Nonparametric Estimation of Splicing Points in Actuarial Loss Distributions via Data Transformation

Benedikt Funke* Masayuki Hirukawa†
TH Köln Ryukoku University

23rd May 2025

Abstract

Identification of splicing points or thresholds in actuarial loss distributions is crucial for risk management, risk-adequate pricing of insurance products, and solvency assessment. Estimating these points, particularly in the right tail where large claims occur, poses challenges due to data sparsity and potentially small discontinuities. This paper introduces a novel nonparametric method for estimating such splicing points. We propose transforming the original non-negative loss data onto the unit interval $[0, 1]$ to enhance the detection of jumps in the underlying density function. The location of the splicing point is then estimated in the transformed scale using the asymmetric beta kernel, which naturally handles boundary issues. We demonstrate that our estimator is strongly consistent and asymptotically normal with a faster convergence rate than $\sqrt{n}$, where n is the sample size. Appealing finite-sample properties and practical relevance of our approach are illustrated through Monte Carlo simulations and applications to real-world non-life insurance loss datasets.

Keywords: beta kernel; cross validation; data transformation; actuarial loss distributions; splicing point.

JEL Classification Codes: C13; C14; G22.

MSC 2010 Codes: 62G07; 62G20; 62P05; 91B30.

*Institute for Insurance Studies, TH Köln - University of Applied Sciences, Gustav-Heinemann-Ufer 54, 50968 Köln, Germany; e-mail: benedikt.funke@th-koeln.de.

†Faculty of Economics, Ryukoku University, 67 Tsukamoto-cho, Fukakusa, Fushimi-ku, Kyoto 612-8577, Japan; e-mail: hirukawa@econ.ryukoku.ac.jp.

1 Introduction

The precise modelling of loss distributions is of fundamental importance in actuarial science, significantly influencing decisions in areas such as solvency assessment, enterprise risk management, and the pricing of insurance products. For solvency assessment, particularly under regulatory frameworks like Solvency II, the accurate representation of the loss distribution, especially its tail, is paramount for developing internal models and determining economic capital. Effective risk management relies on this understanding for setting appropriate risk appetite, establishing exposure limits, and designing optimal reinsurance strategies. A particular challenge is the adequate representation of the distribution's tail, where rare but potentially very high losses like large claims or extreme events occur. These large claims often follow a different dynamic from small or medium-sized attritional claims and are critical for the calculation of risk capital as well as for modelling of pricing risk. This is especially decisive in excess-of-loss reinsurance contracts where the correct determination of layers and attachment points is crucial for both cedent and reinsurer, directly impacting profitability and risk transfer effectiveness.

In actuarial practice, it has become evident that single, simple distribution models are often unable to satisfactorily represent the entire range of insurance losses, from frequent small claims to catastrophic large events. Such inadequacy stems from the fact that underlying risk drivers for attritional losses often differ from those for large or extreme losses. For this reason, actuaries frequently resort to splicing or composite models (see, e.g., Klugman et al., 2019). In this approach, different distribution functions are used for different segments of the loss data, typically one for the bulk of the distribution and another for the tail, which contains the large claims (e.g., Cooray and Ananda, 2005; Scollnik and Sun, 2012; Reynkens et al., 2017). A central and difficult problem in this procedure is to develop an objective and data-driven method of detecting the point at which one distribution model transitions to another. An inaccurate choice of the splicing point or threshold leads to critical model misspecification, with profound financial and strategic consequences for an insurer. These impacts extend beyond reserving adequacy, particularly for incurred but not reported (IBNR) claims in long-tailed lines of business, and the pricing of complex reinsurance structures like stop-loss or aggregate covers, crucially affecting solvency capital determination (e.g., under Solvency II), the strategic design of reinsurance programs, and overall enterprise risk management and underwriting strategy.

The literature on determining thresholds in loss distributions is diverse. Traditionally, heuristic approaches or graphical diagnostic methods were often employed. Heuristic methods include fixing a specific quantile (DuMouchel, 1983) or using formulas depending on the sample size (Loretan and Phillips, 1994). Graphical methods such as the Hill plot and its variants (Resnick, 1997; Kratz and Resnick, 1996) or the mean excess plot (Davison and Smith, 1990) are widely used due to their intuitive nature. Comprehensive overviews of these and other methods are provided, for example, by Scarrott and MacDonald (2012) and Reiss and Thomas (2007). However, these approaches often allow for considerable practitioner discretion, which can be a limitation in contexts requiring auditable and objective methodologies, such as regulatory reporting or internal model validation.

More recent research has increasingly focused on automated and more objective procedures. Many of these methods are based on fitting the generalized Pareto distribution (GPD) to losses exceeding a certain threshold in the context of approaches based on extreme value theory (EVT). Examples include the minimum Kolmogorov-Smirnov distance procedure (Clauset et al., 2009; Drees et al., 2020), sequential goodness-of-fit tests (e.g., Bader et al., 2018; Northrop and Coleman, 2014; Wadsworth, 2016), and criteria based on quantile discrepancies and automated eye-balling methods (Danielsson et al., 2019). Mixture distribution models with the threshold estimated as a part of model parameters have also been proposed (MacDonald et al., 2011; Wadsworth and Tawn, 2012). Although these automated procedures reduce subjectivity, they often depend on the specific assumption of a particular tail model, such as the GPD. Additionally, the stability of the chosen threshold when new data becomes available can present a challenge. An alternative methodological approach to identify structural breaks in distributions is offered by the statistical literature on estimating jump points or change points (e.g., Chu and Cheng, 1996; Couallier, 1999; Huh, 2002). Building upon such techniques, Funke and Hirukawa (2025) propose a nonparametric kernel-based approach for skewed cost distributions, which serves as a related starting point for the methodology developed herein.

This paper addresses the challenge of estimating splicing points in actuarial loss distributions by introducing a novel nonparametric approach. As in Funke and Hirukawa (2025), we interpret the splicing point as a jump location in the underlying distribution. In this context, our proposal may be viewed as a natural extension of theirs. However, there is a certain weakness in the approach by Funke and Hirukawa (2025). When the splicing point is located in the right tail region in which data points are sparse and/or the associated jump size is small, their approach has difficulty in identifying the splicing point in the original scale. To estimate the location of a splicing point precisely under such unfavorable environments, we propose transforming the original, non-negative loss data onto the unit interval $[0, 1]$. The data transformation has the advantage of magnifying the relative jump size at the splicing point and compressing the distances between adjacent observations in the relevant area, thereby facilitating the detection of the jump location.

