Submitted to the Annals of Applied Statistics

ESTIMATING THE TIME-TO-EVENT DISTRIBUTION FOR LOAN-LEVEL DATA WITHIN A CONSUMER AUTO LOAN ASSET-BACKED SECURITY

BY JACKSON P. LAUTIER [1,a], VLADIMIR POZDNYAKOV [2,b] AND JUN YAN [2,c]

[1]*Department of Mathematical Sciences, Bentley University,* [a]*jlautier@bentley.edu*

[2]*Department of Statistics, University of Connecticut,* [b]*vladimir.pozdnyakov@uconn.edu;* [c]*jun.yan@uconn.edu*

The random cash flows of consumer auto asset-backed securities (ABS) depend critically on the time-to-event distribution of its individual, securitized assets. Estimating this distribution has historically been challenged by limited data. Recent regulatory changes reversed this, however, and asset-level auto ABS data is now publicly available to investors for the first time. The idiosyncrasies of this ABS data present new difficulties in estimating the loan-level lifetime distribution due to its discrete-time structure and exposure to left-truncation. We propose a parametric framework for estimating the loan-level lifetime distribution while leaving the left-truncation time distribution unspecified. Through theorems developed to identify the stationary points of the likelihood, we significantly simplify a complex multiparameter constrained optimization problem. These stationary points, shown to be the roots of an estimating equation, enable asymptotic normality and large-sample inference under suitable regularity conditions. For an actuarial policy limit geometric distribution, closed-form maximum likelihood estimates may be derived. These theoretical results are further generalized to accommodate right-censoring and validated through numerical and simulation studies. These methods are then applied to auto ABS data, including a likelihood ratio test to assess model specification, comparing various parametric distributions, and contextualizing these results from the investor perspective.

1. Introduction. If we may treat financial engineering as a subclass of engineering more broadly, then *securitization* is a marvel that would rival any bridge, railway, or expressway. Consider the financial dilemma of an automobile manufacturer: a desire to sell cars to consumers that cannot afford to buy them. The solution is financing through auto loans, and lenders (e.g., Ally Bank) step in to bridge this gap. This eventually creates a problem for lenders: a desire to write loans but a lack of available funds. The solution is again financing whereby the lender will trade a portfolio of long-dated (e.g., 72-month) auto loans for cash. The counter-party in this trade is investors, who have cash to invest and seek returns. Insert the *asset-backed security* (ABS) to make this connection, just as a bridge, railway, or expressway connects allied geographic regions with a shared economic interest.

The formation of an ABS begins with a lender moving individual consumer auto loans into a special purpose vehicle that creates a legal barrier and services these loans. Next, the ABS is formed by establishing a trust (e.g., AART, 2017) with a collection of a specific set of auto loans. This trust also removes any financial dependence of prospective investors on the lender's financial health, which typically allows a lender to obtain a lower borrowing cost than a traditional corporate bond. The ABS also creates various payment priorities (i.e., *tranches*) to offer prospective investors various risk-return options. Critically, the ABS now generates a large, long duration monthly cash flow that may be traded freely on the open market as an investment security (hence the name, *securitization*). For a summary, see Figure 1.

With recent annual issuance of U.S. auto ABS totaling over \$170B (SIFMA, 2024), we see an economic need for an applied study from the perspective of an ABS investor. As

Keywords and phrases: Asset-level disclosures, Credit risk, Incomplete data, Reg AB II, Survival analysis.

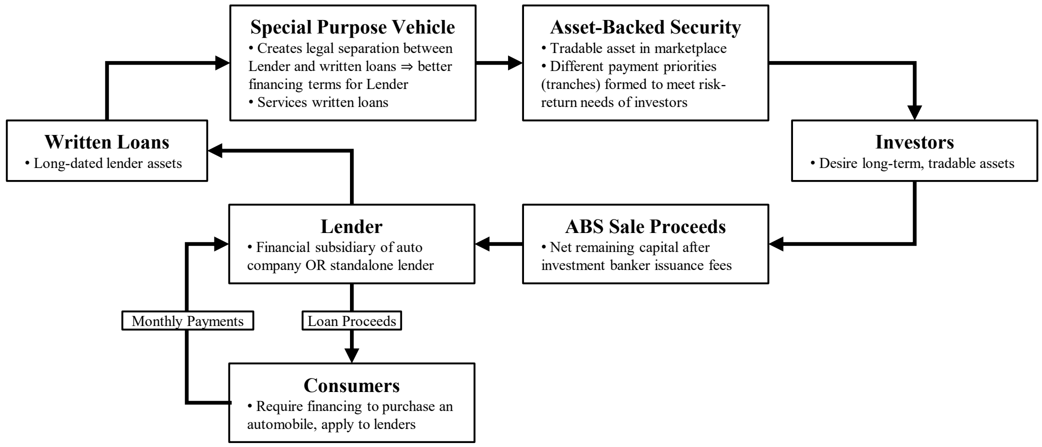

FIG 1. ***Securitization Lifecycle***. *An overview of the economic motivation of the securitization process.*

in most financial modeling exercises, the ultimate end goal is a stochastic ABS cash flow model. Before this can be achieved, however, it is essential to refine methods to estimate the time-to-event distribution of its supporting individual consumer auto loans. To see this, it is informative to visualize the investor level cash flows. Each auto ABS will contain n consumer auto loans of various ages, x_i, $1 \leq i \leq n$. Each loan at age x_i, $L_i(x_i)$, $1 \leq i \leq n$, will then produce a monthly cash flow, either a loan payment, a prepayment, a zero in the event of a missed payment or scheduled termination, or repossession proceeds in the event of default. For any month j the securitization is active, $1 \leq j \leq s$, a column-wise sum of these *loan-level* cash flows, $\text{CF}_{i(x_i+j)}$, $1 \leq i \leq n$, supports the cash flow that is ultimately returned to investors. This total investor cash flow, $\sum_i \text{CF}_{i(x_i+1)}$, then moves through a prescribed payment priority structure known colloquially as a *waterfall*, denoted by f_w, to create its aforementioned tranches (e.g., AART, 2017). This monthly process repeats until there are so few paying loans it is no longer economical to maintain the ABS, which we assume is s total months (see Table 1). Thus, a stochastic ABS cash flow model can be achieved by using a random time-to-termination distribution of each individual auto loan (e.g., Lautier, Pozdnyakov and Yan, 2023a; Agarwal, Marion and Wu, 2024), and it is the resulting statistical problem, estimating this time-to-termination distribution, that is our focus.

TABLE 1
Visualizing Auto ABS Investor Cash Flows. *Monthly ABS investor cash flows (CF) are a function, f_w, of the column-wise sum of the monthly CF of all its individual loans. Hence, ABS CFs can be scholastically modeled using an individual loan-level time-to-termination random variable.*

Loan (L) (Age)	Month 1	Month 2	…	Month s
$\text{L}_1(x_1)$	$\text{CF}_{1(x_1+1)}$	$\text{CF}_{1(x_1+2)}$	…	$\text{CF}_{1(x_1+s)}$
$\text{L}_2(x_2)$	$\text{CF}_{2(x_2+1)}$	$\text{CF}_{2(x_2+2)}$	…	$\text{CF}_{2(x_2+s)}$
⋮	⋮	⋮	⋮	⋮
$\text{L}_n(x_n)$	$\text{CF}_{n(x_n+1)}$	$\text{CF}_{n(x_n+2)}$	…	$\text{CF}_{n(x_n+s)}$
ABS CF	$f_w\left(\sum_{i=1}^{n} \text{CF}_{i(x_i+1)}\right)$	$f_w\left(\sum_{i=1}^{n} \text{CF}_{i(x_i+2)}\right)$	…	$f_w\left(\sum_{i=1}^{n} \text{CF}_{i(x_i+s)}\right)$

Benefiting this time-to-event estimation is a recent explosion in publicly available, loan-level ABS data (see Section 2). Using loan-level data from ABS is not without its complexity, however, because of incomplete data, particularly in the form of left-truncation (Lautier, Pozdnyakov and Yan, 2023b; Katcher et al., 2024). Beyond left-truncation, the financial nature of such data also present challenges related to its discrete-time structure. While left-truncation in continuous time has been extensively studied (e.g., Woodroofe, 1985; Tsai, Jewell and Wang, 1987; Wang, 1989; Huang and Wang, 1995), the combination of left-truncation, discrete-time, and a known, finite support for the lifetime random variable remains relatively unexplored (Lautier, Pozdnyakov and Yan, 2023b). Because this combination is necessary to accurately estimate the loan-level time-to-termination random variable within the auto ABS setting that is our focus, it is important to contribute to this nascent line of work.

Recent approaches leverage the finite, known support of the loan lifetime random variable to estimate each recoverable probability point mass in the distribution as a parameter (Lautier, Pozdnyakov and Yan, 2023b). This allows direct proofs that the classical nonparametric estimators of Woodroofe (1985) and Huang and Wang (1995) are simultaneously parametric maximum likelihood estimates (MLEs) in this setting (Lautier, Pozdnyakov and Yan, 2023b, 2024a). It also yields completely specified asymptotic multivariate normal distributions with a diagonal covariance structure for the vector of hazard rate estimators (Lautier, Pozdnyakov and Yan, 2023a,b, 2024b). Additionally, Lautier, Pozdnyakov and Yan (2023a,b, 2024a,b) impose no assumptions about the shape of the left-truncation random variable, making it less restrictive than methods that require a uniformity assumption (e.g, *length-biased sampling* (Asgharian, M'Lan and Wolfson, 2002; Uña-Álvarez, 2004)). These results have lead to successful applications in financial econometrics (Lautier, Pozdnyakov and Yan, 2024b).

Despite these aforementioned advances, there has not yet be a thorough study of the case the lifetime random variable takes a more traditional parametric form with fewer parameters in the left-truncation, discrete-time, and a known, finite support setting of auto ABS data. Establishing an estimation paradigm in this alternative, fewer parameter case would offer distinctive advantages. First, traditional parametric distributions provide a natural smoothing, whereas the approaches of Lautier, Pozdnyakov and Yan (2023a,b, 2024a,b) can yield zero estimates at months without observations. Second, parametric forms provide an avenue for the eventual incorporation of economic variables, desirable for econometric studies of loan survival times. Finally, the simplicity and stability of parametric forms are attractive properties for ABS investors operating in investment contexts that often require rapid decision-making.

