previously refinanced or modified, we only consider loans younger than 18 months as of the first available ABS reporting month. For loan term, we only include loans with an original term of 72 or 73 months. Pragmatically, the most common loan term in the data was 72/73 months, and so our loan term choice allows us to maximise the sample size.

As a final data integrity check, we remove 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 impacts only 2,630 or 4.3% of the filtered loan population. We are left with 58,118 individual consumer auto loan contracts in total, summary details of which may be found in Section C.2.

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

We now formalise this discussion slightly. Working from Phillips (2013), a borrower's interest rate in risk band $a$, $r_a$, is

$$r_a = r_c + m + l_a,$$

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

$$l_a \equiv f(\text{PTI}, \%\text{ down}, \text{Loan Amt}, \text{Vehicle Val}, \ldots).$$

Rather than attempt to recover this unknown $f$, therefore, we are in effect treating the lender's credit scoring model as an accurate reflection of the borrower's risk.[1] Specifically, we assign 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. In a review of Figure C1 in Section C.2, we can see that the risk bands assigned by interest rate compare favourably to the traditional credit score

[1] Indeed, these models are often quite sophisticated (Einav et al., 2012).

borrower risk band definition (Consumer Financial Protection Bureau, 2019).

## C.2 Summary of Selected Loans

After the data cleaning and filtering of Section C.1, we have payment performance for 58,118 consumer auto loans that span a wide range of borrower credit quality based on the traditional credit score metric. Figure C1 presents a summary of each bond by obligor credit score and interest rate as of loan origination. Judging by credit score, we can see that generally DRIVE is a deep subprime to subprime pool of borrowers, SDART is a subprime to near-prime pool, CARMX is a near-prime to prime pool, and AART is a prime to super-prime pool of borrowers (Consumer Financial Protection Bureau, 2019). As expected, in a risk-based pricing framework, the density plot of each borrower's interest rate has an inverse relationship to the density plot of each borrower's credit score: lower credit scores correspond to higher interest rates (compare the first two rows of Figure C1). As such, we can see the annual percentage rates (APRs) are higher for the DRIVE and SDART bonds, generally sitting within a range around 20% and then declining to under 15% for CARMX and finally under 10% for AART. The bottom two rows of Figure C1 demonstrate that defining risk bands by interest rate corresponds closely to the traditional credit score risk band definitions (Consumer Financial Protection Bureau, 2019), as the expected inverse relationship holds.

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

Table C1 provides a summary of borrower counts by bond and performance. The total pool of 58,118 loans is weighted towards deep subprime and subprime borrowers, which are each 37% of the total and together 74%. Similarly, DRIVE and SDART supply around 85% of the total loans in our sample. The smallest risk band is super-prime, which totals 2,179 loans for 4% of the total of 58,118. Our asymptotic results scale by sample size, so the confidence intervals adjust appropriately.

In terms of loan performance, we can observe some clear trends in Table C1. First, more than half of all deep subprime risk band loans defaulted,[2] and this percentage declines by risk band until super-prime, in which only 4% of loans defaulted during the observation

[2]We use a strict definition of default in that three consecutive missed payments is a default. This was defined within our code (see Section C.3) to ensure a consistent default definition.

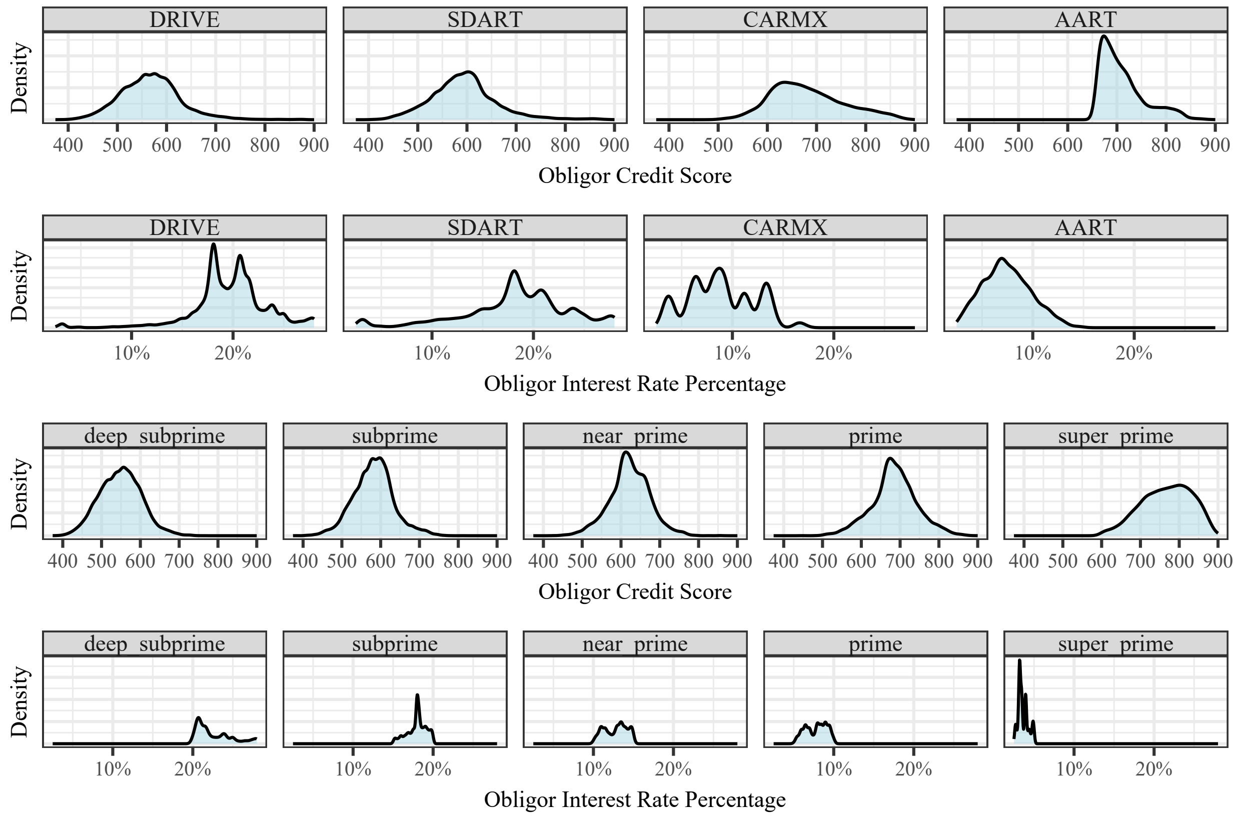


Figure C1: **Borrower Credit Profile and APR by Bond, Risk Band.** Borrower credit profiles (1st row) and charged APR (2nd row) of the 58,118 filtered consumer automobile loans used in the analysis of Sections 3 and 4 by ABS bonds CarMax Auto Owner Trust 2017-2 (CarMax, 2017) (CARMX, 6,835), Ally Auto Receivables Trust 2017-3 (Ally, 2017) (AART, 2,171), Santander Drive Auto Receivables Trust 2017-2 (Santander, 2017b) (SDART, 20,192), and Drive Auto Receivables Trust 2017-1 (Santander, 2017a) (DRIVE, 28,920). Distribution of credit scores (3rd row) and interest rates (4th row), by APR-based risk band classification: super-prime (0-5%), prime (5-10%), near-prime (10-15%), subprime (15-20%), and deep subprime (20%+) for the same set of 58,118 loans.

window. We also see that performance is fairly consistent by risk band, even among different bonds. For example, super-prime default percentages are within a tight range (3-6%) across each bond. The same may be said for deep subprime defaults. We see some wider ranges in the default percentages of the subprime (33-40%), prime (8-19%), and near-prime (17-24%) risk bands by bond, but they remain close enough to suggest there is not a worrisome difference between the credit scoring models employed by each different issuer. Overall, the percentage of defaulted loans declines as the credit quality of the risk band increases. This is further evidence that our APR-based risk band definition has yielded appropriate classification results.