Following Funke and Hirukawa (2025), we take the absolute difference of two nonparametric kernel density estimates – now in the transformed scale – as the diagnostic function. The splicing point estimator can be obtained by back-transforming the maximizer of the diagnostic function into the original scale. To compute density estimates in the transformed scale, we employ the asymmetric beta kernel proposed by Chen (1999). The kernel is defined as the probability density function (pdf) of the beta distribution $Beta\{y/b + 1, (1 - y)/b + 1\}$ and takes the form

$$K_{B(y,b)}(u) = \frac{u^{y/b}(1-u)^{(1-y)/b}}{B\{y/b+1,(1-y)/b+1\}} \mathbf{1}\{u \in [0,1]\} \tag{1}$$

for a data point $u \in [0, 1]$, a design point $y \in [0, 1]$, and a smoothing parameter $b > 0$, where $B(p, q) = \int_0^1 t^{p-1}(1 - t)^{q-1}dt$ for $p, q > 0$ is the beta function, and $\mathbf{1}\{\cdot\}$ denotes an indicator function. This kernel is particularly suitable as it is adapted to the boundaries of the unit interval, thus avoiding boundary bias problems that can occur with standard kernels. Furthermore, like the gamma kernel by Chen (2000), the beta kernel possesses adaptive smoothing properties, offering flexibility in capturing various distributional shapes. The nonparametric nature of our approach offers a key advantage in situations of high model uncertainty or where distributional assumptions for the tail are difficult to verify.

It is demonstrated that our splicing point estimator is strongly consistent and asymptotically normal. Remarkably, the estimator is super-consistent in the sense that its convergence rate exceeds $\sqrt{n}$, where n is the sample size. Super-consistency is beneficial in many actuarial applications, for instance, in two-step estimation procedures for loss distributions or in post-estimation goodness-of-fit testing, as it does not harm the convergence rates of subsequent analyses.

An appealing finite-sample performance of our splicing point estimator is confirmed through Monte Carlo simulations. Its practical relevance to the analysis on actuarial loss distributions is also illustrated in several applications to real non-life insurance datasets. These applications show that our method can be a valuable and flexible addition to the existing toolkit for threshold detection in actuarial practice.

It can be particularly useful in model validation processes or as a complementary tool alongside existing methods, especially when it comes to identifying the point at which the dynamics of large claims change significantly.

The remainder of this paper is organized as follows. Section 2 introduces the estimation of the splicing point based on data transformation and the beta kernel. In Section 3, convergence properties of the proposed estimator, namely, strong consistency and asymptotic normality, are presented. Section 4 conducts Monte Carlo simulations to compare the finite-sample behavior of our splicing point estimator with those of several competing methods. In Section 5, the proposed estimation approach is applied to real-world actuarial datasets. Section 6 concludes. Proofs of theorems and propositions are provided in the Appendix. Proofs of lemmata are deferred to the Supplementary Material, which is available on the second author's webpage.

This paper adopts the following notational conventions: '$a_n \sim b_n$' means that a_n/b_n converges to 1; '$a_n = o(b_n)$' signifies that a_n/b_n converges to 0; '$a_n = O(b_n)$' means that a_n/b_n is bounded; and we say that '$a_n \asymp b_n$' if there exist constants $0 < c_1 < c_2 < \infty$ so that $c_1 a_n \le b_n \le c_2 a_n$. For a function $h(x)$ and a point c, $h(c^-) = \lim_{x\uparrow c} h(x)$, $h(c^+) = \lim_{x\downarrow c} h(x)$ and $h^{(m)}(x) = d^m h(x)/dx^m$ denote the left and right limits, and the mth-order derivative, respectively. The abbreviation '$a.s.$' stands for "almost surely". Finally, the expression '$X \overset{d}{=} Y$' reads "A random variable X obeys the distribution Y."

2 Splicing Point Estimation Using Transformed Data

We tackle the problem of estimating the splicing point under the same setup as in Funke and Hirukawa (2025). It is suspected that $f_X(x)$, the pdf of an actuarial loss variable $X \in \mathbb{R}_+$, is discontinuous at t_0 on a prespecified closed interval $I_0 := [\underline{t}, \overline{t}]$ with $0 < \underline{t} < \overline{t} < \infty$. Throughout it is assumed that the interval I_0 is located in the right tail part of the underlying loss distribution. Prior knowledge on the interval is not at all unrealistic, because quite often practitioners have a rough idea about the location of the threshold through, for example, preliminary threshold estimates from EVT methods, historical experiences, specific policy limits, or empirical quantiles relevant for determining attachment points or layer boundaries. It is also assumed that the pdf $f_X(x)$ for $x \in I_0$ can be modelled as

$$f_X(x) = g_X(x) + d_0 \mathbf{1}\{x < t_0\}, \tag{2}$$

where $g_X(x)$ is a sufficiently smooth function, $t_0 \in I_0$ is the discontinuous, splicing point. Moreover, the jump size

$$d_0 := f_X\left(t_0^-\right) - f_X\left(t_0^+\right) \tag{3}$$

is assumed to satisfy $|d_0| \in (0, \infty)$. This setup allows for a potential discontinuity at the threshold where the behavior of large claims might differ significantly from the bulk of the distribution.

Given n *i.i.d.* observations $\{X_i\}_{i=1}^n$ representing individual insurance losses, we aim to estimate the splicing point t_0 nonparametrically. Suppose that the jump size $|d_0|$ is so small that it is difficult to estimate t_0 precisely in the original scale $\mathbb{R}_+$. This situation can arise in practice, particularly with sparse data in the tail or when the change in density, while potentially significant for risk assessment, is not abrupt. In this environment, it is reasonable to consider a contraction of the data scale via some transformation. Let $T : \mathbb{R}_+ \mapsto [0,1]$ be a known monotone transformation function. An obvious example of T is the cumulative distribution function (cdf) of a non-negative random variable, potentially chosen to reflect typical shapes of loss distributions. Also let $\{Y_i\}_{i=1}^n := \{T(X_i)\}_{i=1}^n \in [0,1]$ be the transformed observations. Accordingly, we reformulate our problem as one in the transformed scale. To be more concrete, assume that $f_Y(y)$, the pdf of the transformed loss variable $Y \in [0,1]$, admits the local structure

$$f_Y(y) = g_Y(y) + d_T \mathbf{1}\{y < t_T\} \tag{4}$$

on a prespecified interval $I_T := [\underline{t}_T, \bar{t}_T] := \left[T(\underline{t}), T(\bar{t})\right] \subset [0,1]$, where $g_Y(y)$ is a sufficiently smooth function, $t_T := T(t_0) \in I_T$ is the splicing point, and

$$d_T := f_Y\left(t_T^-\right) - f_Y\left(t_T^+\right) \tag{5}$$

is the jump size.