We thus present a novel parametric framework for analyzing left-truncated, discrete time-to-event data found in auto ABS. Despite its shared application, this manuscript should not be viewed as an extension of Lautier, Pozdnyakov and Yan (2023a,b, 2024a,b). Rather, it is an entirely new branch of reasoning required by altering the form of the lifetime distribution's admissible families. We find our proposed estimators simplify a complex multidimensional constrained optimization parameter estimation problem into a single or low dimension optimization. This allows for solutions where direct numerical optimization methods may otherwise fail. These estimators can be shown to be asymptotically consistent M-estimators (e.g., van der Vaart, 1998), from which asymptotic inference follows. As a special case, we derive closed-form MLEs for an actuarial *policy limit* geometric distribution (Klugman, Panjer and Willmot, 2012, §8.4, pg. 125). All results also accommodate right-censoring. All methods are vetted through simulation study, and we illustrate these methods with a detailed application to the ABS bonds AART (2017) and AART (2019). In the application, we consider the problem of model specification, of which we propose a likelihood ratio test (LRT) against Lautier, Pozdnyakov and Yan (2023a) and illustrate a vector parameter case with the discrete Weibull distribution of Nakagawa and Osaki (1975). While our focus is ABS, the framework is also applicable where incomplete, discrete time-to-event data are prevalent, such as healthcare, finance, engineering, telecommunication, and insurance.

The paper proceeds as follows. Section 2 introduces the ABS data. Section 3 presents our methods in three parts: establishing preliminaries, providing statements under left-truncation, and then generalizing all statements for right-censoring. Simulation studies follow in Section 4, and Section 5 is a lengthy applied study with discussion of its practical value for ABS investors. Section 6 then concludes. The Supplemental Material provides complete proofs of all major results, and further details to support asymptotic behavior, likelihood construction, implementation, simulation, and the application. For reference, all data and replication code are publicly available at https://github.com/jackson-lautier/consumer-auto-abs-parametric.

2. Auto Asset-Backed Security Data. Estimating the time-to-event distribution for individual consumer auto loans will require access to loan-level ABS data. Historically, such data access has been limited for auto ABS investors (e.g., AART (2010) provides only pool-level summary statistics). This lack of transparency in auto ABS was deemed unacceptable after investors failed to anticipate higher than expected defaults for residential subprime mortgages packaged into ABS, triggering the 2007-2009 financial crisis (Mishkin, 2011). In response, regulators implemented several measures to improve transparency within all ABS. These include the Securities and Exchange Commission's (SEC's) significant revisions to Regulation AB and new rules governing ABS disclosures (Securities and Exchange Commission, 2014). Notably, ABS issuers must now provide detailed loan-level data, including payment performance, on a monthly basis via the Electronic Data Gathering, Analysis, and Retrieval (EDGAR) system (Securities and Exchange Commission, 2016).

Given this new, unprecedented data access, we scraped the SEC's EDGAR system to compile asset-level demographic and payment performance data from the auto ABS bonds Ally Auto Receivables Trust 2017-3 (AART, 2017) and Ally Auto Receivables Trust 2019-3 (AART, 2019) (see Table 2). The bonds AART (2017) and AART (2019) are just two examples of many publicly-issued ABS bonds that may now be analyzed, and similar data have begun to appear in recent studies (e.g., Katcher et al., 2024; Lautier, Pozdnyakov and Yan, 2024b). As such, techniques to model this financial lifetime data have grown in importance.

TABLE 2
ABS Data Summary. *Summary statistics, range, mean (median), of ABS data studied in Section 5.*

Bond	Count	Orig. Date	Credit Score	Orig. Loan Bal.	Loan Term
AART (2017)	67,797	Feb.11'-Apr.17'	725 (719)	\$15,605 (\$14,153)	12-78 Mo.
AART (2019)	67,198	Oct.12'-Jul.19'	722 (718)	\$16,682 (\$14,164)	13-79 Mo.

In Section 5, we define the lifetime random variable as the time-until-loan-payments stop, either due to prepayment or default. We acknowledge differentiating between the two (i.e., *competing risks*) may be relevant for a credit risk analysis of the underlying borrowers (e.g., Lautier, Pozdnyakov and Yan, 2024b). Our focus is estimating the loan lifetime distribution for the eventual purpose of ABS cash flow modeling, however, and so this distinction is less meaningful for two reasons. First, average borrower credit scores in Table 2 are *super-prime* (Lautier, Pozdnyakov and Yan, 2024b), of which only 6% of loans eventually default (Lautier, Pozdnyakov and Yan, 2024b). Second, in the event of default, the automobile will be repossessed, and the repossession proceeds will be paid into the ABS. Taken together, the cash flow difference will likely be negligible when spread over 67,000+ loans.

3. Methods. We begin by detailing how incomplete lifetime data forms within an ABS and establishing our notation. Next, we state major results in the case of left-truncation. The section concludes by generalizing all results to also accommodate right-censoring.

3.1. *Preliminaries.* We now introduce notation and assumptions within the context of estimating the parameters of the loan-level time-to-termination random variable, denoted X, via auto ABS data. Because the lifetime of interest is a loan with monthly payments, X is a discrete random variable over $\mathbb{N}$. Further, the amortization schedule is known at contract signing, so X has a known, finite upper bound, denoted by the nonrandom $\omega \in \mathbb{N}$ (e.g., for a 72-month loan, $\omega = 72$). Left-truncation manifests in the data generating process of auto ABS loan lifetimes through the lengthy legal machinations of its formation (see Figure 1). This length of time, which can take several months to years, is known colloquially as a *warehousing period*. It is thus possible loans will terminate during the warehousing period and never be observed by investors. Hence, loan-level time-to-event data present in an auto ABS is conditional on survival beyond a second time-to-event random variable, denoted Y. The random variable Y represents the random number of months from loan origination to the first month an ABS trust begins paying investors. In other words, we observe X if and only if $X \geq Y$, and so Y is a left-truncation random variable acting upon X.

Given this, an auto ABS bond consists of a fixed size n random sample of consumer auto loans taken from a left-truncated population. For example, AART (2017) consists of 67,797 loans sampled from an outstanding retail contract population of over 2.2 million. We thus assume the population is already subject to left-truncation to more closely reflect the data generation process of loan-level data within an auto ABS pool. This differs from classical treatments (e.g., Woodroofe, 1985), which first assume a bivariate sample of size N, after which a left-truncated random sample of random size $n \leq N$ is drawn. Per Figure 1, this would erroneously imply ABS investors have access to all lender balance sheet loans.

We now formalize the support of X and Y. As with X, Y is a discrete random variable with a known, finite support. Loans that will eventually be included in the auto ABS will be originated for an observable total of nonrandom $m \in \mathbb{N}$ months (Lautier, Pozdnyakov and Yan, 2023a). Once the trust closes to new loans, there will be a second observable period of nonrandom $\Delta \in \mathbb{N} \cup \{0\}$ months that the ABS is marketed to investors. Formally, then, the left-truncation random variable, Y, is finite, discrete with support $v \in \mathcal{V} \equiv \{\Delta + 1, \ldots, \Delta + m\}$. Because we observe X if and only if $X \geq Y$, X is a finite, discrete random variable with support $u \in \mathcal{U} \equiv \{\Delta + 1, \ldots, \omega\}$ (it is assumed $\omega \geq \Delta + m$). As a minor technical point, X may not be completely *recoverable* (Woodroofe, 1985; Lautier, Pozdnyakov and Yan, 2023b) (i.e., if $\Delta \geq 1$). We allow Y to take any shape over its support, which is more flexible than the uniformity assumption of *length-biased sampling* (e.g., Asgharian, M'Lan and Wolfson, 2002; Uña-Álvarez, 2004). This flexibility of Y better matches the cyclical pattern of auto loan originations (Lautier, Pozdnyakov and Yan, 2023b) and so better suits this application to auto ABS. Finally, we further assume X and Y are independent, which is justifiable within an application to auto ABS data (Lautier, Pozdnyakov and Yan, 2023a).

Beyond left-truncation, estimating the parameters of X from ABS data in practice will likely also require accounting for random right-censoring. ABS investors interested in modeling cash flows from an active ABS bond, for example, will need to construct estimates for X from many observations known to still be making ongoing monthly payments with yet unknown termination times. This complicates the left-truncation setting further (Lautier, Pozdnyakov and Yan, 2023a). To formalize, denote $\varepsilon \in \mathbb{N}$, $m + \Delta \leq \varepsilon \leq m + \omega$ to be the present time of the data generation process of an active ABS. If $\varepsilon < m + \omega$, then random right-censoring is present and $\Delta + 1 \leq X \leq \xi \equiv \min(\omega, \varepsilon - 1)$ (Lautier, Pozdnyakov and Yan, 2023a). Specifically, the exact termination time is observed if and only if $X \leq Y + \varepsilon - (m + \Delta + 1)$. In other words, if we define the random variable $C = Y + \varepsilon - (m + \Delta + 1) \equiv Y + \tau$, then the random right-censoring time is a linear shift of Y. This is convenient because C is a linear function of Y and $C \geq Y$ almost surely (Lautier, Pozdnyakov and Yan, 2023a). Thus, C is trivially independent of X. For a visualization of the ABS individual asset lifetime possibilities, see Lautier, Pozdnyakov and Yan (2023a, Fig. 1) or the Supplemental Material, Section A.

A final preliminary detail is that we assume throughout that X is dependent on the parameter, $p \in \mathcal{P}$, where $\mathcal{P}$ is a convex interval of $\mathbb{R}$. (The parameter $p \in \mathcal{P}$ need not be a scalar, though we assume so at present for ease of exposition.) Therefore, we desire to accurately estimate p from ABS data (i.e., subject to left-truncation, discrete-time, a known, finite support, and potentially right-censoring). Estimating p is our focus because it will completely specify X, and it is from X that the random ABS cash flows may be modeled (recall Table 1 or see Lautier, Pozdnyakov and Yan (2023a)). Furthermore, efficient and accurate estimation of p will help ABS investors gain rapid insights into the performance of the individual loans. These insights will help with investment allocation decisions and risk assessment (see Section 5). Our methods begin with the ABS setting first subject to only left-truncation.

3.2. *Left-Truncation.* We consider first the left-truncation ABS setting with no censored observations. Denote the probability mass function (pmf) of X by $f(u;p)$, $u \in \mathcal{U}$, $p \in \mathcal{P}$ and the pmf of Y by $g(v)$, $v \in \mathcal{V}$. The distribution Y is a parametric distribution, where each point of the pmf, $g(v)$, $v \in \mathcal{V}$, may be represented by a parameter, denoted g_v, $v \in \mathcal{V}$, $0 \le g_v \le 1$, such that $\sum_{\mathcal{V}} g_v = 1$. For convenience of notation, we use only g_v, $v \in \mathcal{V}$, in the sequel.

By the assumed independence of X and Y, we obtain the conditional bivariate pmf,

$$h_*(u,v;p) = \Pr(X = u, Y = v \mid Y \le X; p) = \frac{f(u;p)g_v}{\alpha}, \quad p \in \mathcal{P}, (u,v) \in \mathcal{A}, \tag{1}$$

where

$$\alpha \equiv \Pr(Y \le X) = \sum_{u=\Delta+1}^{\omega} f(u;p)\left(\sum_{v=\Delta+1}^{\min(u,\Delta+m)} g_v\right) = \sum_{v=\Delta+1}^{\Delta+m} g_v\left(\sum_{u=v}^{\omega} f(u;p)\right), \tag{2}$$

and $\mathcal{A} = \{\mathcal{U} \times \mathcal{V} : v \le u\}$. The distribution h_* is a parametric distribution with parameter vector $\boldsymbol{\Theta} = (p, \boldsymbol{g})^\top$, where $p \in \mathcal{P}$, $\boldsymbol{g} = (g_{\Delta+1}, \ldots, g_{\Delta+m})^\top \subset \boldsymbol{\mathcal{G}}$, and $\boldsymbol{\mathcal{G}}$ is an m-dimensional hypercube over the unit interval, $\mathcal{I} = (0,1)$.