Table C1: **Borrower Counts by Risk Band, Bond, and Loan Outcome.**
This table reports the summary statistics and loan outcomes of the 58,118 filtered consumer automobile loans summarised in Figure C1. Table specific abbreviations are DRIVE (DRV), SDART (SDT), CARMX (CMX) and AART (AAT). Percentages may not total to 100% due to rounding.

| | | deep subprime | subprime | near-prime | prime | super-prime | Total |
|---|---|---|---|---|---|---|---|
| | Total | 21,630 (37%) | 21,332 (37%) | 6,677 (11%) | 6,300 (11%) | 2,179 (4%) | 58,118 (100%) |
| | DRIVE | 14,079 (65%) | 12,884 (60%) | 1,443 (22%) | 220 (3%) | 294 (13%) | 28,920 (50%) |
| | SDART | 7,551 (35%) | 8,327 (39%) | 2,782 (42%) | 861 (14%) | 671 (31%) | 20,192 (35%) |
| | CARMX | 0 (0%) | 120 (1%) | 2,128 (32%) | 3,752 (60%) | 835 (38%) | 6,835 (12%) |
| | AART | 0 (0%) | 1 (0%) | 324 (5%) | 1,467 (23%) | 379 (17%) | 2,171 (4%) |
| | Total | 21,630 (100%) | 21,332 (100%) | 6,677 (100%) | 6,300 (100%) | 2,179 (100%) | 58,118 (100%) |
| | Defaulted | 11,210 (52%) | 7,900 (37%) | 1,422 (21%) | 624 (10%) | 92 (4%) | 21,248 (37%) |
| | Censored | 3,547 (16%) | 4,599 (22%) | 1,997 (30%) | 2,556 (41%) | 948 (44%) | 13,647 (23%) |
| | Repaid | 6,873 (32%) | 8,833 (41%) | 3,258 (49%) | 3,120 (50%) | 1,139 (52%) | 23,223 (40%) |
| | Total | 21,630 (100%) | 21,332 (100%) | 6,677 (100%) | 6,300 (100%) | 2,179 (100%) | 58,118 (100%) |
| DRV | Defaulted | 7,518 (53%) | 5,115 (40%) | 351 (24%) | 42 (19%) | 14 (5%) | 13,040 (45%) |
| DRV | Censored | 2,214 (16%) | 2,641 (20%) | 324 (22%) | 60 (27%) | 119 (40%) | 5,358 (19%) |
| DRV | Repaid | 4,347 (31%) | 5,128 (40%) | 768 (53%) | 118 (54%) | 161 (55%) | 10,522 (36%) |
| | Total | 14,079 (100%) | 12,884 (100%) | 1,443 (100%) | 220 (100%) | 294 (100%) | 28,920 (100%) |
| SDT | Defaulted | 3,692 (49%) | 2,740 (33%) | 590 (21%) | 105 (12%) | 29 (4%) | 7,156 (35%) |
| SDT | Censored | 1,333 (18%) | 1,915 (23%) | 715 (26%) | 255 (30%) | 299 (45%) | 4,517 (22%) |
| SDT | Repaid | 2,526 (33%) | 3,672 (44%) | 1,477 (53%) | 501 (58%) | 343 (51%) | 8,519 (42%) |
| | Total | 7,551 (100%) | 8,327 (100%) | 2,782 (100%) | 861 (100%) | 671 (100%) | 20,192 (100%) |
| CMX | Defaulted | 0 | 45 (38%) | 427 (20%) | 296 (8%) | 25 (3%) | 793 (12%) |
| CMX | Censored | 0 | 43 (36%) | 854 (40%) | 1,736 (46%) | 392 (47%) | 3,025 (44%) |
| CMX | Repaid | 0 | 32 (27%) | 847 (40%) | 1,720 (46%) | 418 (50%) | 3,017 (44%) |
| | Total | 0 | 120 (100%) | 2,128 (100%) | 3,752 (100%) | 835 (100%) | 6,835 (100%) |
| AAT | Defaulted | 0 | 0 (0%) | 54 (17%) | 181 (12%) | 24 (6%) | 259 (12%) |
| AAT | Censored | 0 | 0 (0%) | 104 (32%) | 505 (34%) | 138 (36%) | 747 (34%) |
| AAT | Repaid | 0 | 1 (100%) | 166 (51%) | 781 (53%) | 217 (57%) | 1,165 (54%) |
| | Total | 0 | 1 (100%) | 324 (100%) | 1,467 (100%) | 379 (100%) | 2,171 (100%) |

## C.3 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 aggregate 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 proceeds as follows. For each remaining bond after the filtering of Section C.1, we extract three vectors, each of which was the same length as the number of months a trust was active and paying. The first vector represents the ordered monthly balance, the second is the ordered monthly payments, and the third is the ordered monthly amount of payment applied to principal. We then consider 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 is 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 is less than the first month's outstanding loan balance, we then consider a loan outcome to be either right-censored or defaulted. To make this determination, we search the monthly payments received vector for three consecutive zeros (i.e., three straight months of missed payments). If we find three consecutive missed payments, we assume the loan to be defaulted with a time-of-default set to be the month in which the first of three zeros is observed. If we do not find three consecutive months of missed payments, the loan is 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 C2.

## C.4 Estimating Recovery Upon Default

Consumer auto loans are secured with the collateral of the attached automobile. In the event of a defaulted loan, the lender has legal standing to repossess the vehicle to make up the outstanding balance of the loan. In most cases, particularly for deep subprime and subprime borrowers, the estimated value of a repossessed automobile in the event of default is an important component in the initial pricing of a loan. In this section, therefore, we briefly discuss our process to estimate a recovery assumption by loan age, which is ultimately defined as a percentage of the initial loan balance. Our estimates are used in the analysis of Section 4.1, but we acknowledge the empirical results may also be of interest to readers more generally. We thus present our estimated recovery curve for the 2017 issuance (see Section C.2) in Figure C3.