The correspondence between (2)-(3) and (4)-(5) can be immediately found. Assume that T has a continuous and uniformly bounded first-order derivative. Then, by (4) and a straightforward calculation,

$$\begin{aligned} f_X(x) &= g_X(x) + d_0 \mathbf{1}\{x < t_0\} \\ &= g_Y\{T(x)\} T^{(1)}(x) + d_T T^{(1)}(x) \mathbf{1}\{x < t_0\}. \end{aligned} \tag{6}$$

Because of continuity of $T(\cdot)$ and $T^{(1)}(\cdot)$ at t_0, we also have

$$d_0 = d_T T^{(1)}(t_0). \tag{7}$$

As will be documented in Assumption 3(ii) shortly, if $0 < T^{(1)} < 1$ on $[0,1]$ is the case, then $|d_0| < |d_T|$ holds and thus it becomes easier to estimate the splicing point in the transformed scale.

The above argument ensures that we can recover the splicing point in the original scale by concentrating only on estimating the one in the transformed scale. In what follows, we develop a nonparametric method of estimating t_T using transformed observations $\{Y_i\}_{i=1}^n \in [0,1]$. On the one hand, t_T is likely to be estimated fairly precisely if it is located in the middle part of $[0,1]$. On the other hand, even if t_T is moved away from two boundaries 0 and 1, the region above t_T is originally the tail part and thus the number of observations falling into this region is small. Therefore, splitting the entire sample into two sub-samples near t_T results in imbalance in sample sizes of two sub-samples and an imprecise density estimate from the right sub-sample. Inevitably, it is highly desirable to take measures to estimate $f_Y\left(t_T^{\pm}\right)$ using the entire sample. Then, following Funke and Hirukawa (2025), we propose to use 'shifted' beta kernels $K_{B(y,b;\pm\Delta)}(\cdot)$, which are defined as pdfs of beta distributions $Beta\left\{(y \pm \Delta)/b + 1, (1 - y \mp \Delta)/b + 1\right\}$, i.e.,

$$K_{B(y,b;\pm\Delta)}(u) = \frac{u^{(y\pm\Delta)/b}(1-u)^{(1-y\mp\Delta)/b}}{B\left\{(y \pm \Delta)/b + 1, (1 - y \mp \Delta)/b + 1\right\}}\mathbf{1}\left\{u \in [0,1]\right\},$$

where $b\,(= b_n > 0)$ is the smoothing parameter, and $\Delta\,(= \Delta_n > 0)$ plays the role of a shift parameter. Notice that each parameter is common across two kernels and shrinks toward zero at a certain rate; see Assumption 4 below for more details on their rate requirements. Obviously, $K_{B(y,b;\pm\Delta)}(\cdot)$ collapse to Chen's (1999) original beta kernel (1) when $\Delta = 0$. The kernels can be interpreted as those designed to smooth the data off the target design point y by a margin of Δ. In addition, they put the maximum weight at sightly left or right of y because they have their modes at $y \pm \Delta$.

We exploit as the source of our splicing point estimator the difference between two density estimates that is created by the shift parameter Δ. Let the shifted density estimators be

$$\hat{f}_Y^{\pm}(y) := \frac{1}{n}\sum_{i=1}^{n} K_{B(y,b;\pm\Delta)}(Y_i).$$

Also define

$$\hat{J}(y) := \hat{f}_Y^{-}(y) - \hat{f}_Y^{+}(y).$$

We take $\left|\hat{J}(y)\right|$ as the diagnostic function for our splicing point estimation in the transformed scale. Then, $\hat{t}_T$, the estimator of the splicing point t_T, is defined as the maximizer of $\left|\hat{J}(y)\right|$ on $y \in I_T$, i.e.,

$$\hat{t}_T := \arg\max_{y \in I_T} \left|\hat{J}(y)\right|.$$

Let $\hat{t}_B \left(= \hat{t}_B(T)\right)$ be the estimator of t_0, the splicing point in the original scale, where the subscript "B" signifies dependence of the estimator on the beta kernel. Because $t_T = T(t_0)$, it is natural to define the estimator for t_0 as $T^{-1}\left(\hat{t}_T\right)$, i.e.,

$$\hat{t}_B := T^{-1}\left(\hat{t}_T\right).$$

3 Convergence Properties of the Splicing Point Estimator

This section presents large-sample properties of the splicing point estimator $\hat{t}_B$. Our particular focus is on its strong consistency and asymptotic normality. In addition, existence of the unique maximum in the probability limit of $\hat{J}(y)$ on $y \in I_T$ is demonstrated. This is a key condition for consistency of $\hat{t}_B$.

3.1 Regularity Conditions

Convergence results rely on the fact that $\hat{J}(y)$ can be approximated by the difference between two incomplete beta function ratios; see equations (12)-(14) below for more details. To deliver the results, we impose the following regularity conditions.

Assumption 1. $\{X_i\}_{i=1}^n \in \mathbb{R}_+$ are *i.i.d.* random variables.

Assumption 2.

(i) The pdf $f_Y(y)$ is uniformly bounded on $y \in [0,1]$.

(ii) The local structure (4) holds, $g_Y^{(2)}(y)$ is uniformly bounded on $y \in [0,1]$, and $g_Y^{(3)}(y)$ is Lipschitz continuous and bounded on $y \in I_T$.

Assumption 3. The transformation $T : \mathbb{R}_+ \mapsto [0,1]$ satisfies the followings.

(i) $T(x)$ is injective with $T(0) = 0$ and $T(t_M) = 1/2$ for $t_M := \left(\underline{t} + \bar{t}\right)/2$.