Given an independent and identically distributed (i.i.d.) sample of pairs of left-truncated observations, $\mathcal{S}_n = \{(X_i, Y_i)\}_{1 \le i \le n}$, it is of interest to estimate the parameters of h_*. From (1) and (2), the likelihood is

$$\mathcal{L}(\boldsymbol{\Theta} \mid \mathcal{S}_n) = \prod_{v=\Delta+1}^{\Delta+m} \prod_{u=v}^{\omega} \left[\frac{f(u;p)g_v}{\alpha}\right]^{\sum_{i=1}^{n} \mathbf{1}_{(X_i,Y_i)=(u,v)}}. \tag{3}$$

If we denote the convex subset, $\mathcal{C} = \{\mathcal{P} \times \boldsymbol{\mathcal{G}} : \sum_v g_v = 1\} \subset \mathcal{P} \times \boldsymbol{\mathcal{G}}$, then we seek

$$\sup_{p, \boldsymbol{g} \in \mathcal{C}} \mathcal{L}(\boldsymbol{\Theta} \mid \mathcal{S}_n). \tag{4}$$

An approach to solve (4) without the assistance of computational programming is not immediate. Further, as the parameter space grows in dimension, performing the multidimensional constrained optimization numerically can become computationally demanding, complex in its implementation, and even potentially unfeasible (e.g., Murtagh and Saunders, 1978). We now show (4) may be reduced to a single-parameter optimization problem.

THEOREM 3.1 (Stationary points of $\mathcal{L}$ over $\mathcal{C}$). *Let $\mathcal{S}_n$ be an i.i.d. sample from the distribution $h_*(u,v;p)$ defined in (1) such that $\hat{h}_{\bullet v} := \sum_u \hat{h}_{uv} > 0$ and $\hat{h}_{u\bullet} := \sum_v \hat{h}_{uv} > 0$, where*

$\hat{h}_{uv} = \sum_i \mathbf{1}_{(X_i,Y_i)=(u,v)}/n$. *Further assume* $\partial f(u;p)/\partial p$ *exists and is finite for all* $p \in \mathcal{P}$, $u \in \mathcal{U}$. *Then the stationary points of* $\mathcal{L}(\boldsymbol{\Theta} \mid \mathcal{S}_n)$ *are*

$$\hat{g}_v = \frac{\hat{h}_{\bullet v}}{S(v;\hat{p})}\left[\sum_{k=\Delta+1}^{\Delta+m} \frac{\hat{h}_{\bullet k}}{S(k;\hat{p})}\right]^{-1}, \quad v \in \mathcal{V}, \tag{5}$$

where $S(\cdot)$ *denotes the* survival function*,*

$$S(x;p) \coloneqq \Pr(X \geq x; p) = \sum_{u=x}^{\omega} f(u;p), \tag{6}$$

and $\hat{p}$ *is any* $p \in \mathcal{P}$ *such that*

$$\sum_{v=\Delta+1}^{\Delta+m}\left(\frac{\hat{h}_{\bullet v}}{\sum_{u=v}^{\omega} f(u;p)}\right)\left(\sum_{u=v}^{\omega}\frac{\partial}{\partial p} f(u;p)\right) = \sum_{v=\Delta+1}^{\Delta+m}\sum_{u=v}^{\omega}\frac{\hat{h}_{uv}}{f(u;p)}\frac{\partial}{\partial p} f(u;p). \tag{7}$$

Further, $\hat{p} \in \mathcal{C}$ *and* $\hat{g}_v \in \mathcal{C}$ *for all* $v \in \mathcal{V}$*.*

PROOF. See the Supplemental Material, Section B.1. □

REMARK. The conditions $\hat{h}_{\bullet v} > 0$ and $\hat{h}_{u\bullet} > 0$ in Theorem 3.1 are necessary to avoid vacuous identifiability concerns for the boundaries of $\mathcal{A}$. These conditions are generally not a concern for the large n commonly found in ABS (see Table 2).

REMARK. The solution space (7) is a single equation consisting of finite sums for a single unknown, and differentiability conditions are imposed on f. This suggests that (7) can likely be solved using standard numeric optimization techniques (e.g., R Core Team, 2023, `optimize`) for a wide range of distributions, such as those admitted under the standard regularity conditions (e.g., van der Vaart, 1998, §5.3, pg. 51; Mukhopadhyay, 2000, §12.2, pg. 539). In some cases, closed-form solutions may exist (e.g., Theorem 3.3). Nonetheless, solving (7) may be challenging, especially in the multivariate case introduced momentarily.

The solution space (7) is a general form of an estimator for the parameter $p \in \mathcal{P}$ under the setting of Theorem 3.1 not yet derived to our knowledge. Once an estimate of p is found, $\hat{p}$, the lifetime distribution density estimate becomes $f(\cdot;\hat{p})$. Conveniently, the estimator (7), along with the closed-form solutions (5), reduce the multidimensional problem of (4) to a single dimension problem. This is valuable because a large parametric space for $\boldsymbol{g}$ is common when the lifetime of interest is consumer automobile monthly loan payments from ABS data. For example, a subset of data from AART (2017) generates $\mathcal{V} = \{4, \ldots, 24\}$ (see Section 5). This implies $\boldsymbol{\Theta}$ will contain 21 parameters to estimate. Without the dimension reduction of Theorem 3.1, solving (4) for this data is a daunting computational task. Indeed, we illustrate in Section 4 that numeric techniques can fail in this case.

We may also allow f to depend on a finite, r-dimensional parameter vector $\boldsymbol{p} \equiv (p_1, \ldots, p_r)^\top \subset \boldsymbol{\mathcal{P}}$, where $r < (\omega - \Delta)$ and $\boldsymbol{\mathcal{P}}$ is an r-dimensional convex set of $\mathbb{R}^r$. This allows for greater flexibility in modeling the lifetime distribution. The estimator (7) in Theorem 3.1 may be generalized appropriately, and we now provide discussion for completeness. Denote the pmf of X by $f(u;\boldsymbol{p})$, $u \in \mathcal{A}$. The equivalent notation for the conditional bivariate pmf in (1) then becomes $h_*(u,v;\boldsymbol{p})$, $(u,v) \in \mathcal{A}$. Let $\mathcal{S}_n' = \{(X_i, Y_i)\}_{1 \leq i \leq n}$ be an i.i.d. sample of left-truncated observations from the distribution $h_*(u,v;\boldsymbol{p})$. The multidimensional, constrained optimization problem is then to find $\sup \mathcal{L}(\boldsymbol{g}, \boldsymbol{p} \mid \mathcal{S}_n')$ such that $(\boldsymbol{p}, \boldsymbol{g}) \in \boldsymbol{\mathcal{C}}$, where

$$\boldsymbol{\mathcal{C}} = \left\{\boldsymbol{\mathcal{P}} \times \boldsymbol{\mathcal{G}} : \sum_{v \in \mathcal{V}} g_v = 1\right\} \subset \boldsymbol{\mathcal{P}} \times \boldsymbol{\mathcal{G}}, \tag{8}$$

and

$$\mathcal{L}(\boldsymbol{g},\boldsymbol{p}\,|\,\mathcal{S}_n') = \prod_{v=\Delta+1}^{\Delta+m}\prod_{u=v}^{\omega}\left[\frac{f(u;\boldsymbol{p})g_v}{\alpha}\right]^{\sum_{i=1}^{n}\mathbf{1}_{(X_i,Y_i)=(u,v)}}. \tag{9}$$

We have dropped the parametric notation, $\boldsymbol{\Theta}$, of (3) in the lead up to (9) to emphasize the replacement of p with the more general $\boldsymbol{p}$ in the parametric space. For completeness, α in (9) takes the same form as (2) but with $f(\cdot;\boldsymbol{p})$ replacing $f(\cdot;p)$. The formal result is stated in Corollary 3.1.1. For reference, Lautier, Pozdnyakov and Yan (2023b) study the limiting case, where each probability mass, $f(u)$, is itself represented by a parameter, f_u, $\Delta+1 \le u \le \omega$.

COROLLARY 3.1.1 (Stationary points of $\mathcal{L}$ over $\boldsymbol{\mathcal{C}}$). *Let $\mathcal{S}_n'$ be an i.i.d. sample of left-truncated observations from the distribution $h_*(u,v;\boldsymbol{p})$ subject to the same identifiability conditions of Theorem 3.1. Further assume $\partial f(u;\boldsymbol{p})/\partial p_j$ exists and is finite for all $j = 1,\dots,r$. Then the stationary points of $\mathcal{L}(\boldsymbol{g},\boldsymbol{p}\,|\,\mathcal{S}_n')$ are*

$$\hat{g}_v = \frac{\hat{h}_{\bullet v}}{S(v;\hat{\boldsymbol{p}})}\left[\sum_{k=\Delta+1}^{\Delta+m}\frac{\hat{h}_{\bullet k}}{S(k;\hat{\boldsymbol{p}})}\right]^{-1}, \quad v\in\mathcal{V}, \tag{10}$$

where $\hat{\boldsymbol{p}}$ is any $\boldsymbol{p}\in\boldsymbol{\mathcal{P}}$ that satisfies the system of equations

$$\sum_{v=\Delta+1}^{\Delta+m}\left(\frac{\hat{h}_{\bullet v}}{\sum_{u=v}^{\omega} f(u;\boldsymbol{p})}\right)\left(\sum_{u=v}^{\omega}\frac{\partial}{\partial p_j}f(u;\boldsymbol{p})\right) = \sum_{v=\Delta+1}^{\Delta+m}\sum_{u=v}^{\omega}\frac{\hat{h}_{uv}}{f(u;\boldsymbol{p})}\frac{\partial}{\partial p_j}f(u;\boldsymbol{p}), \tag{11}$$

for all $j = 1,\dots,r$. Further, $\hat{\boldsymbol{p}}\in\boldsymbol{\mathcal{C}}$ and $\hat{g}_v\in\boldsymbol{\mathcal{C}}$ for all $v\in\mathcal{V}$.

PROOF. See the Supplemental Material, Section B.2. □

REMARK. For an unspecified f, the $r\ge 2$ dimension of $\boldsymbol{p}$ in (11) will increase the complexity of the optimization problem in comparison to (7). It may also raise questions of uniqueness and/or identifiability. To illustrate these challenges, Section 5 provides a case study for the two-parameter discrete Weibull distribution of Nakagawa and Osaki (1975).