The results of Figure C3 utilise the detailed reporting of the loan level data of Securities and Exchange Commission (2016) to perform the estimation for both the filtered sample of 58,118 loans issued in 2017 and summarised in Section C.2 and the filtered sample of 65,802 loans issued in 2019 and summarised in Section D.1. Specifically, we calculate 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). We then divide the total `recoveredAmount` by the `originalLoanAmount` for each defaulted loan.

```
B ← bond_data                                   ▷ bond_data is a row of the loan performance data
bal_vec ← each month's sequential outstanding principal balance
pmt_vec ← each month's sequential actual payment
prc_vec ← each month's sequential payment applied to principal
init_bal ← current balance as of the first trust month
paid_princ ← sum(prc_vec)                       ▷ plus $10 pad to avoid odd tie behaviour
if paid_princ >= init_bal then
    D = 0
    R = 1
    C = 0
    X ← location of first zero in bal_vec      ▷ loan repaid
else
    z ← starting time of three consecutive zero payments in pmt_vec
    if z empty then
        D = 0
        R = 0
        C = 1
        X ← length of pmt_vec                   ▷ loan censored
    else
        D = 1
        R = 0
        C = 0
        X ← z                                   ▷ loan defaults
    end if
end if
```

Figure C2: **Determination of Loan Outcome.**
We first extract three vectors, each of which is the same length as the number of months the trust was active and paying. The first vector (`bal_vec`) represents the ordered monthly balance, the second (`pmt_vec`) is the ordered monthly payments, and the third (`prc_vec`) is the ordered monthly amount of payment applied to principal. We consider a loan to be repaid if the sum total principal received is greater than the outstanding loan balance as of the first month the trust was actively paying. In this case, the timing of a repayment is set to be the first month with a zero outstanding principal balance. If the sum total principal received is less than the first month's outstanding loan balance, we consider a loan outcome to be either right-censored or defaulted. To make this determination, we search the monthly payments received vector for three consecutive zeros (i.e., three straight months of missed payments). If we find three consecutive missed payments, we assume the loan to be defaulted with a time-of-default set to be the month in which the first of three zeros is observed. If we do not find three consecutive months of missed payments, the loan is assumed to be a right-censored observation and assigned an event time as of the last month the trust was actively paying.

Finally, we take an average of these recovery percentages by age of default in months. The point estimates may be found in Figure C3. Next, for convenient use within the lender profitability analysis of Section 4.1, we nonparametrically smooth the point estimates using the `loess()` function in `R` (R Core Team, 2022). See the dashed line in Figure C3. This nonparametric `loess` curve is then fitted to a gamma-kernel via ordinary minimization of a sum-of-squared differences, which allows for extrapolation beyond the recoverable sample

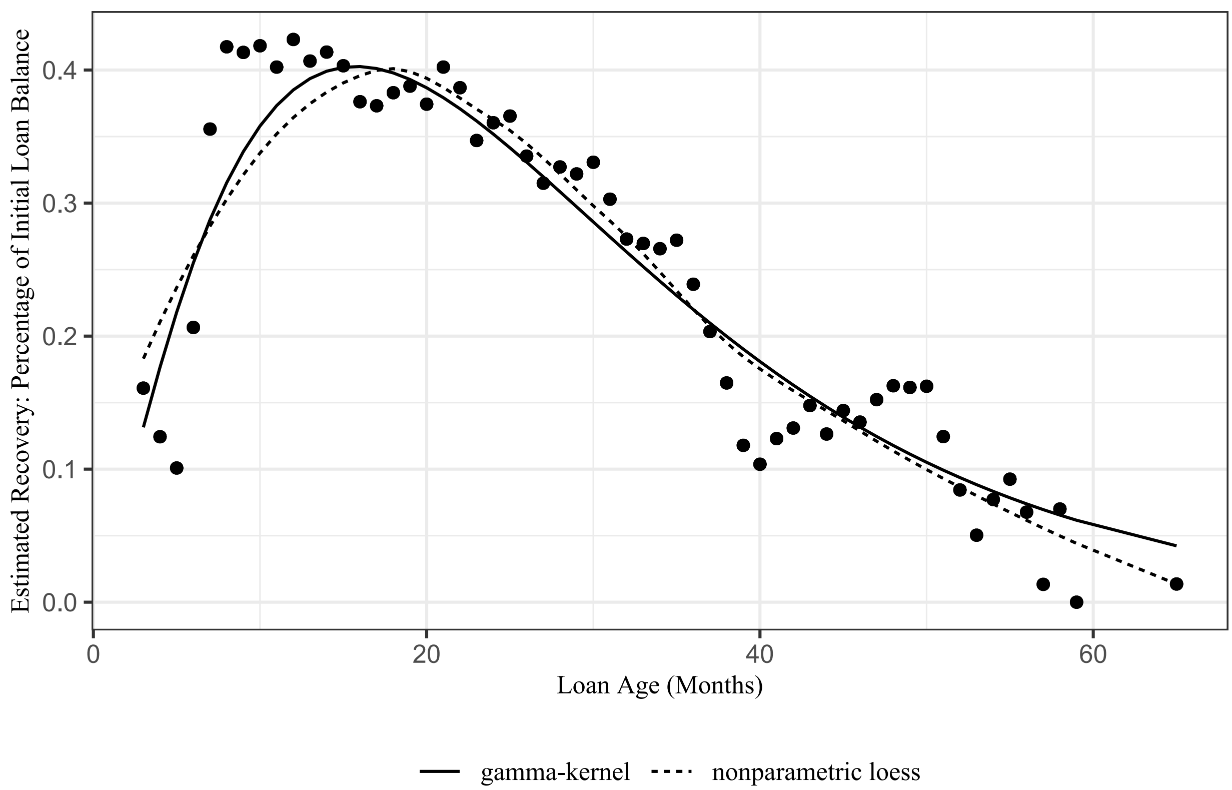


Figure C3: **Estimation of the Recovery Upon Default Assumption.**
The point estimates are formed using the asset-level data of Securities and Exchange Commission (2016) for the 58,118 filtered loans summarised in Section C.2. Specifically, they are the monthly average of the sum total of the `recoveredAmount` 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. Smoothing techniques are also presented. The shape of the recovery curve is similar for the sample of 65,802 loans issued in 2019.

space. See the solid line in Figure C3.

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 or removed collision insurance. Overall, it can be difficult to recover a meaningful amount in these circumstances. The recovery percentage then peaks at month 12 at just over 42% before declining towards zero as the loan age approaches termination (72–73 months). Since all vehicles in our sample are used, the decline in recoveries reflects the typical depreciating value of the automobile over time (e.g., Storchmann, 2004).

We close this section by noting the economic welfare of an automobile repossession has attracted the attention of researchers. Generally, the results are mixed. On the one hand, Pollard et al. (2021) discuss a vicious cycle of subprime auto lending where the same car may be bought, sold, and repossessed 20-30 times. This suggests repossessions may negatively impact economic welfare. A earlier result by Cohen (1998) finds that manufacturers prefer to offer prospective borrowers interest discounts over equivalent cash rebates because a legal technicality finds such a discount is financially beneficial to the lender in the event of repossession. In this case, the legal circumstances of a repossession may influence market behaviour. 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.

# D Robustness Analysis

We first examine the sensitivity of the credit risk convergence results to the economic impact of COVID-19. Next, we examine the sensitivity of the convergence results to collateral type (i.e., new autos versus used) and the business model of the lender.