(ii) $T^{(1)}(x)$ is Lipschitz continuous on $x \in \mathbb{R}_+$, and there are constants $0 < \underline{T}^{(1)} \leq 1/2 \leq \overline{T}^{(1)} < 1$ so that $\underline{T}^{(1)} \leq T^{(1)}(x) \leq \overline{T}^{(1)}$ holds for all $x \in I_0$.

Assumption 4. Tuning parameters b and Δ satisfy $b, \Delta \to 0$,

$$\frac{b^{3/4}}{\Delta} + \frac{\Delta}{b^{1/2+\delta_1}} + \frac{b^{1/2-4\delta_1}}{n^{1-\delta_2}\Delta^2} \to 0 \tag{8}$$

for some arbitrarily small $\delta_1, \delta_2 > 0$, and

$$\frac{\ln n}{nb^{3/2-\kappa}} = O(1) \tag{9}$$

for some $\kappa \in [0, 1)$, as $n \to \infty$.

Assumptions 1, 2 and 4 are standard for uniform approximations to asymmetric kernel estimators with support on $[0, 1]$. Similar conditions can be found, for example, in Hirukawa et al. (2022). Assumption 1 ensures that $\{Y_i\}_{i=1}^n \in [0, 1]$ are also *i.i.d.* random variables. Assumption 2(i) is a key condition for uniform approximation of $\hat{f}_Y^{\pm}(y)$ on $y \in I_T$ in Proposition 1. It also follows from Assumption 2(ii) that $f_Y^{(1)}(t_T^-) = f_Y^{(1)}(t_T^+)$. This type of condition has often been imposed in simulation studies on change point detection (e.g., Chu and Cheng, 1996). Two boundedness conditions on derivatives of the smoothed component $g_Y(\cdot)$ also serve as important ingredients for approximations to $E\left\{\hat{J}^{(p)}(y)\right\}$ on $y \in I_T$ for $p = 0, 1, 2$. These approximations in turn play an important role in establishing asymptotic normality of $\hat{t}_B$.

Assumption 3(i) means that T is a one-to-one mapping, and thus the inverse mapping T^{-1} exists. While $T(0) = 0$ is a normalization, $T(t_M) = 1/2$ means that T maps the midpoint of I_0 to that of $[0, 1]$ (and as a result, the entire part of the transformed interval I_T can be located roughly in the middle part of $[0, 1]$). The fact that t_T is located away from two boundaries 0 and 1 enables us to estimate t_T more easily, as will be explained after Proposition 2. Positivity of $T^{(1)}(x)$ in Assumption 3(ii) implies monotonicity of T. It follows from $T^{(1)}(x) < 1$ for $x \in I_0$ and (7) that $|d_0| < |d_T|$ holds. This ensures that estimating t_T in the transformed scale is easier than estimating t_0 in the original scale. There are many transformations that satisfy Assumption 3. Table 1 lists examples of such transformations.

TABLE 1 ABOUT HERE

Assumption 4 controls the shrinkage rates of tuning parameters b and Δ. The condition (8) draws the following important conclusions: (i) $b = o(\Delta)$; (ii) $b^{1/2} =$

$o(\Delta^2/b)$; and (iii) $\Delta = o\left(\Delta^3/b^{3/2}\right)$. These are frequently used to controls remainder terms in the asymptotic expansions. It also follows from $b = o(\Delta)$ and $\Delta = o\left(b^{1/2}\right)$ that although the shift parameter Δ should shrink to zero more slowly than the smoothing parameter b, the convergence rate of Δ must not be too slow (or must be faster than $b^{1/2}$, to be more precise). The condition (8) also implies that

$$\frac{\ln n}{nb^{1/2}} = \left(\frac{b^{1/2-4\delta_1}}{n^{1-\delta_2}\Delta^2}\right)\left(\frac{\ln n}{n^{\delta_2}}\right)\left(\frac{\Delta}{b^{1/2+\delta_1}}\right)^2 b^{6\delta_1} \to 0.$$

This result serves as a prerequisite for Proposition 1, as will be revealed shortly. The other condition (9) is an additional technical requirement for strong uniform consistency of $\hat{t}_B$.

3.2 Approximation to the Diagnostic Function

Below asymptotic properties of the splicing point estimator $\hat{t}_B$ are explored. Our analysis starts from a uniform approximation to $\hat{f}_Y^{\pm}(y)$ on I_T, which is documented in the next proposition. To save space, we adopt the shorthand notation $K_y^{\pm}(u) = K_{B(y,b;\pm\Delta)}(u)$ whenever no confusion may arise.

Proposition 1. *If Assumptions 1-4 hold, then*

$$\sup_{y\in I_T}\left|E\left\{\hat{f}_Y^{\pm}(y)\right\} - \left\{g_Y(y) \pm g_Y^{(1)}(y)\Delta + d_T\int_0^{t_T} K_y^{\pm}(u)\,du\right\}\right| = O(b), \tag{10}$$

and

$$\sup_{y\in I_T}\left|\hat{f}_Y^{\pm}(y) - E\left\{\hat{f}_Y^{\pm}(y)\right\}\right| = O\left(\sqrt{\frac{\ln n}{nb^{1/2}}}\right) \ a.s., \tag{11}$$

as $n \to \infty$.

A direct outcome from Proposition 1 is that

$$\sup_{y\in I_T}\left|\hat{J}(y) - E\left\{\hat{J}(y)\right\}\right| = O\left(\sqrt{\frac{\ln n}{nb^{1/2}}}\right) \ a.s.$$

Because $\left|\left|\hat{J}(y)\right| - \left|E\left\{\hat{J}(y)\right\}\right|\right| \le \left|\hat{J}(y) - E\left\{\hat{J}(y)\right\}\right|$, it holds that

$$\left|\hat{J}(y)\right| = \left|E\left\{\hat{J}(y)\right\}\right| + O\left(\sqrt{\frac{\ln n}{nb^{1/2}}}\right) \ a.s. \tag{12}$$

uniformly on I_T. In short, $\left|E\left\{\hat{J}(y)\right\}\right|$ constitutes the dominant term in the diagnostic function $\left|\hat{J}(y)\right|$, or the effect of the location y on $\left|\hat{J}(y)\right|$ is governed by the value of $\left|E\left\{\hat{J}(y)\right\}\right|$ in a first-order asymptotic sense. This result also plays a key role in the proof of Theorem 1 in the Appendix.