Beyond the point estimation results of Theorem 3.1 and Corollary 3.1.1, it is of interest to examine the asymptotic properties of (7) to assess estimation precision. (Recall, it is the lifetime distribution X that is of most practical importance to auto ABS investors.) To this end, a further advantage of (7) is that it takes the form of an asymptotically consistent M-estimator (van der Vaart, 1998, §5.3, pg. 51). Therefore, under the standard regularity conditions (e.g., van der Vaart, 1998, §5.3, pg. 51; Mukhopadhyay, 2000, §12.2, pg. 539) that span a wide range of parametric distributions with finite support, we may derive the exact form of its asymptotically normal distribution and provide practical techniques to estimate the asymptotic variance. We have used these properties to derive asymptotic confidence intervals (e.g., Table 3). For details, see the Supplemental Material, Sections B.3 and B.4. For an example of an extension to a vector $\boldsymbol{p}$ via (11), see the Supplemental Material, Section G.

We have thus far made only differentiation requirements on f. For specific choices of f, it is possible we can increase the claim in Theorem 3.1 from identifying stationary points of (4) to finding the MLEs of $p\in\mathcal{P}$ and $\boldsymbol{g}\in\boldsymbol{\mathcal{G}}$. Consider first Theorem 3.2.

THEOREM 3.2. *Assume the conditions of Theorem 3.1. Then (7) is satisfied if and only if*

$$\frac{\partial}{\partial p}\frac{\prod_{v=\Delta+1}^{\Delta+m} S(v;p)^{\hat{h}_{\bullet v}}}{\prod_{u=\Delta+1}^{\omega} f(u;p)^{\hat{h}_{u\bullet}}} = 0. \tag{12}$$

PROOF. See the Supplemental Material, Section B.5. □

A close study of (12) suggests candidates for f to yield a direct solution and thus an MLE. See, for example, Theorem 3.3, which yields an f with a closed-form MLE of $\boldsymbol{\Theta}$.

THEOREM 3.3 (MLE of $\boldsymbol{\Theta}$, PL geometric). *Define the policy limit (PL) geometric distribution with parameter,* $0 < p < 1$, *as*

$$f_G(u;p) = \begin{cases} p(1-p)^{u-(\Delta+1)} & \Delta+1 \le u \le \omega-1, \\ (1-p)^{u-(\Delta+1)} & u = \omega. \end{cases} \tag{13}$$

Then, for the conditional bivariate probability mass function, h_*, *defined in* (1) *under the sampling conditions of Theorem 3.1, the MLE of the parameter* p *is*

$$\hat{p}_{\text{MLE}} = \frac{b}{b-a}, \tag{14}$$

where

$$a = \sum_{v=\Delta+1}^{\Delta+m} \{v-(\Delta+1)\}\hat{h}_{\bullet v} - \sum_{u=\Delta+1}^{\omega} \{u-(\Delta+1)\}\hat{h}_{u\bullet}, \tag{15}$$

and

$$b = \sum_{u=\Delta+1}^{\omega-1} \hat{h}_{u\bullet}. \tag{16}$$

Further, the MLE of $\mathbf{g}$ *is*

$$\{\hat{g}_{v,\text{MLE}}\}_{v\in\mathcal{V}} = \hat{h}_{\bullet v}\left(1-\frac{b}{a}\right)^{v-(\Delta+1)}\left[\sum_{k=\Delta+1}^{\Delta+m} \hat{h}_{\bullet k}\left(1-\frac{b}{a}\right)^{k-(\Delta+1)}\right]^{-1}. \tag{17}$$

PROOF. See the Supplemental Material, Section B.6. □

The density function f_G defined in Theorem 3.3 is motivated by actuarial applications of statistical analysis to insurance policy limits. Following Klugman, Panjer and Willmot (2012, §8.4, pg. 125), a policy limit of ζ entitles an insured to the full repayment of losses for any amount below ζ with a maximum repayment of ζ for any loss greater than or equal to ζ. Hence, if L_1 denotes the loss random variable before the limit and L_2 denotes the loss random variable after the limit, the cdf of losses to the insurer becomes

$$F_{L_2}(\ell) = \begin{cases} F_{L_1}(\ell), & \ell < \zeta, \\ 1, & \ell \ge \zeta. \end{cases}$$

This setting motivates (13), and it has a natural application to the loan-level auto ABS analysis that motivates our study. Our reasoning stems from the observation that any probability for $u \ge \omega$ is loaded onto the final point, ω. Because many auto loans will stop making payments at the termination time dictated by the amortization schedule, such a weighting can be reasonable within an application to auto loan analysis (especially for the high-credit borrowers of Table 2). For reference, the Supplemental Material, Section B.7 contains a restatement of Theorem 3.3 with an alternative parameterization via a discretized, PL exponential distribution (i.e., $p > 0$). An alternative $\mathcal{P}$ space may have utility in generalized linear model (GLM) regression analysis build from the model of (1). For a further discussion, see Section 6.

3.3. *Right-Censoring.* We generalize the results of Section 3.2 for random right-censoring. The observed data now takes the triple $\mathcal{S}_{\tau,n} = \{(Y_i, \min(X_i, C_i), D_i)\}_{1 \le i \le n}$, where $D_i = 1$ if $X_i \le C_i$ and 0 otherwise, for $1 \le i \le n$. For convenience of notation, we define $Z_i = \min(X_i, C_i)$, for $1 \le i \le n$, and so $\mathcal{S}_{\tau,n} \equiv \{(Y_i, Z_i, D_i)\}_{1 \le i \le n}$. If there is no right-censoring present in the data, $D_i = 1$, for all $1 \le i \le n$, and $\mathcal{S}_{\tau,n}$ reduces to $\mathcal{S}_n$. The subscript τ appears in the present section to emphasize right-censoring is assumed present.

Let us now derive the likelihood. If $D_i = 0$ for any i, $1 \le i \le n$, which implies an observation is censored (i.e., $X_i > C_i$), the contribution to the likelihood is the probability

$$\begin{aligned}\Pr(Y_i = v, Z_i = u, D_i = 0) &= \Pr(Y_i = v, \min(X_i, C_i) = u, X_i > C_i) \\ &= \Pr(Y_i = v, X_i > u)\mathbf{1}(v + \tau = u) \\ &= \bar{h}_*(u, v; p),\end{aligned}$$

where

$$\bar{h}_*(u, v; p) = \Pr(Y = v, X \ge u + 1 \mid X \ge Y; p) = \frac{S(u+1; p) g_v}{\alpha},$$

for $p \in \mathcal{P}, (u, v) \in \mathcal{A}$. We may drop the indicator $\mathbf{1}(v + \tau = u)$ because $D_i = 0$ implies $v + \tau = u$ (for any i, $1 \le i \le n$, $D_i = 0 \implies X_i > C_i \implies Z_i = C_i = Y_i + \tau$). By the same reasoning, the contribution to the likelihood for $D_i = 1$, for some $1 \le i \le n$, is (1). (The Supplemental Material, Section C provides an illustrative example that h_* and $\bar{h}_*$ together form a valid density for all possible outcomes of a single sample, (Y_i, Z_i, D_i), for any, i, $1 \le i \le n$.) Thus, the likelihood for $\mathcal{S}_{\tau,n}$ becomes

$$\begin{aligned}\mathcal{L}_\tau(\boldsymbol{\Theta} \mid \mathcal{S}_{\tau,n}) &= \prod_{\{\mathcal{S}_{\tau,n}: D_i = 1\}} h_*(Z_i, Y_i; p) \prod_{\{\mathcal{S}_{\tau,n}: D_i = 0\}} \bar{h}_*(Z_i, Y_i; p) \\ &= \alpha^{-n} \prod_{v=\Delta+1}^{m+\Delta} g_v^{n\hat{\gamma}_n(v)} \prod_{i=1}^{n} f(Z_i; p)^{D_i} S(Z_i + 1; p)^{1 - D_i}, \qquad (18)\end{aligned}$$

where $\hat{\gamma}_n(v) = \sum_i \mathbf{1}_{Y_i = v}/n$. If $D_i = 1$ for all i, $1 \le i \le n$, then $\mathcal{L}_\tau(\boldsymbol{\Theta} \mid \mathcal{S}_{\tau,n})$ reduces to $\mathcal{L}(\boldsymbol{\Theta} \mid \mathcal{S}_n)$ of Section 3.2. As with Theorem 3.1, we seek

$$\sup_{p, \boldsymbol{g} \in \mathcal{C}} \mathcal{L}_\tau(\boldsymbol{\Theta} \mid \mathcal{S}_{\tau,n}). \qquad (19)$$

REMARK. It is assumed in (18) and all following statements that terms involving $f(\cdot; p)$ only appear when $D_i = 1$ and, conversely, terms involving $S(\cdot; p)$ only appear when $D_i = 0$. This avoids any complications when $D_i = 1$ and $S(\cdot; p) = 0$. It is understood this convention may be coded easily, and we assume the form of (18) for ease of exposition.

THEOREM 3.4 (Stationary points of $\mathcal{L}_\tau$ over $\mathcal{C}$). *Let $\mathcal{S}_{\tau,n}$ be an i.i.d. sample of left-truncated observations from the distribution $h_*(u, v; p)$ defined in (1) under the additional incomplete data setting of right-censoring. Assume the identifiability and differentiability conditions of Theorem 3.1. Then the stationary points of $\mathcal{L}_\tau(\boldsymbol{\Theta} \mid \mathcal{S}_{\tau,n})$ are*

$$\hat{g}_{\tau,v} = \frac{\hat{\gamma}_n(v)}{S(v; \hat{p}_\tau)} \left[\sum_{k=\Delta+1}^{\Delta+m} \frac{\hat{\gamma}_n(k)}{S(k; \hat{p}_\tau)} \right]^{-1}, \quad v \in \mathcal{V},$$

where $S(\cdot)$ denotes the survival function *defined in (6), and $\hat{p}_\tau$ is any $p \in \mathcal{P}$ such that*

$$\sum_{v=\Delta+1}^{\Delta+m} \left(\frac{\hat{\gamma}_n(v)}{\sum_{u=v}^{\xi} f(u; p)} \right) \left(\sum_{u=v}^{\xi} \frac{\partial}{\partial p} f(u; p) \right) \qquad (20)$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{D_i}{f(Z_i;p)}\frac{\partial}{\partial p}f(Z_i;p) + \frac{1-D_i}{S(Z_i+1;p)}\frac{\partial}{\partial p}S(Z_i+1;p)\right).$$

Further, $\hat{p}_\tau \in \mathcal{C}$ and $\hat{g}_{\tau,v} \in \mathcal{C}$, for all $v \in \mathcal{V}$.

PROOF. See the Supplemental Material, Section B.8. □

The solution space (20) is a general form of an estimator for the parameter $p \in \mathcal{P}$ under the left-truncation and right-censoring incomplete data setting of Theorem 3.4 not yet derived to our knowledge. As with Theorem 3.1, Theorem 3.4 reduces a potentially complex and computationally demanding multidimensional constrained optimization problem, (19), into a single parameter optimization problem, (20). These comments echo those following Theorem 3.1, and we omit further discussion to avoid unnecessary repetition.