## D.1 Impact of COVID-19

As alluded to in Section 3, we have attributed the large increase around loan age 40 for the default CSH rate estimate observable in Figure 1 to the Spring 2020 economic shutdown resulting from the initial rapid spread of the Coronavirus disease. Because the point of credit risk convergence occurs after month 40 for some pairs of risk bands in Table 1 (e.g., deep subprime and prime credit risk convergence occurs by loan age 50), there is a concern that the point estimate of default risk converging for disparate risk bands is due to the filtering effect of the shock of the economic shutdown rather than due to some inherent property of loan risk behaviour. In other words, only the strongest credits could survive such a shock, and credit risk convergence may occur later or not at all 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 1, we can see that pairs of risk bands converge earlier than loan age 40 (e.g., deep subprime and subprime, near-prime and prime, near-prime and super-prime, and prime and super-prime). Thus, we have examples of risk bands that converge in conditional monthly default risk prior to the onset of the Spring 2020 economic shutdown. Second, if credit risk convergence is completely driven by the Spring 2020 economic shutdown, we would expect to see it occur much earlier in a sample of bonds issued closer to Spring 2020 when subject to the same loan selection process and risk band definitions of Section C.1. Hence, we obtained loan level data from the same four consumer auto loan ABS issuers but from bonds issued closer to Spring 2020: SDART 2019-3 (Santander, 2019b), DRIVE 2019-4 (Santander, 2019a), CARMX 2019-4 (CarMax, 2019), and AART 2019-3 (Ally, 2019).[3] These bonds began paying in late Summer 2019, whereas the bonds introduced in Section C began paying in Spring 2017.

Figure D1 is a repeat of Figure 1; it presents the estimated CSH rates for default plus asymptotic 95% confidence intervals for the 2019 sample. As expected, we see the large spike in the CSH rate for defaults in subprime loans around 10 months, which, when adjusted for left-truncation, corresponds to the Spring 2020 economic shutdown. We also display the estimated credit risk convergence matrix in Table D1 for direct comparison to Table 1. In reviewing the matrix, we see evidence of earlier convergence. Hence, the shock of the economic shutdown of Spring 2020 has likely played some role. It is not the whole story, however. For example, the subprime risk band in the 2019 issuance does not converge with the super-prime risk until loan age 42. In the 2017 issuance, the subprime risk band converges with the super-prime risk band at loan age 48. This suggests that loan age or loan seasoning also plays a role. Similarly, while convergence between risk bands occurs earlier for the 2019 sample, it takes more months after the shutdown shock for most disparate risk bands to converge than after the same shock in the 2017 sample. For example, the subprime and prime risk bands converge by loan 25 in the 2019 sample, which is 15 months after the economic shutdown shock. For the 2017 sample, however, the subprime risk band converges with the prime risk band at loan age 42, which is only 2 months after the economic shutdown. This further suggests that the converge results of Table 1 are not solely attributable to the economic event of COVID. For completeness, we plot the full five-by-five matrix of CSH rate estimates for default in Figure D2 for the sample of 65,802 loans issued in 2019. It is a complete extension of the subprime versus prime plot in Figure D1. That is, Figure D1 is a zoomed-in view of the subprime-prime cell (row 4, column 2) in Figure D2.

[3] The filtered 2019 sample mirrors the distribution of the 2017 filtered sample summarised in Table C1. For example, there are 31,221 DRIVE 2019-4 loans, 19,962 SDART 2019-3 loans, 11,724 CARMX 2019-4 loans, and 2,895 AART 2019-3 loans, for a total of 65,802. By risk band, there are 24,107 (37%) deep subprime loans, 20,874 (32%) subprime loans, 9,930 (15%) near-prime loans, 8,625 (13%) prime loans, and 2,266 (3%) super-prime loans.

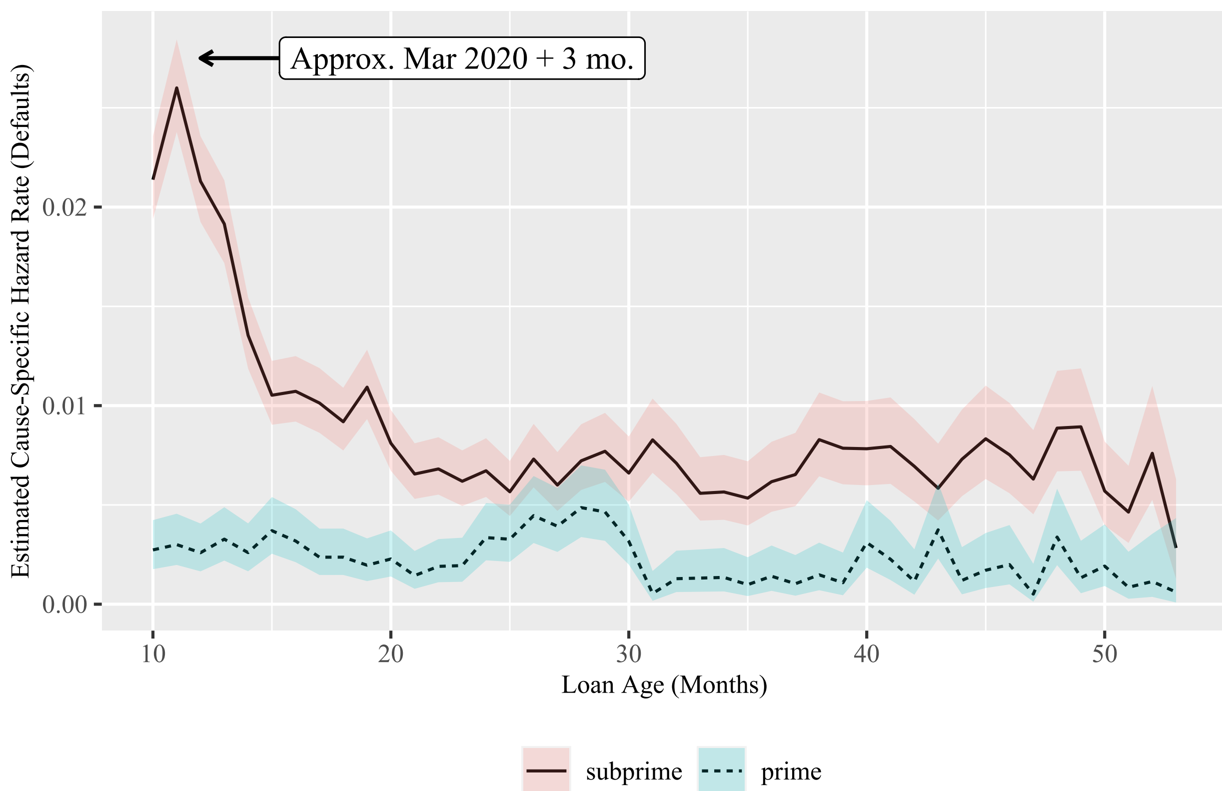


Figure D1: **Credit Risk Convergence: COVID Sensitivity.**
A plot of $\hat{\lambda}^{01}_{\tau,n}$ (defaults) defined in (5) by loan age for the subprime and prime risk bands within the sample of 65,802 loans issued in 2019, plus 95% confidence intervals using Lemma 1. We may use the hypothesis test described in (7) by searching for the minimum age that the confidence intervals overlap between two disparate risk bands. Because the 2019 bonds were issued closer to Spring 2020, the large upward spike in $\hat{\lambda}^{01}_{\tau,n}$ occurs much earlier for the subprime risk band, closer to loan age 10 (compare with Figure 1). We see some evidence of earlier credit risk convergence around loan age 25 in comparison to Figure 1.

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, credit risk convergence may be perpetually present, even if it may be partially explained by the filtering effects of an economic crisis.