It follows from (10) that $\left|E\left\{\hat{J}(y)\right\}\right|$ can be further approximated by

$$\left|E\left\{\hat{J}(y)\right\}\right| := |d_T|\,J(y) + O(\Delta) \tag{13}$$

uniformly on I_T, where

$$\begin{aligned} J(y) &= \left|\int_0^{t_T} K_y^-(u)\,du - \int_0^{t_T} K_y^+(u)\,du\right| \\ &= \int_0^{t_T} K_y^-(u)\,du - \int_0^{t_T} K_y^+(u)\,du \\ &=: B\left(\frac{y-\Delta}{b}+1, \frac{1-y+\Delta}{b}+1; t_T\right) \\ &\quad - B\left(\frac{y+\Delta}{b}+1, \frac{1-y-\Delta}{b}+1; t_T\right), \end{aligned} \tag{14}$$

and

$$B(p,q;r) = \frac{1}{B(p,q)}\int_0^r y^{p-1}(1-y)^{q-1}\,dy$$

for $p, q > 0$ and $r \in [0,1]$ is the incomplete beta function ratio.

It is natural to verify whether $J(y)$ on $y \in I_T$ indeed has a unique maximum at t_T (or within a shrinking neighborhood of t_T even if it is not maximized exactly at this point). In reality, however, it is quite cumbersome to look into the local property of $J(y)$ analytically. It is well known that as $p, q \to \infty$, the pdf of $Beta(p,q)$ can be approximated by a normal pdf. Lemma A1 in the Appendix formally offers such an approximation, which is a refinement of Lemma A.1 of Moscovich, Nadler and Spiegelman (2016). Based on Lemma A1, the next proposition refers to properties of the approximation and the maximizer of the approximated function.

Proposition 2. *If Assumption 4 holds, then the followings hold true.*

(i) *Define*

$$\begin{aligned} Q(y) &:= \left[\frac{(1+2b)\sqrt{1+3b}\,\{(1-2y)\,t_T + (y+b)\}}{\{(y+b)(1-y+b)\}^{3/2}}\right] \\ &\quad \times \phi\left\{\sqrt{\frac{1+3b}{b\,(y+b)(1-y+b)}}\,(t_T - y + (2t_T-1)\,b)\right\}, \end{aligned}$$

where $\phi(\cdot)$ is the pdf of $N(0,1)$. Then,

$$\sup_{y \in I_T} \left| J(y) - Q(y)\left(\frac{\Delta}{b^{1/2}}\right) \right| = O\left(\frac{\Delta^3}{b^{3/2}}\right),$$

as $n \to \infty$.

(ii) *$Q(y)$ on I_T has a unique maximum at $y = t_T^* = t_T + 2(2t_T - 1)b + O(b^2)$, as $n \to \infty$.*

As a consequence of Propositions 1 and 2, it holds that

$$\left|\hat{J}(y)\right| \sim |d_T| J(y) \sim |d_T| Q(y)\left(\frac{\Delta}{b^{1/2}}\right) \tag{15}$$

on $y \in I_T$. Furthermore, Proposition 2(ii) suggests that $t_T^* = t_T + O(b^2)$ when $t_T = 1/2$. In this case, the maximizer of $Q(y)$ is considerably close to the true splicing point in the transformed scale, so are the maximizers of $J(y)$ and $\left|\hat{J}(y)\right|$. Although the location of the splicing point in the original scale is unknown, we have a high chance to estimate it precisely in the transformed scale as long as it is mapped to (a neighborhood of) 1/2. This is a rationale of mapping the midpoint of I_0 to 1/2 in Assumption 3(i).

3.3 Consistency

The theorem below documents strong consistency of $\hat{t}_B$ for t_0.

Theorem 1. *Let $c_n := b^{1/2+\delta_1}$ for δ_1 defined in Assumption 4. If Assumptions 1-4 hold, then $\left|\hat{t}_B - t_0\right| = O(c_n)$ a.s. as $n \to \infty$.*

It follows from (15) that maximizing the diagnostic function $\left|\hat{J}(y)\right|$ on $y \in I_T$ is a well-defined problem, and the relation is utilized as a fundamental part of the proof of this theorem. As an intermediate product of the proof, we can also obtain $\left|\hat{t}_T - t_T\right| = O_p(c_n) = o_p\left(b^{1/2}\right)$. As will be seen in Lemma A7 in the Appendix, the weak consistency of $\hat{t}_T$ with this rate plays a key role in establishing the asymptotic normality of $\hat{t}_T$ (and thus that of $\hat{t}_B$).

3.4 Asymptotic Normality

The theorem below documents asymptotic normality of $\hat{t}_B$. As in Theorem 1 of Funke and Hirukawa (2025), we derive the limiting distribution indirectly. The indirect derivation comes from the fact that $\hat{t}_T$ solves the first-order condition $\hat{J}^{(1)}\left(\hat{t}_T\right) = 0$.

Then, by a mean-value expansion of the left-hand side around $\hat{t}_T = t_T$ and suitable approximations to the incomplete beta function ratio, digamma and polygamma functions, we can obtain the asymptotic normality of $\hat{t}_T$ as an intermediate result. This is possible because unlike $f_Y(y)$, its estimates $\hat{f}_Y^{\pm}(y)$ are smooth functions even at t_T due to differentiability of shifted beta kernels $K_y^{\pm}(\cdot)$ with respect to y. The asymptotic normality of $\hat{t}_B$ can be reached by back-transforming each part of the intermediate result into the original scale.

Theorem 2. *If Assumptions 1-4 hold, then*

$$\sqrt{\frac{n}{b^{1/2}}}\left[\hat{t}_B - t_0 - \left\{\frac{1-T(t_0)/2}{T^{(1)}(t_0)}\right\} b\left\{1+o_p(1)\right\}\right] \xrightarrow{d} N(0, V_B)$$

as $n \to \infty$, *where*

$$V_B\left(= V_B(T)\right) := \frac{3\sqrt{\pi}\sqrt{T(t_0)\left\{1-T(t_0)\right\}}}{4d_0^2 T^{(1)}(t_0)}\left\{\frac{f_X(t_0^-)+f_X(t_0^+)}{2}\right\}.$$

Remark 1. While it is difficult to obtain asymptotic bias and variance of $\hat{t}_B$ in light of the indirect nature, the asymptotic distribution in Theorem 2 implies the first two moments of $\hat{t}_B$. It can be immediately found that both bias and variance terms are free of the shift parameter Δ, as in Theorem 2 of Funke and Hirukawa (2025). In short, Δ does not affect convergence properties of $\hat{t}_B$ in a first-order asymptotic sense.