We may also allow f to depend on a finite, r-dimensional parameter vector, $\boldsymbol{p}$ (i.e., generalizing Corollary 3.1.1 to accommodate right-censoring). Under $D_i = 0$, $1 \le i \le n$, we denote the likelihood contribution, $\bar{h}_*(u,v;\boldsymbol{p})$. We therefore seek $\sup \mathcal{L}_\tau(\boldsymbol{g},\boldsymbol{p} \mid \mathcal{S}_{\tau,n})$ such that $(\boldsymbol{p},\boldsymbol{g}) \in \boldsymbol{\mathcal{C}}$, where $\boldsymbol{\mathcal{C}}$ follows (8), and

$$\mathcal{L}_\tau(\boldsymbol{g},\boldsymbol{p} \mid \mathcal{S}_{\tau,n}) = \alpha^{-n}\prod_{v=\Delta+1}^{m+\Delta} g_v^{n\hat{\gamma}_n(v)}\prod_{i=1}^{n}\left(f(Z_i;\boldsymbol{p})^{D_i}S(Z_i+1;\boldsymbol{p})^{1-D_i}\right).$$

We again drop the parametric notation, $\boldsymbol{\Theta}$, of $\mathcal{L}_\tau$ to emphasize the replacement of p with the more general $\boldsymbol{p}$ in the parametric space. The formal result is Corollary 3.4.1. For reference, Lautier, Pozdnyakov and Yan (2023a) study the limiting case, where each probability mass, $f(u)$, is itself represented by a parameter, f_u, $\Delta + 1 \le u \le \xi$.

COROLLARY 3.4.1 (Stationary points of $\mathcal{L}_\tau$ over $\boldsymbol{\mathcal{C}}$). *Let $\mathcal{S}_{\tau,n}$ be an i.i.d. sample of left-truncated observations from the distribution $h_*(u,v;\boldsymbol{p})$ under the additional incomplete data setting of right-censoring. Assume the identifiability and differentiability conditions of Theorem 3.1. Then the stationary points of $\mathcal{L}_\tau(\boldsymbol{g},\boldsymbol{p} \mid \mathcal{S}_{\tau,n})$ are*

$$\hat{g}_{\tau,v} = \frac{\hat{\gamma}_n(v)}{S(v;\hat{\boldsymbol{p}}_\tau)}\left[\sum_{k=\Delta+1}^{\Delta+m}\frac{\hat{\gamma}_n(k)}{S(k;\hat{\boldsymbol{p}}_\tau)}\right]^{-1}, \quad v \in \mathcal{V}, \tag{21}$$

where $\hat{\mathbf{p}}_\tau$ is any $\mathbf{p} \in \boldsymbol{\mathcal{P}}$ that satisfies the system of equations

$$\sum_{v=\Delta+1}^{\Delta+m}\left(\frac{\hat{\gamma}_n(v)}{\sum_{u=v}^{\xi} f(u;\boldsymbol{p})}\right)\left(\sum_{u=v}^{\xi}\frac{\partial}{\partial p_j}f(u;\boldsymbol{p})\right) \tag{22}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{D_i}{f(Z_i;\boldsymbol{p})}\frac{\partial}{\partial p_j}f(Z_i;\boldsymbol{p}) + \frac{1-D_i}{S(Z_i+1;\boldsymbol{p})}\frac{\partial}{\partial p_j}S(Z_i+1;\boldsymbol{p})\right),$$

for all $j = 1,\ldots,r'$. Further, $\hat{\mathbf{p}}_\tau \in \boldsymbol{\mathcal{C}}$ and $\hat{g}_{\tau,v} \in \boldsymbol{\mathcal{C}}$ for all $v \in \mathcal{V}$.

PROOF. See the Supplemental Material, Section B.9. □

REMARK. The same Remark following Corollary 3.1.1 also applies for (22) versus (20).

Conveniently, the estimator (20) also takes the form of an asymptotically consistent M-estimator (van der Vaart, 1998, §5.3, pg. 51). Therefore, under the standard regularity conditions (e.g., van der Vaart, 1998, §5.3, pg. 51; Mukhopadhyay, 2000, §12.2, pg. 539), asymptotic normality with practical variance estimation techniques follow to assess estimation precision. For details, see the Supplemental Material, Sections B.10, B.11, and, for the vector extension, $\boldsymbol{p}$, Section G. We close this section by generalizing Theorem 3.3.

COROLLARY 3.4.2 (MLE of $\boldsymbol{g}$, p, PL geometric, right-censoring). *Assume f_G follows (13). Then, for the conditional bivariate probability mass function, h_*, defined in (1) under the sampling conditions of Theorem 3.4, the MLE of the parameter p is*

$$\hat{p}_{\tau,\text{MLE}} = \frac{b_\tau}{b_\tau - a_\tau},$$

where

$$a_\tau = \sum_{v=\Delta+1}^{\Delta+m} \{v-(\Delta+1)\}\hat{\gamma}_n(v) - \frac{1}{n}\sum_{i=1}^{n}\big(\{Z_i-(\Delta+1)\}D_i + \{Z_i+1-(\Delta+1)\}(1-D_i)\big),$$

and

$$b_\tau = \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}(Z_i \neq \xi)D_i.$$

Further, the MLE of $\boldsymbol{g}$ is

$$\{\hat{g}_{\tau,v,\text{MLE}}\}_{v\in\mathcal{V}} = \hat{\gamma}_n(v)\left(1-\frac{b_\tau}{a_\tau}\right)^{v-(\Delta+1)}\left[\sum_{k=\Delta+1}^{\Delta+m}\hat{\gamma}_n(k)\left(1-\frac{b_\tau}{a_\tau}\right)^{k-(\Delta+1)}\right]^{-1}.$$

PROOF. See the Supplemental Material, Section B.12. □

4. Simulation Studies. We first use simulation to provide an example for which the reduction in dimension of the constrained optimization problem of (4) through Theorem 3.1 accurately estimates the true parameter values for $\boldsymbol{\Theta}$ while a full dimension, direct numeric optimization technique fails. Next, we validate the asymptotic properties of the estimators proposed in Section 3 for (13) via a robustness coverage probability simulation study. For completeness, the Supplemental Material, Section F provides a numeric validation of all theoretical results in Section 3, including each estimator's asymptotic properties (the latter stated formally in the Supplemental Material, Section B). These additional simulation studies in the supplement also include parametric distributions beyond (13), such as the binomial distribution and the two-parameter PL discrete Weibull distribution introduced later in Section 5.

To begin, we select a trapezoid $\mathcal{A}$ reflective of what may be encountered in an application to auto ABS (see Section 5). Specifically, we assume $m = 20$, $\Delta = 0$, and $\omega = 24$. The lifetime distribution, X, follows (13) with $p = 0.05$. The left-truncation random variable is a weighted-mixture of shifted binomial distributions with 9 trials and success probability of 0.35, denoted $\mathcal{B}$. That is, $\Pr(Y = v) = 0.4 \times \mathbf{1}(1 \le v \le 10)\mathcal{B}(v-1) + 0.6 \times \mathbf{1}(11 \le v \le 20)\mathcal{B}(v-11)$. We simulate a size $n = 1{,}000$ sample of pairs $(Y_i, X_i)_{1\le i\le 1{,}000}$. It is instructive to see if (4) can be solved with a direct, constrained numeric optimization using `constroOptim` via R Core Team (2023). Because there are 20 parameters to be estimated, this is not a straightforward task. The initial values were selected to be uninformative with $p_{\text{INIT}} = 0.5$ and $g_{\text{INIT},v} = 0.05$ for $1 \le v \le 20$. After 25.07 minutes under default settings and producing a result, `constrOptim` was unable to recover the true parameter values. Conversely, Theorems 3.1 and 3.3 were able to recover the true parameter values after 0.40 and 0.11 seconds, respectively. For details and computer specifications, see the Supplemental Material, Section F. This demonstrates that the closed-form solutions (5) in Theorem 3.1 that reduce the multidimensional problem of (4) to a single dimension problem can rapidly find accurate estimates when a direct, constrained numeric optimization may fail.

Our second simulation study is a traditional robustness analysis of the methods we propose to sample size and level of right-censoring. The distributions for X and Y are unchanged from the previous paragraph. Because an application to ABS data is primarily focused on the lifetime distribution of the loans, X, we focus on estimation of p. We consider sample sizes $n \in \{50, 100, 250, 500\}$. We may control for right-censoring by varying the current time, ε, $m + \Delta + 1 \le \varepsilon \le m + \omega$. The equivalent censoring rate probability, $\mathcal{C}_\varepsilon$, is then

$$\mathcal{C}_\varepsilon = \sum_{u=v+\tau} \bar{h}_*(u, v). \tag{23}$$

As $\varepsilon \downarrow (m + \Delta + 1)$, $\mathcal{C}_\varepsilon \uparrow 1$. We consider $\varepsilon \in \{26, 32, 38\}$ and a case of no right-censoring. For each combination of n and $\mathcal{C}_\varepsilon$, we simulate a random sample from h_*. We then estimate p using either Corollary 3.4.2 or Theorem 3.3, as appropriate. This process repeats for 1,000 replicates. We then calculate the empirical mean and standard deviation of the 1,000 estimates of p. We compare the empirical mean with the true parameter value, $p = 0.05$, and we compare the empirical standard deviation with the theoretical value we calculate using the asymptotically consistent M-estimator theory (see the Supplemental Material, Section B). To evaluate the practical utility of these asymptotic results, we report an empirical coverage probability of the 95% asymptotic confidence intervals. The complete results may be found in Table 3. The methods we propose are robust to small samples and right-censoring rate, which is an attractive property for an application to ABS data (and more generally).