## D.2 Additional Sensitivity Analysis

With some rudimentary data sorting, the techniques of Section 2 may be used for sensitivity testing. To illustrate, we now consider an additional robustness analysis. We instead sort the data for new cars at the point of sale. This will give us exposure to a potentially different borrower profile and depreciating collateral value pattern. It will also greatly reduce our exposure to the CARMX bond. Reduced exposure to CARMX is of interest because the

Table D1: **Credit Risk Convergence: 2019 Transition Matrix.**
This table reports a summary matrix of the estimated month of credit risk convergence for the sample of 65,802 72-73 month consumer automobile loans issued in 2019 (see Section D.1). For conservatism, the month of credit risk convergence is defined as the earlier of (1) the first of two consecutive months after ten months that the asymptotic confidence intervals for $\hat{\lambda}^{01}_{\tau,n}$ overlap or (2) once $\hat{\lambda}^{01}_{\tau,n}$ is consistently zero for both risk bands. Visually, it is helpful to compare Figure D1 with the subprime-prime cell below. Full comparisons may be made with Figure D2.

| | deep subprime | subprime | near-prime | prime | super-prime |
|---|---|---|---|---|---|
| deep subprime | 10 | 31 | 51 | 58 | 58 |
| subprime | | 10 | 23 | 25 | 42 |
| near-prime | | | 10 | 15 | 15 |
| prime | | | | 10 | 10 |
| super-prime | | | | | 10 |

parent company, CarMax, has an entirely different business model and therefore financing incentive than either Santander or Ally, the origination banks of the DRIVE, SDART, and AART ABS bonds. Because of this, it is possible that CARMX loans behave differently than loans originated by banks.

We again return to the original collective pool of over 275,000 consumer auto loans of the 2017 issuance of the four bonds introduced in Section C: CARMX, AART, DRIVE, and SDART. We then perform the identical risk band APR-based sorting and loan filtering of Section C.1, except rather than used cars we restrict our sample to new cars. This leaves a total sample of 16,412 loans, with bond exposures of DRIVE (7,692), SDART (7,369), ALLY (1,342) and CMAX (9). As expected, restricting the sample to new cars has eliminated almost all loans from CMAX, whose parent company, CarMax, specialises in used auto sales. Thus, the current sample of 16,412 loans consists of loans originated by traditional banks, Santander and Ally. In terms of risk band, we maintain dispersed exposure with deep subprime (3,892), subprime (8,242), near-prime (2,132), prime (1,407), and super-prime (739). Finally, all loans consist of a new vehicle at the point of sale, and so we are now considering an entirely different collateral depreciation pattern and even potentially borrower profile. We present an update of both Figure 1 and Figure D1 in Figure D3.

Immediately, we see that the overall pattern of Figure D3 closely mirrors that of Figure 1. The subprime loans have a default CSH rate estimate that is consistently higher than prime loans in the early months of a loan's age. We also see the large increase in the CSH rate for subprime loans around loan age 40, which correspond to the timing of the economic shutdown due to COVID-19 in Spring of 2020. As with the used cars-at-the-point-of-sale loans, there

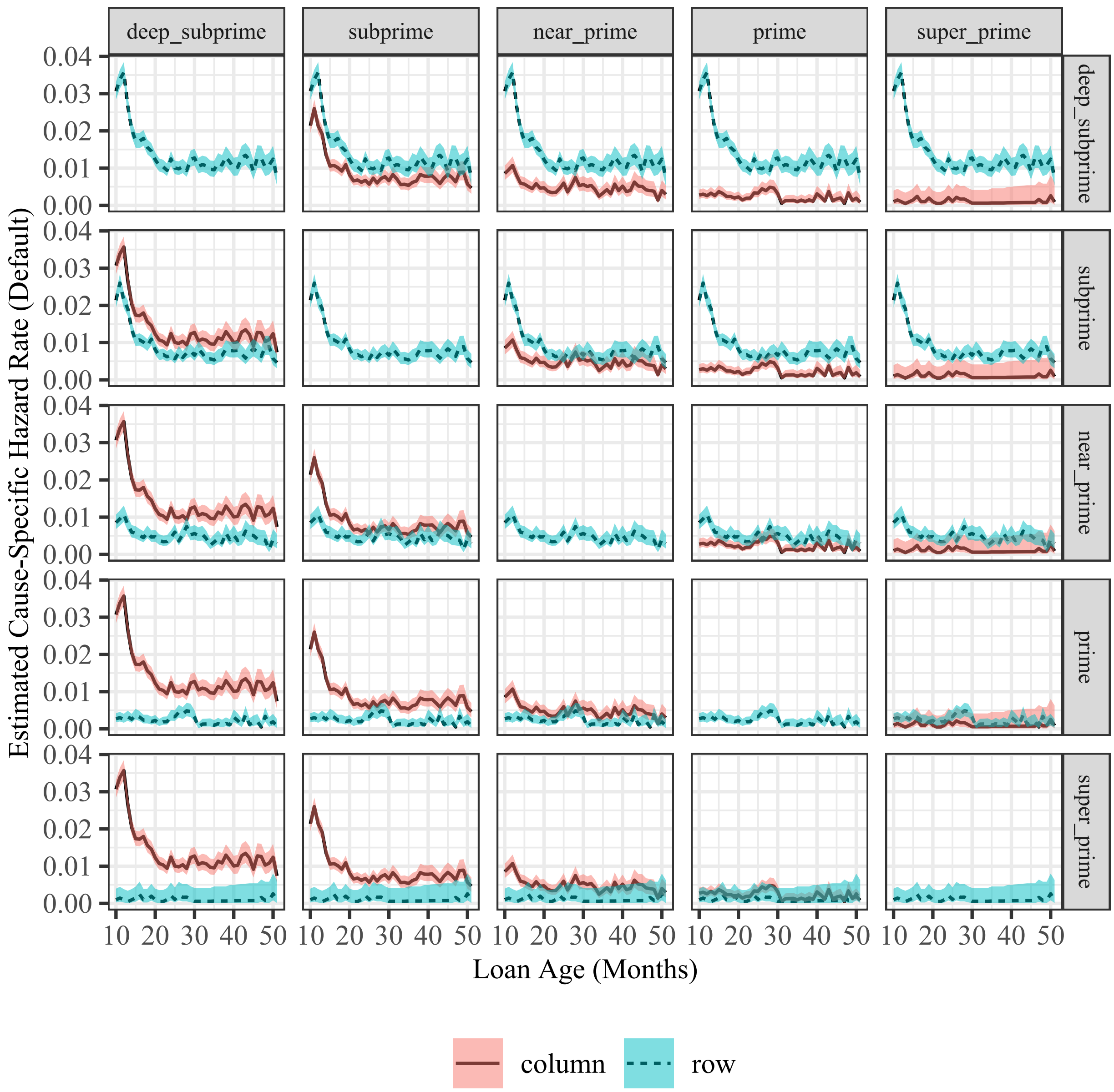


Figure D2: **Credit Risk Convergence: All Risk Bands (2019).**
A plot of $\hat{\lambda}^{01}_{\tau,n}$ (defaults) defined in (5) by loan age for all five risk bands within the sample of 65,802 loans (Section D.1), plus 95% confidence intervals using Lemma 1. It is a repeat of Figure 6 for the 2019 issuance as a sensitivity check that the economic shock of COVID-19 is not the sole reason for the estimated timing of credit risk convergence between disparate risk bands.