In addition, the dominant bias term of $\hat{t}_B$ depends on an unknown quantity t_0. This is a sharp contrast to Theorem 2 of Funke and Hirukawa (2025), in which the leading bias coefficient on $\hat{t}_G$, the splicing point estimator in the original scale based on the gamma kernel, is free of unknowns. Therefore, an elementary bias correction is applicable to $\hat{t}_G$ with no price of additional spread. On the other hand, because of the dependence of the bias of $\hat{t}_B$ on t_0, we do not pursue bias correction for $\hat{t}_B$.

Moreover, V_B, the coefficient of the dominant variance term, suggests that the larger the magnitude of discontinuity $|d_0|$, the easier the estimation of t_0. The part $T(t_0)\left\{1-T(t_0)\right\} = t_T(1-t_T)$ indicates that V_B tends to be smaller as the splicing point in the transformed scale is closer to either boundary of $[0,1]$. However, this interpretation is misleading. As shown in Lemma of Chen (1999), if the design point at which smoothing is made by the beta kernel is located in the vicinity of a boundary, then the convergence rate of the beta density estimator slows down. As a result, the convergence rate of $\hat{t}_B$ is also expected to decelerate.

Remark 2. Theorem 2 also yields an approximation to the mean squared error (AMSE) of $\hat{t}_B$ as

$$AMSE\left(\hat{t}_B\right) = \left\{\frac{1 - T\left(t_0\right)/2}{T^{(1)}\left(t_0\right)}\right\}^2 b^2 + \frac{b^{1/2}}{n} V_B = O\left(b^2 + \frac{b^{1/2}}{n}\right),$$

where $O\left(b^2\right)$ and $O\left(b^{1/2}/n\right)$ terms are leading squared bias and variance of $\hat{t}_B$, respectively. Observe that orders of magnitude in leading squared bias and variance of $\hat{t}_B$ are the same as those of $\hat{t}_G$ in Theorem 2 of Funke and Hirukawa (2025). In particular, the $b^{1/2}/n$ variance convergence rate is faster than the $1/n$ parametric rate, or $\hat{t}_B$ is super-consistent. Furthermore, no bias-variance trade-off arises, because a smaller b makes both squared bias and variance terms smaller.

Remark 3. One might naturally ask: can we estimate the splicing point more precisely in the original scale or in the transformed scale? A comparison of asymptotic variances of $\hat{t}_B$ and $\hat{t}_G$ could help answer this question. It follows from Theorem 2 of Funke and Hirukawa (2025) that

$$Var\left(\hat{t}_G\right) \sim \frac{b^{1/2}}{n} V_G := \frac{b^{1/2}}{n} \frac{3\sqrt{\pi} t_0^{1/2}}{4 d_0^2} \left\{\frac{f_X(t_0^-) + f_X(t_0^+)}{2}\right\}.$$

Notice that the bias-corrected estimator $\tilde{t}_G = \hat{t}_G + b$, which is advocated by Funke and Hirukawa (2025), has the same asymptotic variance. Then, $Var\left(\hat{t}_G\right)/Var\left(\hat{t}_B\right) \sim V_G/V_B = T^{(1)}\left(t_0\right)\sqrt{t_0/\left[T\left(t_0\right)\left\{1 - T\left(t_0\right)\right\}\right]}$. Now suppose that $t_0 \approx t_M$ so that $T\left(t_0\right)\left\{1 - T\left(t_0\right)\right\} \approx 1/4$ by Assumption 3(i). In this case, $Var\left(\hat{t}_G\right)/Var\left(\hat{t}_B\right) \approx 2T^{(1)}\left(t_M\right)\sqrt{t_M}$. Transformations $T_1 - T_4$ in Table 1 yield values of the right-hand side as

$$2T^{(1)}\left(t_M\right)\sqrt{t_M} = \begin{cases} 2/\left(\pi\sqrt{t_M}\right) & \text{for } T_1 \\ \ln 2/\sqrt{t_M} & \text{for } T_2 \\ 1/\left(2\sqrt{t_M}\right) & \text{for } T_3 \\ 3\ln 3/\left(4\sqrt{t_M}\right) & \text{for } T_4 \end{cases}.$$

For each of the four transformations, this quantity is less than unity when $t_M \geq 1$, and thus it appears that $\hat{t}_B$ is less efficient (and thus less appealing) than $\hat{t}_G$ in realistic settings with the splicing point on the right tail.

However, this comparison is based on the assumption that the smoothing parameter value b is common across two splicing point estimators. This is not the case in reality. Once different values of b are allowed for the original and transformed scales, efficiency comparison between $\hat{t}_G$ and $\hat{t}_B$ becomes unclear. In the Monte Carlo study in Section 4, we will make their efficiency comparison in finite samples.

Remark 4. Some readers may wonder how to pick b and Δ for super-consistency of $\hat{t}_B$ under the constraints (8) and (9). For arbitrarily small $\delta_1, \delta_2 > 0$ as given in Assumption 4, put $\Delta \asymp b^{\alpha}$ for some $\alpha \in (1/2 + \delta_1, 3/4)$ and $b \asymp n^{-\beta}$ for some $\beta \in (0, (1-\delta_2) / (2\alpha - 1/2 + 4\delta_1))$. It is straightforward to see that such Δ and b jointly satisfy (8). In addition, when $\alpha \in (1/2, 3/4)$, we have

$$\frac{1-\delta_2}{1+4\delta_1} < \frac{1-\delta_2}{2\alpha - 1/2 + 4\delta_1} < \frac{2(1-\delta_2)}{1+12\delta_1},$$

where the two bounds $(1-\delta_2)/(1+4\delta_1)$ and $2(1-\delta_2)/(1+12\delta_1)$ are slightly below 1 and 2, respectively. Using this, we may draw the following three conclusions on the convergence rate of $\hat{t}_B$:

1. We are always allowed to pick $\beta > 1/2$. Then, $AMSE\left(\hat{t}_B\right) = o\left(n^{-1}\right)$, or $\hat{t}_B$ becomes super-consistent.