TABLE 3
Robustness Simulation Study. *A robustness analysis of §3 estimation methods for $X \sim$ (13), $Y \sim \mathcal{B}$, $m = 20$, $\Delta = 0$, and $\omega = 24$ by n and $\mathcal{C}_\varepsilon$. We report the empirical mean (eMean), empirical standard deviation (eSD), and coverage probability (CP) for 95% asymptotic confidence intervals.*

		$\varepsilon = 26$ ($\mathcal{C}_\varepsilon = 73.4\%$)				$\varepsilon = 32$ ($\mathcal{C}_\varepsilon = 32.0\%$)			
n	p_0	eMean	eSD	Thm B.2	CP	eMean	eSD	Thm B.2	CP
50	0.05	0.0503	0.0137	0.0134	0.923	0.0509	0.0105	0.0105	0.943
100	0.05	0.0502	0.0095	0.0095	0.936	0.0503	0.0072	0.0074	0.953
250	0.05	0.0501	0.0063	0.0060	0.938	0.0503	0.0046	0.0047	0.961
500	0.05	0.0500	0.0042	0.0042	0.959	0.0501	0.0034	0.0033	0.943
		$\varepsilon = 38$ ($\mathcal{C}_\varepsilon = 20.0\%$)				No Censoring ($\mathcal{C}_\varepsilon = 0\%$)			
n	p_0	eMean	eSD	Thm B.2	CP	eMean	eSD	Thm B.1	CP
50	0.05	0.0506	0.0098	0.0097	0.943	0.0503	0.0098	0.0095	0.947
100	0.05	0.0504	0.0068	0.0069	0.951	0.0507	0.0066	0.0067	0.951
250	0.05	0.0500	0.0043	0.0043	0.953	0.0501	0.0042	0.0043	0.950
500	0.05	0.0500	0.0031	0.0031	0.940	0.0501	0.0029	0.0030	0.954

5. Application.

We apply the methods of Section 3 to the ABS bonds AART (2017, 2019), introduced in Section 2. When proposing a parametric f, it is desirable to assess model specification. Section 5.1 thus provides a LRT that may be used to formally evaluate any f against the unrestricted approach of Lautier, Pozdnyakov and Yan (2023a). Section 5.2 then fits and evaluates (13) and a two-parameter PL version of the discrete Weibull distribution of Nakagawa and Osaki (1975). For the latter, we illustrate the potential challenges in the multidimensional optimization required by Corollaries 3.1.1 and 3.4.1. Section 5.3 then contextualizes these results from the perspective of an ABS investor.

5.1. *Model Specification.*

The methods of Section 3 apply for any discrete, parametric f with finite support (under the differentiation conditions of Theorem 3.1). Hence, an investor is free to choose any f that meets these criteria. Left unanswered to this point, however, is how to evaluate one f versus another. This is the purpose of the present section.

In the discrete-time with a known, finite support setting, Lautier, Pozdnyakov and Yan (2023a) assume each probability point mass of f is a parameter. In other words, just as g_v, $\Delta+1 \le v \le \Delta+m$ is a parameter in (1), $f(u;p)$ may be described by a parameter, f_u, $\Delta+1 \le u \le \xi$. The form of (1) is then $h_*(u,v) = f_u g_v/\alpha$ for $(u,v) \in \mathcal{A}$. In this case, f has $\xi - \Delta$ free parameters (by requiring $\sum_u f_u = 1$). Hence, for any f such that the dimension of $\boldsymbol{p}$ is less than $\xi - \Delta$, it may be considered a restricted parameterization in comparison to Lautier, Pozdnyakov and Yan (2023a). This motivates a classical LRT as follows.

1. Use Lautier, Pozdnyakov and Yan (2023a) to estimate $\boldsymbol{f} = (f_{\Delta+1}, \ldots, f_\xi)^\top$. Specifically,

$$\hat{f}_u = \hat{\lambda}_{\tau,n}(u) \prod_{k=\Delta+1}^{u-1} (1 - \hat{\lambda}_{\tau,n}(k)), \quad \Delta+1 \le u \le \xi,$$

   where

$$\hat{\lambda}_{\tau,n}(x) = \frac{\sum_i \mathbf{1}(D_i = 1)\mathbf{1}(Z_i = x)}{\sum_i \mathbf{1}(Y_i \le x \le Z_i)}.$$

   Then estimate $\boldsymbol{g}$ using (3.4). Denote this unrestricted estimate as $\hat{\boldsymbol{\Theta}}_1 = (\hat{\boldsymbol{f}}, \hat{\boldsymbol{g}}_1)^\top$.
2. For restricted $f(u;\boldsymbol{p})$, use Theorem 3.4 to estimate $\boldsymbol{\Theta}_0 = (\boldsymbol{p}, \boldsymbol{g})$, denoted $\hat{\boldsymbol{\Theta}}_0 = (\hat{\boldsymbol{p}}_\tau, \hat{\boldsymbol{g}}_0)^\top$.
3. Calculate the LRT statistic, Λ_n, with $\mathcal{L}_\tau$ from (18). Under suitable regularity conditions (Lehmann and Romano, 2006, Theorem 12.4.2, pg. 515), Λ_n converges in distribution to a chi-square distribution with $\delta = \xi - (\Delta+1) - \#\{\boldsymbol{p}_\tau\}$ degrees of freedom, χ^2_δ,

$$\Lambda_n = 2\log\left(\frac{\mathcal{L}_\tau(\hat{\boldsymbol{\Theta}}_1 \mid \mathcal{S}_{\tau,n})}{\mathcal{L}_\tau(\hat{\boldsymbol{\Theta}}_0 \mid \mathcal{S}_{\tau,n})}\right) \longrightarrow_d \chi^2_\delta.$$

4. For significance level, $0 < \mu < 1$, let $\chi^2_{\mu,\delta}$ denote the $100 \times (1-\mu)$th percentile of χ^2_δ. Then to test $H_0 : \boldsymbol{\Theta} = \boldsymbol{\Theta}_0$ at significance level μ, reject H_0 if $\Lambda_n > \chi^2_{\mu,\delta}$. Rejecting H_0 implies $f(u;\boldsymbol{p})$ is less preferable to describe $\mathcal{S}_{\tau,n}$ than Lautier, Pozdnyakov and Yan (2023a).

For restricted f following (13), an attractive feature of this LRT is that Λ_n may be constructed from closed-form parameter estimates, which significantly simplifies implementation. For reference, $\Lambda_n \longrightarrow_d \chi^2_\delta$ has been verified numerically (Supplemental Material, Section G).

The above LRT assumes both left-truncation and right-censoring are present. If no right-censoring occurs, Section 3.3 results simplify to Section 3.2, and the above holds (set $D_i = 1$ for all $1 \le i \le n$). Furthermore, the LRT approach may be used to compare two parametric choices of f with subset parameterizations (e.g., see Table 4). In addition to the LRT, a visual comparison may also be informative (e.g., see Figure 2).

5.2. *Empirical Analysis.* We now apply our methods to the auto ABS bonds AART (2017) and AART (2019) introduced in Section 2. Because the loan terms in AART (2017) range from 12 to 78 months, and each loan term must be treated separately, we initially focus on a subset of $n = 151$ loans with a loan term of 25 months. For these 25-month loans, we have $m = 21$, $\Delta = 3$, and $\omega = 26$. There are thus 21 parameters to be estimated, which can make obtaining a solution to (4) or (19) numerically unfeasible without the methods we propose (see Section 4). We use the full 43 months of performance data, which sets $\varepsilon = 67 > m + \omega$. Hence, there is no right-censoring present, and we use the methods from Section 3.2. As a minor data adjustment, any observations with a termination time of 27, 28, or 29 months are considered terminated at $\omega = 26$ months. Such small extensions may be an artifact of financial reporting and generally have a small impact on estimated profitability.

We apply Theorem 3.3 and the asymptotic results of the Supplemental Material, Section B to estimate $\hat{p}^{\text{17-25}}_{\text{MLE}} = 0.0313$ with a 95% asymptotic confidence interval of $(0.0226, 0.0399)$.

Let us now consider the question of model specification. We first investigate visually by comparing it to Lautier, Pozdnyakov and Yan (2023a), which makes no structural assumptions about the shape of f. That is, we estimate the discrete hazard rate, $\lambda(x) = \Pr(X = x \mid X \geq x)$, plus 95% asymptotic confidence intervals for each loan age via Lautier, Pozdnyakov and Yan (2023a). Because (13) assumes a constant hazard rate, we would expect $\hat{p}_{\text{MLE}}^{\text{17-25}} = 0.0313 \pm 0.0086$ to fall between the 95% asymptotic confidence intervals of Lautier, Pozdnyakov and Yan (2023a). A visual of this comparison occurs in the top left panel of Figure 2. The point estimates of the hazard rate via Lautier, Pozdnyakov and Yan (2023a) are the solid blue line, and the blue ribbon represents each estimate's 95% asymptotic confidence intervals. The horizontal red dashed line plus red ribbon represents $\hat{p}_{\text{MLE}}^{\text{17-25}} = 0.0313 \pm 0.0086$. Because the 95% asymptotic confidence intervals for each method consistently overlap, it is initial evidence that (13) with a parameter of $\hat{p}_{\text{MLE}}^{\text{17-25}} = 0.0313 \pm 0.0086$ is a reasonable choice to model this loan lifetime data.

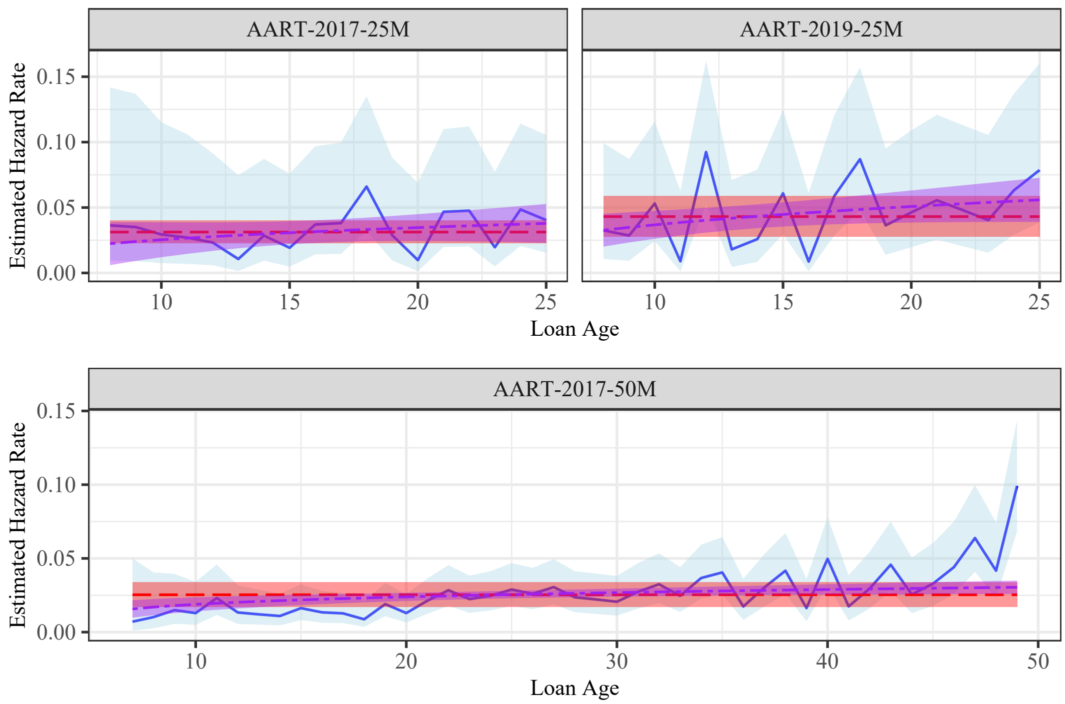

FIG 2. ***Model Specification, Visual Assessment***. *A comparison of $\hat{\lambda}$ via Lautier, Pozdnyakov and Yan (2023a) (solid, blue) against (13) (dashed, red) and (24) (two-dashed, purple) by bond AART (2017, 2019) and original loan term (25, 50 months). A consistent overlap of the asymptotic 95% confidence intervals between Lautier, Pozdnyakov and Yan (2023a) and proposed models provides initial visual evidence that a proposed parametric model may obtain an adequate fit to this data.*

REMARK. Asymptotic confidence intervals for λ follow from the Delta Method (e.g., Lehmann and Casella, 1998, Theorem 8.12, pg. 58) (see Supplemental Material, Section G).