appears to be minimal impact from COVID-19 for prime loans. The two CSH rates for the subprime and prime risk bands eventually converge, however, which we see at the lower right corner of Figure D3. The asymptotic confidence intervals begin to consistently overlap beginning shortly after loan age 40, which corresponds to row two, column four of the top matrix of Table 1. Thus, our credit risk convergence point estimates appear to be robust in

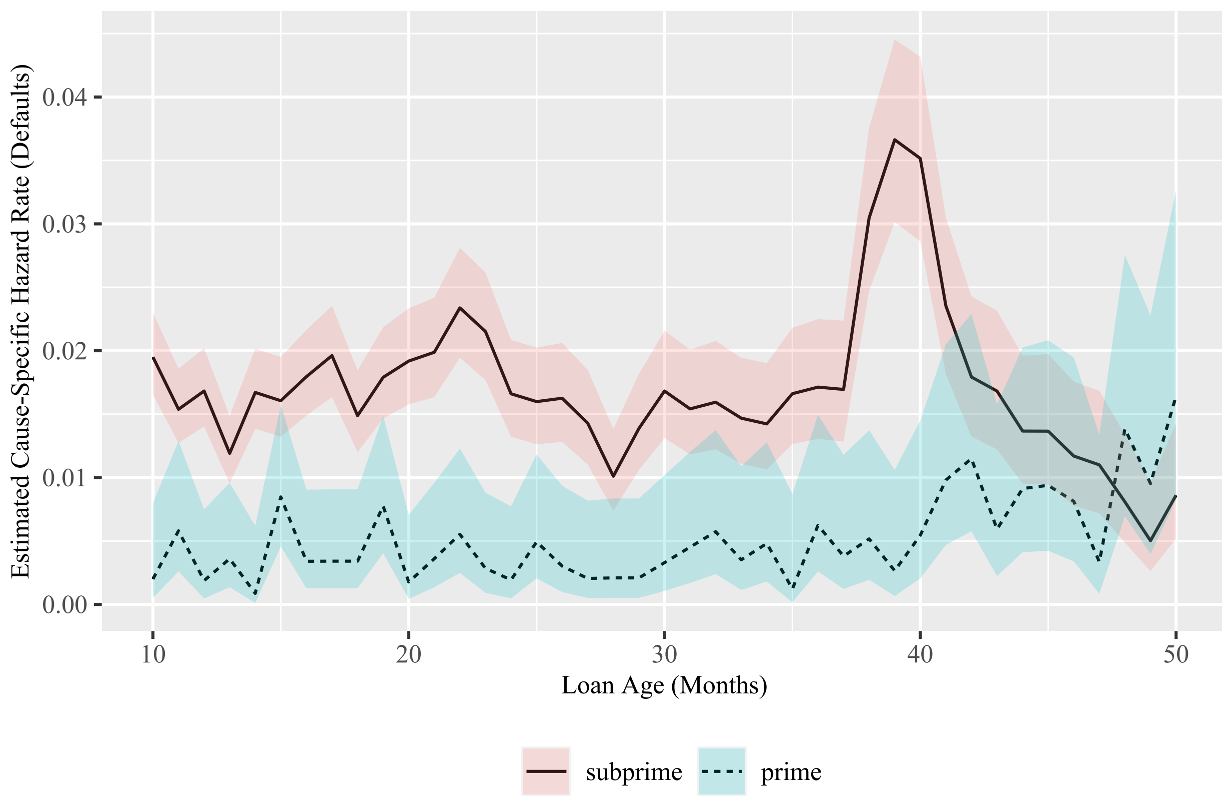


Figure D3: **Credit Risk Convergence: Collateral Sensitivity.**
A plot of $\hat{\lambda}^{01}_{\tau,n}$ (defaults) defined in (5) by loan age for the subprime and prime risk bands within the sample of 16,412 loans issued in 2017 with new cars at the point of sale, plus 95% confidence intervals using Lemma 1. We may use the hypothesis test described in (7) by searching for the minimum age that the confidence intervals overlap between two disparate risk bands. Because of the smaller sample, the asymptotic confidence interval for the CSH rate of prime loans is wider. The overall pattern is very similar to Figure 1, however, and so the point estimates of credit risk convergence appear to be robust to collateral type at the point of sale (i.e., new or used). The sample of 16,412 new car loans also has minimal exposure to CARMX. Thus, the CSH estimates further appear robust to different business incentives of the loan originator.

consumer auto loans to the collateral type at the point of sale (i.e., new or used). Because the sample of 16,412 new car loans has such minimal exposure to CARMX, we also see that our credit risk convergence point estimates appear to be robust to potentially different business incentives of the parent company to the loan originator (i.e., used car sales versus traditional banking).

As a final note on collateral type, a close inspection of Figure D3 in comparison to Figure 1 reveals wider asymptotic confidence intervals for the CSH rate for default in prime loans. This is driven by the smaller sample size, and it is exacerbated for super-prime loans written on new cars (i.e., there are very few defaults for super-prime loans written on new cars in our sample of 739). Hence, we have avoided reporting the credit risk convergence point estimate matrix of Table 1 for the sample of 16,412 new car loans to avoid potentially