2. It is even possible to set $\beta = 2/3$, in particular. This value balances orders of magnitude in the squared bias and variance so that $O\left(b^2\right) = O\left(b^{1/2}/n\right) = O\left(n^{-4/3}\right)$. As a consequence, $AMSE\left(\hat{t}_B\right) = O\left(n^{-4/3}\right)$. It is also clear that the AMSE convergence rate of $\hat{t}_B$ is determined by the exponent β. $AMSE\left(\hat{t}_B\right) = O\left(b^2\right)$ (i.e., the squared bias dominates) for $\beta \le 2/3$, and $AMSE\left(\hat{t}_B\right) = O\left(b^{1/2}/n\right)$ (i.e., the squared bias becomes asymptotically negligible) otherwise. The latter case corresponds to an 'undersmoothing' scenario so that $nb^{3/2} \to 0$ holds. As a consequence, the asymptotic normality statement in Theorem 2 reduces to $\sqrt{n/b^{1/2}}\left(\hat{t}_B - t_0\right) \xrightarrow{d} N\left(0, V_B\right)$.

3. The best possible rate is $AMSE\left(\hat{t}_B\right) = O\left(n^{-2+\varepsilon}\right)$ for an arbitrarily small $\varepsilon > 0$. The rate can be attained by setting α and β slightly above 1/2 and slightly below 2, respectively.

Furthermore, it is not hard to see that for Δ and b defined above, we can always find some $\kappa \in [0, 1)$ satisfying (9). To see this, observe that (9) holds if $nb^{3/2-\kappa} \to \infty$ at a polynomial rate. The rate requirement is attained for case 1 by setting β slightly above 1/2 and $\kappa = 0$. For case 2, $\beta = 2/3$ and any $\kappa \in (0, 1)$ can jointly establish a polynomial divergence of $nb^{3/2-\kappa}$. Finally, for case 3, β slightly below 2 and κ slightly below 1 lead to $nb^{3/2-\kappa} \to \infty$ at a polynomial rate.

Remark 5. Remark 4 indicates that $\hat{t}_B$ becomes super-consistent when suitably implemented. Super-consistency is beneficial in many applications. For example, Reynkens et al. (2017) study a two-step estimation of actuarial loss distributions. In their procedure, the splicing point alone is initially estimated, and all remaining model parameters are estimated in the second step. Our splicing point estimator $\hat{t}_B$ fits well with their procedure, because its convergence rate is faster than the parametric one. Furthermore, Clauset et al. (2009) argue that fitting a power law distribution to the tail part has nothing to do with a plausible match of the distribution with the data, and they recommend a goodness-of-fit test as a post-estimation analysis. Super-consistency of $\hat{t}_B$ does no harm to convergence rates of the test statistic, either.

4 Finite-Sample Performance

4.1 Monte Carlo Design

In the simulation study below, three alternative models are considered as the distribution of the univariate random variable $X \in \mathbb{R}_+$. For each distribution, the common splicing point $t_0 = 4$ is maintained.

In the first model, labelled as Model A, X is drawn from a *log-normal*-like distribution. What differs from a usual log-normal distribution is that a quadratic term is added to the pdf on the interval $[0, t_0) = [0, 4)$. Specifically, the pdf $f(x)$ is

$$f_X(x) = \left\{ \frac{1}{1 + (2/3) D t_0} \right\} \left[\frac{1}{x\sigma\sqrt{2\pi}} \exp\left\{ -\frac{(\ln x - \mu)^2}{2\sigma^2} \right\} + S(x) \right], \ (\mu, \sigma) = \left(\frac{3}{5}, \frac{1}{2} \right),$$

where

$$S(x) := D \left\{ 1 - \left(\frac{x - t_0}{t_0} \right)^2 \right\} \mathbf{1}\{x < t_0\}$$

and

$$d_0 = f_X(t_0^-) - f_X(t_0^+) = \frac{D}{1 + (2/3) D t_0}.$$

Specifically, we take $D = 3/52$ so that the jump size becomes $d_0 = 0.05$. It also follows from $f_X^{(1)}(t_0^-) = f_X^{(1)}(t_0^+)$ that (2) is satisfied in the neighborhood of t_0.

The second and third models, labelled as Models B and C, share common structure. The distribution for each model is spliced exactly at $t_0 = 4$, and the pdf of X takes the general form

$$f_X(x) = f_L(x) \mathbf{1}\{x < t_0\} + (1 - c_L) f_R(x) \mathbf{1}\{x \geq t_0\},$$

where $f_L(x)$ is some density function truncated at t_0, $f_R(x)$ is another density function with support on $[t_0, \infty)$, and $c_L := \int_0^{t_0} f_X(x)\, dx = \int_0^{t_0} f_L(x)\, dx$ ensures unity of the integral of $f_X(x)$ over its entire support $\mathbb{R}_+$. To put it another way, $f_L(x)$ and $f_R(x)$ represent bulk and tail models, respectively. The bulk model $f_L(x)$ is common but the tail model $f_R(x)$ differs across Models B and C. The bulk part is modelled commonly as the *Weibull* distribution with density

$$f_L(x) = \frac{\rho}{\lambda}\left(\frac{x}{\lambda}\right)^{\rho-1} \exp\left\{-\left(\frac{x}{\lambda}\right)^{\rho}\right\},\ (\rho, \lambda) = \left(\frac{9}{4}, \frac{5}{2}\right).$$

The tail part for Model B is the *GPD* with density

$$f_R(x) = \frac{1}{s}\left\{1 + \frac{\xi(x - t_0)}{s}\right\}^{-(1+1/\xi)} \mathbf{1}\{x \geq t_0\}, (\xi, s) = \left(\frac{1}{4}, \frac{3}{2}\right),$$

whereas the one for Model C is the (shifted) *half-normal* distribution with density

$$f_R(x) = \frac{1}{\varsigma}\sqrt{\frac{2}{\pi}} \exp\left\{-\frac{(x - t_0)^2}{2\varsigma^2}\right\} \mathbf{1}\{x \geq t_0\}, \varsigma = \frac{3}{\sqrt{2\pi}}.$$

These setups produce the common jump size $d_0 = f_X(t_0^-) - f_X(t_0^+) \approx 0.05$. Drees et al. (2020, p.83) argue that discontinuity of the density at the threshold is an easy scenario for the threshold detection method by Clauset et al. (2009). Models B and C violate (2) by construction but may be more realistic, because it is hard to judge whether the local structure (2) indeed holds in real data. Considering that the GPD is chosen as the tail model, Model B can be thought of as most favorable to existing threshold detection methods. In contrast, Model C has a normal-type thin tail. Our aim is to see how $\hat{t}_B$ will behave when an important regularity condition is violated and/or when tail thickness changes.