After passing the visual inspection, we next test (13) using the LRT of Section 5.1. We calculate $\Lambda_n^{\text{17-25}} = 18.357$, which follows a chi-square distribution with $26 - (3+1) - 1 = 21$ degrees of freedom. The resulting p-value is 0.6263, and so we fail to reject (13) with parameter $\hat{p}_{\text{MLE}}^{\text{17-25}} = 0.0313 \pm 0.0086$ as an adequate choice to describe 25-month loans from AART (2017) at any reasonable significance level. This confirms the visual analysis of Figure 2.

In some cases, (13) cannot adequately describe AART (2017, 2019) data, however. We now consider 25-month loans from AART (2019), issued two years later than AART (2017). The sample size is $n_{2019} = 178$ with $m_{2019} = 19$, $\Delta_{2019} = 2$, and $\omega_{2019} = 26$. This results in 19 parameters to be estimated. AART (2019) was actively paying for 46 months, which sets $\varepsilon_{2019} = 67 > m_{2019} + \Delta_{2019}$. Thus, no right-censoring is present. As with AART (2017), we set any loan with termination times beyond month 26 to $\omega_{2019} = 26$. We next apply Theorem 3.3 and the asymptotic results of the Supplemental Material, Section B to estimate $\hat{p}^{19\text{-}25}_{\text{MLE}} = 0.0431$ with a 95% asymptotic confidence interval of $(0.0337, 0.0524)$. A visual analysis appears in Figure 2. The 95% asymptotic confidence intervals of $\hat{p}^{19\text{-}25}_{\text{MLE}}$ and Lautier, Pozdnyakov and Yan (2023a) generally overlap, which suggests (13) may be adequate. To assess further, we apply the LRT of Section 5.1. We calculate $\Lambda^{19\text{-}25}_n = 37.229$, which follows a chi-square distribution with 22 degrees of freedom. The resulting p-value is 0.0223, and so we reject (13) as an adequate choice for this data at significance level $\mu = 0.05$.

A visual analysis can intimate alternative choices for f. The apparent increasing linear hazard rate suggests an f with this feature. One example of a discrete f capable of modeling an increasing hazard rate is the discrete Weibull of Nakagawa and Osaki (1975),

$$\Pr(X = x; p_1, p_2) = p_1^{x^{p_2}} - p_1^{(x+1)^{p_2}}, \quad x = 0, 1, 2, \dots, \tag{24}$$

where $0 < p_1 < 1$ and $p_2 > 0$. It has two parameters, and so Corollaries 3.1.1 and 3.4.1 apply. A policy limit version of (24) (i.e., Klugman, Panjer and Willmot, 2012, §8.4, pg. 125) is

$$f_W(u; p_1, p_2) = \begin{cases} p_1^{(u-(\Delta+1))^{p_2}} - p_1^{(u-\Delta)^{p_2}}, & \Delta + 1 \le u \le \omega - 1 \\ p_1^{(u-(\Delta+1))^{p_2}}, & u = \omega. \end{cases} \tag{25}$$

Per Corollary 3.1.1, it is necessary to find p_1 and p_2 such that for both $i \in \{1, 2\}$,

$$\sum_{v=\Delta+1}^{\Delta+m} \left(\frac{\hat{h}_{\bullet v}}{\sum_{u=v}^{\omega} f_W(u; p_1, p_2)} \right) \left(\sum_{u=v}^{\omega} \frac{\partial}{\partial p_i} f_W(u; p_1, p_2) \right) = \sum_{v=\Delta+1}^{\Delta+m} \sum_{u=v}^{\omega} \frac{\hat{h}_{uv}}{f_W(u; p_1, p_2)} \frac{\partial}{\partial p_i} f_W(u; p_1, p_2).$$

The increase in dimension from a single equation to optimize over a single variable to two equations to optimize over two parameters increases complexity. For example, the risk increases that a unique solution may not exist or that the individual components of $\boldsymbol{p}$ cannot be identified for a $\hat{\boldsymbol{p}}$ satisfying (20). Even more fundamentally, numeric optimization is well-suited to locate local max/minima, but it is uncertain if such points are global max/minima.

Because there are many possibilities of f, it is difficult to recommend a universal approach, but we offer suggestions for (25) that may help inspire approaches for other vector-parameter distributions. First, Figure 2 suggests an increasing hazard rate. This requires $p_2 > 1$. Further, for 25-month loans, f_W needs to span over 20 integers without rapid decay. Hence, if $p_2 > 1$, both p_1 and p_2 will need to be close to 1. These observations can shrink the initial search bounds for numeric optimization techniques, such as $0.5 < p_1 < 1$ and $1 < p_2 < 2$. Then, each search region can be cut into a set of discrete points, and a cross product can be taken. This allows for a visual plot of the two-dimensional optimization equation. This plot can be used to narrow the search regions further, and the process can be repeated until the bounds are suitably narrow for the numeric optimization program to produce a precise solution set (for details, see the Supplemental Material, Section G). Visual approaches may be less viable for higher dimensions. The field of mathematical optimization is vast, however, and classical texts can offer guidance (e.g., Luenberger and Ye, 2021). The asymptotic theory may also be expanded for a vector parameter (see the Supplemental Material, Section G).

Applying this approach, we estimate $\hat{p}_1^{19\text{-}25} = 0.9877$ and $\hat{p}_2^{19\text{-}25} = 1.3888$. The top right panel of Figure 2 shows the desired increasing hazard rate (purple two-dashed line and ribbon). For the LRT, we calculate $\Lambda_n^{19\text{-}25} = 32.671$, which follows a chi-square distribution with 21 degrees of freedom. The resulting p-value is 0.1231, and so we fail to reject (25) with parameters $\hat{p}_1^{19\text{-}25} = 0.9877$ and $\hat{p}_2^{19\text{-}25} = 1.3888$ as a viable choice to model 25-month loans from AART (2019). The added flexibility of (25) in comparison to (13) is sufficient enough to improve the fit, and its simplicity in comparison to Lautier, Pozdnyakov and Yan (2023a) is preferred by the LRT. We may continue this process for various loan terms. As an example in which neither (13) nor (25) are adequate is for 50-month loans from AART (2017), as the LRT rejects both at any reasonable significance level. For a summary, see Table 4.

An advantage of the methods we propose is that alternatives for f may be selected, and the results of Section 3 will continue to substantially simplify the parameter estimation problem. Then, the LRT approach of Section 5.1 may be used to formally assess the new choice of f. This flexibility is attractive in that more complex choices of f may be proposed, such as a spliced distribution that produces a different constant hazard rate at different time points. Such a distribution will still be a parametric subset of Lautier, Pozdnyakov and Yan (2023a), and the LRT will apply. While the optimization challenges may increase as the number of parameters in f increases, as demonstrated with (25), Section 3 provides direct formulas to recover $\boldsymbol{g}$. As the length of $\boldsymbol{g}$ can be large, this simplification helps to estimate $\boldsymbol{\Theta}$.

TABLE 4

Model Specification, Empirical Results. *An application of the LRT of Section 5.1 for different loan terms from AART (2017, 2019). The notation $\boldsymbol{f}$ denotes the unrestricted Lautier, Pozdnyakov and Yan (2023a), f_G denotes (13), and f_W denotes (25). A p-value above 0.05 indicates the simpler model cannot be rejected as a viable choice to model the ABS loan data at the 5% significance level.*

Sample Details			$\log \mathcal{L}$			LRT p-values (H_0/H_1)		
Bond	Term	n	$\boldsymbol{f}$	f_G	f_W	$f_G/\boldsymbol{f}$	$f_W/\boldsymbol{f}$	f_G/f_W
AART (2017)	25-mo.	151	−668.01	−677.19	−675.30	0.6263	0.7996	0.0522
AART (2019)	25-mo.	178	−865.15	−883.77	−879.47	0.0223	0.1231	0.0034
AART (2017)	50-mo.	692	−3,926.47	−3,989.69	−3,970.00	<0.00001	0.00009	<0.00001

5.3. *Investor Perspectives.* For ABS investors, a stochastic cash flow model is necessary for pricing and risk management. The ABS cash flows are a function of the total monthly cash flow of its individual loans (see Table 1). Further, the monthly cash flow of each individual loan is a function of its random time-to-termination. In other words, we can use the random variable, X, of which its estimation has been the focus of this manuscript. Specifically, for any consumer auto loan at age x, its expected cash flow at age $x+1$ can be approximated by

$$\mathbf{E}[\mathrm{CF}_{x+1}] = \mathrm{PMT} \times (1 - \lambda(x)) + \mathrm{BAL}_{x+1} \times \lambda(x), \tag{26}$$

where PMT and BAL_{x+1} represent the monthly payment and outstanding principle balance at age $x+1$, respectively. Because PMT and BAL are observable, we can construct a stochastic loan-level cash flow model by estimating X and building up to the ABS cash flows via Table 1. This is the framework for cash flow models constructed at the individual asset level, which is the preferred method to model credit risk (Lautier, Pozdnyakov and Yan, 2023a).

The applied methods we propose use a traditional parametric model for X, which can offer ABS investors a few advantages. The first is its simplicity. For example, (13) offers closed-form MLE solutions (Theorem 3.3, Corollary 3.4.2). Furthermore, the parameter of (13) is equivalent to its hazard rate, which flows directly into (26). A larger value of p will shorten the duration of the monthly payments in (26), thereby potentially shortening the duration of the ABS cash flows. This shorter duration can result in a lower return to ABS investors. Thus, with little effort, an ABS investor can get an immediate insight into its potential cash flows. Drawing such inference quickly is valuable in the fast-paced environment of financial trading, where many decisions must be made quickly amid shifting prices in the open market.

Similarly, it is often desirable for ABS investors to analyze trends over time. Our methods can help with this goal. To demonstrate, observe that AART (2017, 2019) are from the same parent and are similar by pool level summary statistics (see Table 2). Hence, investors would expect the lifetime distribution of individual loans estimated from each bond to be fairly similar. Recall for 25-month loans, we estimated $\hat{p}_{\text{MLE}}^{17\text{-}25} = 0.0313\ (0.0226, 0.0399)$ and $\hat{p}_{\text{MLE}}^{19\text{-}25} = 0.0431\ (0.0337, 0.0524)$. While the 95% asymptotic confidence intervals overlap, it is only slightly so. This could suggest a possible structural change in the nature of the underlying individual consumer auto loans within these two bonds or reflect a changing economic environment. The LRT analysis of Table 4 also supports this sentiment. In other words, (13) is a reasonable choice to model 25-month loans from AART (2017), but it is less so for 25-month loans from AART (2019) at the 5% significant level. In a setting where buy and sell decisions must be made quickly, this change could be enough to give investors pause.