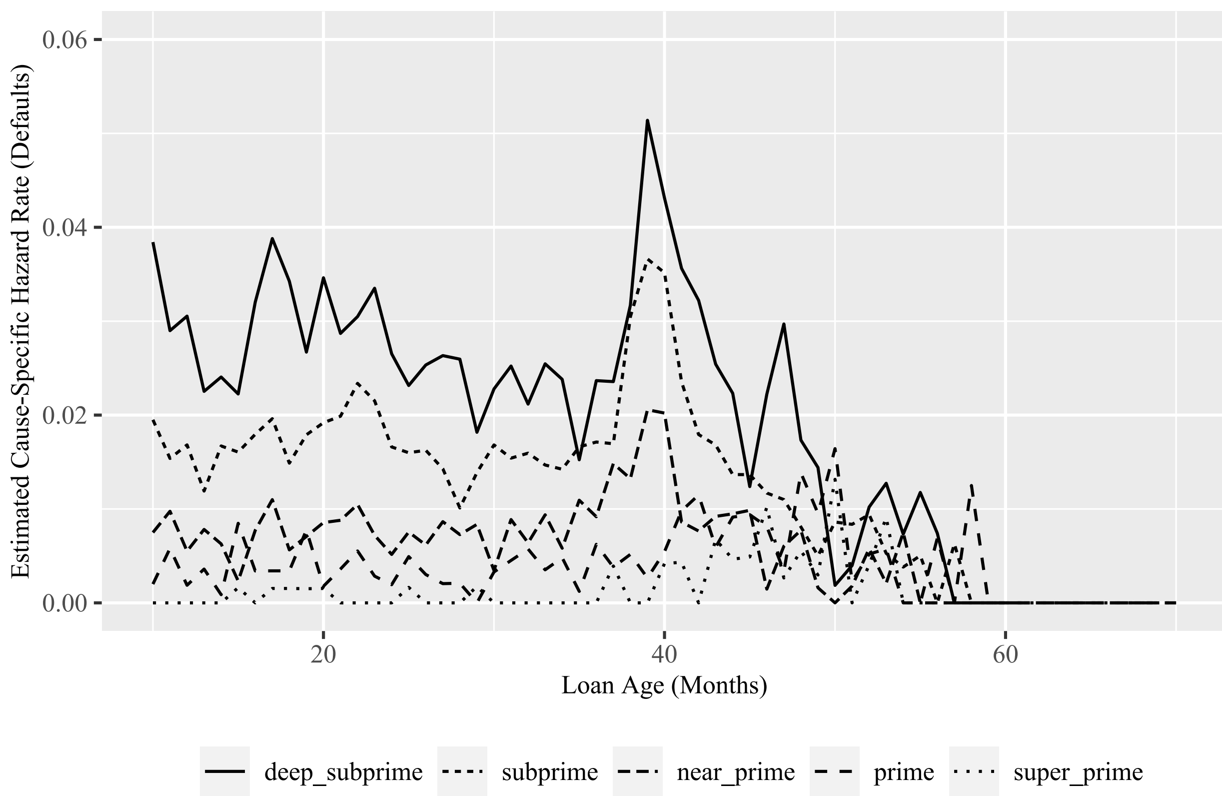


Figure D4: **Credit Risk Convergence: All Risk Bands, Point Estimates.** A plot of $\hat{\lambda}^{01}_{\tau,n}$ (defaults) defined in (5) by loan age for all risk bands within the sample of 16,412 loans issued in 2017 with new cars at the point of sale. As expected, the CSH rates are the highest for deep subprime loans and then trend downwards until super-prime loans at the onset of loan lifetimes. As the loans mature and stay current, however, we see that all CSH rates eventually converge towards zero at the bottom right. This is an alternative visualization of loan seasoning, to be compared with Figure A1.

erroneously conclusions due to faulty asymptotic statistics stemming from a small default sample. Instead, we report the point CSH rate estimates for default for all five risk bands in Figure D4. In this case, a simple line plot speaks volumes. In the young ages of a loan, we see that the CSH rates for default is the highest for deep subprime loans, and it progresses sequentially downward by risk band until super-prime loans, of which there are very few defaults. This pattern is expected. As the loans age, however, we see all CSH rates for default for each risk band converge together in the bottom right of Figure D4 near loan age 50. Given consumer auto loans on new car sales are also collateralised with rapidly depreciating assets, these results similarly cannot be explained by traditional LTV optionality arguments found in mortgages (e.g., Campbell and Cocco, 2015).

# E Large Sample Simulation Study

We present a simulation study in support of Theorem 2.1 and Lemma 1. Let the true distribution for the lifetime random variable $X$ and bivariate distribution of $(X, Z_X)$ be as in Table E1. 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, $X$.
4. Simulate a Bernoulli event with probability $p(x)$ to determine if the event $X$ was caused by type 1 with probability $p(x)$ or type 2 with probability $1 - p(x)$.

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

To validate the asymptotic results of Theorem 2.1, we compare the empirical covariance matrix against the derived asymptotic covariance matrix, $\mathbf{\Sigma}^{0i}$, by examining estimates of the confidence intervals using Lemma 1. Figure E1 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 Theorem 2.1 and estimated quantities using Theorem 2.1 but replacing all quantities with their respective estimates from Section 2. This agreement further confirms Theorem 2.1.

Table E1: **Simulation Study Lifetime of Interest Probabilities.**
The true probabilities of the lifetime random variable, $X$, for the simulation study results of Figure E1. The probabilities $p(x)$ and $\Pr(X = x)$ for $x \in \{1, \ldots, 10\}$ are selected at onset, and the remaining probabilities in this table may be derived from these quantities. Not summarised here is the truncation random variable, $Y$, which was assumed to be discrete uniform over the integers $\{1, \ldots, 5\}$.

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

# F Lifetime Risk-Adjusted Return

We present an expansion of the actuarial methods in Section 4.1 to consider the full remaining lifetime of a loan rather than assuming a prepayment in the next month. Denote the risk-adjusted rate of return for a loan in risk band $a$ as $\rho_a$. Given reliable estimates of borrower default and prepayment probabilities, such as those in Section 2, we may estimate $\rho_a$ for a given loan in risk band $a$. In particular, we may estimate $\rho_a$ for each month a loan is still active and paying to find a *conditional risk-adjusted rate of return* over a loan's full remaining lifetime. Contrast this with Section 4.1, in which we calculate a one-month risk-adjusted return. Pleasingly, $\rho_a$ equals the loan contract effective rate of return in the event the future loan payments will proceed as scheduled with no uncertainty, which we state formally in Theorem F.1.

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

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

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

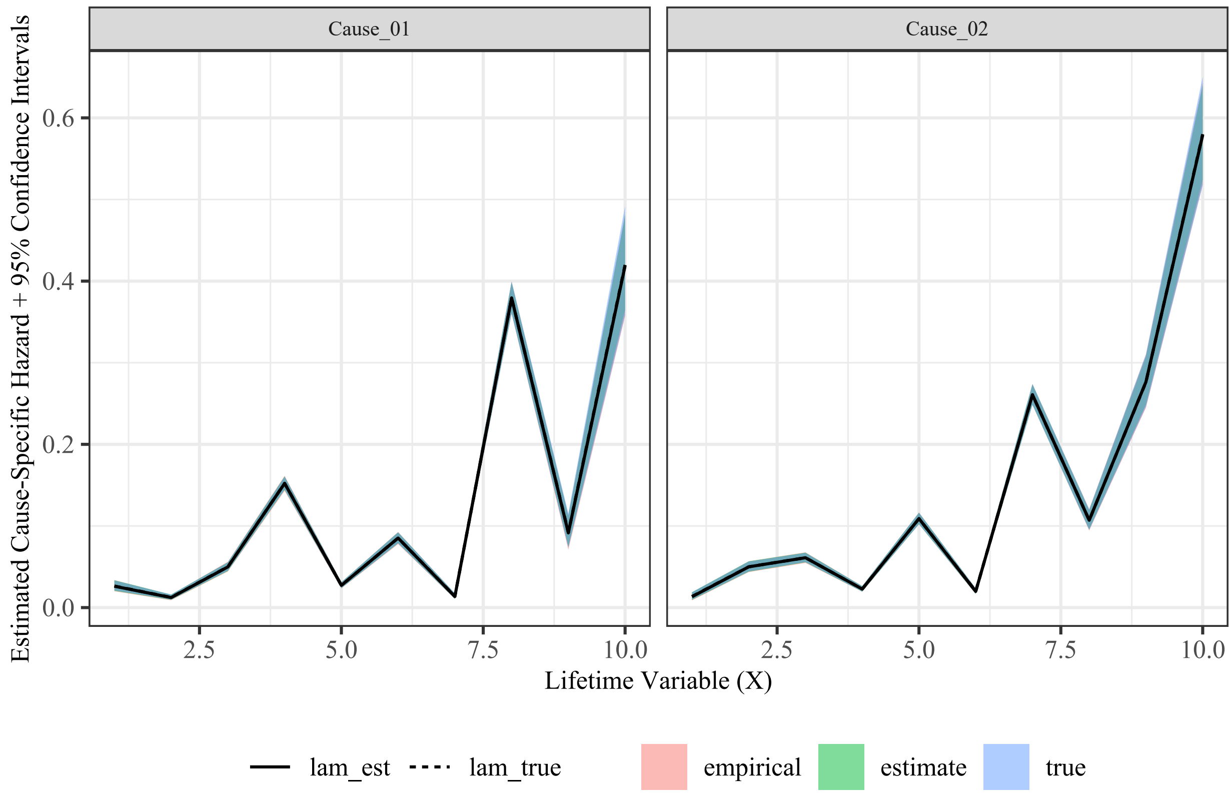


Figure E1: **Simulation Study Results.**
A comparison of true $\lambda_{\tau}^{0i}(x)$ (`lam_true`) and estimated $\hat{\lambda}_{\tau,n}^{0i}(x)$ (`lam_est`), including confidence intervals, for the distribution in Table E1 and $i = 1, 2$. The "true" values are from Theorem 2.1 and Lemma 1. The "estimate" values use the formulas from Theorem 2.1 and Lemma 1 but replace the true values with the estimates from Section 2 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.