A final advantage of modeling X with a more traditional parametric distribution is its stability in long-term analysis. For example, while choosing between two ABS bonds, like AART (2017) and AART (2019) is of obvious interest to investors, a larger decision is whether to invest in an ABS bond versus other asset classes, like corporate or treasury bonds. Our methods may be used to specify a parametric distribution for X, and this choice can be employed as a stable, long-term engine for ABS cash flow modeling. Contrast this to Lautier, Pozdnyakov and Yan (2023a), which can be prone to large fluctuations in its estimate of X between samples (see Figure 2). This is the downside of a large number of parameters in long-term trend analysis, though its upside is clear when its flexibility offers a better data fit to a specific data set (e.g., 50-month loans in Table 4).

6. Discussion. Securitization is a marvel of financial engineering brought about by the economic necessities of financing consumer automobile purchases (i.e., Figure 1). An important participant in the securitization process is the ABS investor, who currently invests over \$170B annually into newly issued U.S. auto asset-backed securities (SIFMA, 2024). This large scale motivates the search for improvements in financial valuation techniques for auto ABS, which will require a stochastic cash flow model. Because the ABS cash flow is a function of the total monthly cash flow of its individual loans (see Table 1), ABS cash flow models can be constructed from specifying a time-to-termination loan-level distribution. It is thus the problem of estimating this lifetime distribution that is our focus.

Historically, any estimation of a loan lifetime distribution was challenged by a lack of available loan-level data (e.g., AART, 2010). Recently, the SEC adopted significant revisions to regulations surrounding auto ABS (Securities and Exchange Commission, 2014), which took effect beginning 2017 (e.g., AART, 2017). Investors now have free access to pertinent loan-level and payment performance data on a monthly basis (Securities and Exchange Commission, 2016). This newly available data further motivates an applied statistical study.

Care must be taken in analyzing this auto ABS data, however, because its data generating process is subject to the incomplete data challenge of left-truncation. Though many classical models exist to model continuous left-truncated data (e.g., Woodroofe, 1985; Tsai, Jewell and Wang, 1987; Wang, 1989), the ABS setting also requires assuming discrete-time over a known, finite support. A series of recent results identifies this combination as largely unexplored and offers methods and applications that do not assume a traditional parametric form of the lifetime distribution (Lautier, Pozdnyakov and Yan, 2023a,b, 2024a,b). Despite these results, the problem of assuming a more traditional parametric form for the lifetime distribution remains open. As summarized in Section 5.3, a solution would provide ABS investors with an alternative asset-level stochastic engine for cash flow modeling, a simplified tool to facilitate the rapid comparison of ABS bonds (valuable in the fast pace of active investment management), and a more stable sample-to-sample approach than Lautier, Pozdnyakov and Yan (2023a) (valuable for long-term trend investment allocation analysis).

Our results are as follows. Under random left-truncation (§3.2), we first reduce a complex multidimensional constrained optimization parameter estimation problem into a single parameter optimization problem (Theorem 3.1). When the parameter space is large – often the case for modeling consumer loan lifetimes from ABS, brute force numeric optimization methods can fail (see Section 4). Conversely, our results can recover the true parameter values accurately and quickly for the distributions we consider. Theorem 3.1 is a general statement for a lifetime random variable X that depends on a single parameter. We then generalize this to a vector parameter for X (Corollary 3.1.1). The form of these estimators takes an asymptotically consistent M-estimator (van der Vaart, 1998, §5.3, pg. 51). Therefore, under the standard regularity conditions (e.g., van der Vaart, 1998, §5.3, pg. 51; Mukhopadhyay, 2000, §12.2, pg. 539), attractive asymptotic properties follow and are detailed in the Supplemental Material. As a special case, we find a closed-form MLE when X follows a PL geometric distribution (13) (Theorem 3.3). In Section 3.3, we generalize all of the left-truncation results to also accommodate random right-censoring. This requires first deriving an updated likelihood equation. Throughout, the form of the left-truncation random variable, Y, is unspecified. Complete proofs are in the Supplemental Material, Section B. All theoretical results are verified numerically and via simulation (see Section 4 and the Supplemental Material).

In Section 5, we apply our methods to data from the ABS bonds AART (2017, 2019). We begin by specifying a LRT that can be used to assess model selection against the flexible Lautier, Pozdnyakov and Yan (2023a). This, along with a visual analysis (e.g., Figure 2), can be used to evaluate model specification questions for various choices of f. We found that for 25-month loans from AART (2017), (13), with its closed-form MLE solution, was viable. For loan terms where this was not the case, we evaluated (25), of which (13) is a special case. Its vector parameter case can introduce concerns of uniqueness and/or indentifiability, and it also adds complexity to the optimization problem of Corollaries 3.1.1 and 3.4.1. Through our detailed applied study, we hope to illustrate these potential challenges and provide a road map of ways they can be overcome. For 25-month loans from AART (2019), the added flexibility of (25) is preferable to (13). For other loan terms where neither was deemed to be preferable to Lautier, Pozdnyakov and Yan (2023a) (e.g., 50-month auto loans from AART (2017)), the process of Section 5 can be repeated until a suitable choice is found. One suggestion is a spliced PL geometric distribution, but we at present leave this problem open to further study.

We close with suggestions for further research. First, a *competing risks* extension can expand applicability of these methods to mortgage-backed securities, where separating defaults from prepayments is a more vital component of ABS cash flow modeling than for AART (2017, 2019). Second, a large portion of auto ABS are on leased assets (e.g., Lautier, Pozdnyakov and Yan, 2023b), of which our methods are directly applicable and thus suggests an applied study. Third, linking the parameter(s) of X to a set of covariates in the form of a GLM can support economic modeling, a problem that remains open. Finally, we postulate our methods will be applicable to other fields where incomplete, discrete time-to-event data frequently occur, such as healthcare, finance, engineering, telecommunication, and insurance.

SUPPLEMENTARY MATERIAL

The supplemental material provides proofs of all major results, additional details related to Sections 3, 4, and 5, and other material referenced herein. All data and replication code is publicly available at https://github.com/jackson-lautier/consumer-auto-abs-parametric.

REFERENCES

AART (2010). Ally Auto Receivables Trust Prospectus No. 2010-1, Ally Auto Assets LLC.

AART (2017). Ally Auto Receivables Trust Prospectus No. 2017-3, Ally Auto Assets LLC.

AART (2019). Ally Auto Receivables Trust Prospectus No. 2019-3, Ally Auto Assets LLC.

AGARWAL, N., MARION, S. and WU, J. (2024). Moody's Global Approach to Rating Auto Loan and Lease-Backed ABS Rating Methodology No. 426427, Moody's Ratings.

ASGHARIAN, M., M'LAN, C. E. and WOLFSON, D. B. (2002). Length-Biased Sampling with Right Censoring. *Journal of the American Statistical Association* **97** 201–209. https://doi.org/10.1198/016214502753479347

HUANG, Y. and WANG, M.-C. (1995). Estimating the Occurrence Rate for Prevalent Survival Data in Competing Risks Models. *Journal of the American Statistical Association* **90** 1406–1415.

KATCHER, B., LI, G., MEZZA, A. and RAMOS, S. (2024). One Month Longer, One Month Later? Prepayments in the Auto Loan Market Finance and Economics Discussion Series (FEDS) No. 2024.056, Federal Reserve Board, Washington, D.C.

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

LAUTIER, J. P., POZDNYAKOV, V. and YAN, J. (2023a). Pricing Time-to-Event Contingent Cash Flows: A Discrete-Time Survival Analysis Approach. *Insurance: Mathematics and Economics* **110** 53–71. https://doi.org/10.1016/j.insmatheco.2023.02.003

LAUTIER, J. P., POZDNYAKOV, V. and YAN, J. (2023b). Estimating a Discrete Distribution Subject to Random Left-Truncation with an Application to Structured Finance. *Econometrics and Statistics*. Forthcoming. https://doi.org/10.1016/j.ecosta.2023.05.005

LAUTIER, J. P., POZDNYAKOV, V. and YAN, J. (2024a). On the Maximum Likelihood Estimation of a Discrete, Finite Support Distribution under Left-Truncation and Competing Risks. *Statistics & Probability Letters* **207** 109973. https://doi.org/10.1016/j.spl.2023.109973

LAUTIER, J. P., POZDNYAKOV, V. and YAN, J. (2024b). On the Convergence of Credit Risk in Current Consumer Automobile Loans. *Journal of the Royal Statistical Society Series A: Statistics in Society*. Forthcoming. https://doi.org/10.1093/jrsssa/qnae137

LEHMANN, E. L. and CASELLA, G. (1998). *Theory of Point Estimation, 2nd Edition*. Springer.

LEHMANN, E. L. and ROMANO, J. P. (2006). *Testing Statistical Hypotheses*. Springer New York.

LUENBERGER, D. G. and YE, Y. (2021). *Linear and Nonlinear Programming, Fifth Edition*. Springer.

MISHKIN, F. S. (2011). Over the Cliff: From the Subprime to the Global Financial Crisis. *Journal of Economic Perspectives* **25** 49-–70. https://doi.org/10.1257/jep.25.1.49

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

MURTAGH, B. A. and SAUNDERS, M. A. (1978). Large-Scale Linearly Constrained Optimization. *Mathematical Programming* **14** 41–72.

UÑA-ÁLVAREZ, D. (2004). Nonparametric Estimation Under Length-Biased Sampling and Type I Censoring: A Moment Based Approach. *Annals of the Institute of Statistical Mathematics* **56** 667–681. https://doi.org/10.1007/BF02506482

NAKAGAWA, T. and OSAKI, S. (1975). The Discrete Weibull Distribution. *IEEE Transactions on Reliability* **R-24** 300–301. https://doi.org/10.1109/TR.1975.5214915

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.

SIFMA (2024). US ABS securities: Issuance, trading volume, outstanding. https://www.sifma.org/resources/research/us-asset-backed-securities-statistics/. Online; accessed 10 December 2024.

R CORE TEAM (2023). R: A Language and Environment for Statistical Computing R Foundation for Statistical Computing, Vienna, Austria.

TSAI, W.-Y., JEWELL, N. P. and WANG, M.-C. (1987). A Note on the Product-Limit Estimator under Right Censoring and Left Truncation. *Biometrika* **74** 883–886. https://doi.org/10.1093/biomet/74.4.883

VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press.

WANG, M.-C. (1989). A Semiparametric Model for Randomly Truncated Data. *Journal of the American Statistical Association* **84** 742–748.

WOODROOFE, M. (1985). Estimating a Distribution Function with Truncated Data. *The Annals of Statistics* **13** 163–177. https://doi.org/10.1214/aos/1176346584