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

$$\begin{aligned} B_x &= B(1+r_a)^x - P\left[\frac{(1+r_a)^x - 1}{r_a}\right] \\ &= B(1+r_a)^x - B\left[\frac{1-(1+r_a)^{-\xi}}{r_a}\right]^{-1}\left[\frac{(1+r_a)^x - 1}{r_a}\right]. \end{aligned} \tag{2}$$

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

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

Use (2) and solve for $\rho_{a|x}$ to complete the proof. $\square$

We now formalise the estimation of $\rho_{a|x}$, as defined in Theorem F.1. 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. Assume we consider a loan with a $\psi$-month schedule. Denote the current age of a loan by $x$, $1 \leq x \leq \psi$.[4] 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 $i = 1, 2$, $x \leq j \leq \psi$,

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

again with the convention $\prod_{k=x}^{x-1}\{1 - \lambda(k)\} = 1$. For convenience, denote $p_x^{0i}(j) = \Pr(X = j, Z_j = i \mid X \geq x)$ for $i = 1, 2$, $x \leq j \leq \psi$. Hence,

$$p_x^{0i}(j) = \begin{cases} \lambda^{0i}(x), & j = x \\ \lambda^{0i}(j) \prod_{k=x}^{j-1}\{1 - \lambda(k)\}, & j > x, \end{cases} \qquad i = 1, 2.$$

One may verify $\sum_{j=x}^{\psi} \sum_{i=1}^{2} p_x^{0i}(j) = 1$ for every $x$.[5]

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

[4] Depending on the impact of left-truncation and right-censoring, the recoverable range of $X$ may not be the entire original loan termination schedule (see Section 2 for details). In such an instance, assumptions about the probability distribution may be necessary. Assuming a geometric right-tail (i.e., a constant hazard rate that follows the last recoverable value) is common in survival analysis (Klugman et al., 2012, Section 12.1). We will proceed as though the full distribution is recoverable and allow readers to adjust as needed.

[5] It may be of help to review the numeric example of Table E1 in Online Appendix E.

future default paths is

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

Note that row 1 of **DEF** would be the cash flows assuming a default at loan age $x$, which occurs with probability $p_x^{01}(x)$. Similarly, row 2 of **DEF** would be the cash flows assuming a default at loan age $x+1$, which occurs with estimated probability $p_x^{01}(x+1)$, and so on and so forth. In the same way, we can define the prepayment matrix at loan age $x \le \psi - 1$ as

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

As with defaults, row 1 of **PRE** would be the cash flows assuming a prepayment at loan age $x$, which occurs with estimated probability $p_x^{02}(x)$. Similarly, row 2 of **PRE** would be the cash flows assuming a prepayment at loan age $x+1$, which occurs with estimated probability $p_x^{02}(x+1)$, and so on and so forth. Therefore, if we denote the $(\psi - x + 1) \times 1$ dimensional discount vector assuming the unknown monthly rate of $\rho_x$ as

$$\left(\boldsymbol{\nu}_x\right)^\top = \Big((1+\rho_x)^{-1} \quad (1+\rho_x)^{-2} \quad \dots \quad (1+\rho_x)^{-(\psi-x+1)}\Big)^\top,$$

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

$$\left(\boldsymbol{p}_x^{0i}\right)^\top = \Big(p_x^{0i}(x) \quad p_x^{0i}(x+1) \quad \dots \quad p_x^{0i}(\psi)\Big)^\top,$$

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

$$\mathrm{EPV}_x = \left(\boldsymbol{p}_x^{01}\right)^\top \mathbf{DEF}_x \boldsymbol{\nu}_x + \left(\boldsymbol{p}_x^{02}\right)^\top \mathbf{PRE}_x \boldsymbol{\nu}_x.$$

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

$$\{\rho_x : B_x = \text{EPV}_x\}. \tag{3}$$

In words, $\rho_x$ represents the expected return realised by lending $B_x$ and taking into account the original monthly payments $P$ and default and prepayment risk over the remaining lifetime of the loan. We have $\rho_x \leq r$ for a given contract, with equality only in the circumstances of Theorem F.1. Finally, we of course do note know the true distribution of $X$. We do have the estimators in (5), however, and Theorem 2.1. Thus, we may estimate $\rho_x$ by replacing the cause-specific hazard rates $\lambda^{0i}$ with the estimate in (5). For completeness, we close this section with the following lemma.

**Lemma 1** ($\hat{\rho}_{n,x}$ *Asymptotic Properties*)**.** *Replace the cause-specific hazard rates in* (3) *with the estimators from* (5)*. Define the estimated risk-adjusted rate of return over the remaining lifetime given a loan has survived to month $x$ as $\hat{\rho}_{n,x}$. Then,*

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

*Proof.* The result follows by Theorem 2.1, part $(i)$ and the Continuous Mapping Theorem (Mukhopadhyay, 2000, Theorem 5.2.5, pg. 249). □

# References

M. Adelino, K. Gerardi and B. Hartman-Glaser (2019). "Are lemons sold first? Dynamic signaling in the mortgage market." *Journal of Financial Economics* **132**, 1–25.

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.

J. J. Assunção, E. Benmelech and F. S. S. Silva (2013). "Repossession and the democratization of credit." *The Review of Financial Studies* **27**, 2661–2689.

J. Beyersmann, A. Latouche, A. Buchholz and M. Schumacher (2009). "Simulating competing risks data in survival analysis." *Statistics in Medicine* **28**, 956–971.

J. Y. Campbell and J. F. Cocco (2015). "A model of mortgage default." *The Journal of Finance* **70**, 1495–1554.

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.

L. Cohen (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).

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

L. Einav, M. Jenkins and J. Levin (2012). "Contract pricing in consumer credit markets." *Econometrica* **80**, 1387–1432.

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

J. P. Lautier, V. Pozdnyakov and J. Yan (2023). "Pricing time-to-event contingent cash flows: A discrete-time survival analysis approach." *Insurance: Mathematics and Economics* **110**, 53–71.

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

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

R. Phillips (2013). "Optimizing prices for consumer credit." *Journal of Revenue & Pricing Management* **12**.

J. Pollard, E. Blumenberg and S. Brumbaugh (2021). "Driven to debt: Social reproduction and (auto)mobility in Los Angeles." *Annals of the American Association of Geographers* **111**, 1445–1461.

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

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.

Securities and Exchange Commission (2016). "17 CFR S 229.1125 (Item 1125) Schedule AL - Asset-Level Information."

K. Storchmann (2004). "On the depreciation of automobiles: An international comparison." *Transportation* **31**, 371–408.