We simulate 1000 Monte Carlo replications of $\{X_i\}_{i=1}^n$ with sample size $n \in \{250, 500\}$ from each model, and n observations $\{X_i\}_{i=1}^n$ in the original scale are transformed into $\{Y_i\}_{i=1}^n$ via $T_1 - T_4$ in Table 1. Invoke that each of four transformation $T_1 - T_4$ depends on the midpoint t_M of the prespecified interval I_0 in the original scale. It is also expected that $t_M = t_0$ would be the most ideal scenario for our splicing point estimation method in the transformed scale. Because it is not possible to know whether t_M coincides with (or lies in the vicinity of) t_0, we are motivated to investigate how a particular choice of I_0 will influence the finite-sample performance of $\hat{t}_B$. Then, the following three cases are examined:

$$I_0 = \begin{cases} [t_0 - 1, t_0 + 1] = [3, 5] & : \text{Case (i)} \\ [t_0 - 1/2, t_0 + 3/2] = [3.5, 5.5] & : \text{Case (ii)} \\ [t_0 - 3/2, t_0 + 1/2] = [2.5, 4.5] & : \text{Case (iii)} \end{cases}.$$

Observe that $t_0 = t_M = 4$, $t_0 = 4 < 4.5 = t_M$ and $t_0 = 4 > 3.5 = t_M$ hold for Cases (i)-(iii), respectively. For convenience, Table 1 provides the magnification factor $d_T/d_0 = 1/T^{(1)}(t_0)$ (see (7)) and its values in three cases for each transformation.

Before proceeding, Table 2 presents the mode of the distribution of X, the constant c_L, left and right limits of the density at the splicing point $f_X\left(t_0^{\pm}\right)$, and the jump size $d_0 = f_X\left(t_0^{-}\right) - f_X\left(t_0^{+}\right)$. In each model, roughly 95% of observations concentrate on the interval $[0, t_0)$ (i.e., in the bulk region). The densities for Models A and B have a polynomial decaying tail. These features reasonably mimic properties of cost distributions.

TABLE 2 ABOUT HERE

Our estimation procedure for t_0 is implemented as follows. For each of four transformations $T_1 - T_4$ in Table 1, there are two optimizations required, namely, (i) the one for tuning parameters (b, Δ) and (ii) the other for the search of the splicing point. For (i), Remark 4 suggests $\alpha \in (1/2, 3/4)$, and thus we restrict our attention to four values, namely, $\alpha \in \{0.55, 0.60, 0.65, 0.70\}$. A few cross-validation (CV) methods for b are investigated, and their details are deferred to the next section. For each CV method, candidates of b are taken from 50 equally-spaced grids over the interval $[0.005, 0.250]$. For (ii), after (b, Δ) are determined, the location of the splicing point is searched via a numerical optimization routine for the diagnostic function $\left|\hat{J}(y)\right|$ on the interval I_T in the transformed scale.[1]

Finite-sample performance of $\hat{t}_B$ is compared with those of splicing point estimators in the original scale and automated threshold detection methods. For the former, the splicing point estimator in the original scale using the shifted gamma kernels $\hat{t}_G$ and its bias-corrected version $\tilde{t}_G = \hat{t}_G + b$ by Funke and Hirukawa (2025) are investigated. For the latter, the followings are examined: (a) the minimum KS distance procedure between the empirical and GPD-based distribution functions by Clauset et al. (2009) [KS]; (b) the minimum quantile discrepancy criterion for the mean absolute deviation between empirical and GPD-based quantiles by Danielsson et al. (2019) [Q-MAD]; (c) the minimum quantile discrepancy criterion for the sup-norm between empirical and GPD-based quantiles by Danielsson et al. (2019) [Q-SUP]; (d) the automated Eye-Balling method based on tail index estimates by Danielsson et al.

[1] The reason why different algorithms are utilized for (i) and (ii) is as follows. While a numerical optimization routine substantially reduces computation time for (i), it often finds local extrema and corner solutions because of a high degree of nonlinearity in CV criteria. A grid search can circumvent these issues. In contrast, as (15) indicates, $\left|\hat{J}(y)\right|$ is concave on I_T, and a numerical optimization routine helps expedite computation for (ii).

(2019) [AEB]; and (e) the Anderson-Darling sequential testing procedure by Bader et al. (2018) [ADST]. For (e), candidates of thresholds are 20 empirical percentiles from 50.0% until 97.5% with an increment of 2.5%, i.e., $\{50.0\%, 52.5\%, \ldots, 95.0\%, 97.5\%\}$. The 5% level of significance is used for testing, and p-values for multiple tests are adjusted by the ForwardStop procedure.

All simulations are conducted on R. In particular, R-packages "`poweRlaw`", "`tea`" and "`eva`" are employed to implement automated threshold detection methods (a), (b)-(d) and (e), respectively..

4.2 Smoothing Parameter Selection

Selecting the smoothing parameter b is the most important practical issue. In our context, values of (b, Δ) must be determined before threshold location search so that the diagnostic function can be fixed on I_T. The implementation methods below largely follow those discussed in Section 4.2 of Funke and Hirukawa (2025).

Before proceeding, based on the dependence of Δ on b, put $\Delta = b^{\alpha}$ for a given α. Accordingly, $\hat{f}_Y^{\pm}(y)$ are rewritten as $\hat{f}_{Y,b}^{\pm}(y;\alpha)$, which signify the dependence of density estimates on (b, α). Also let

$$\hat{f}_{Y,b,-i}^{\pm}(y;\alpha) := \frac{1}{n-1} \sum_{j=1, j\neq i}^{n} K_y^{\pm}(Y_j)$$

be density estimates using the sample with the ith observation eliminated. Finally, denote the number of observations falling into I_T as $n_0 := \sum_{i=1}^{n} \mathbf{1}\{Y_i \in I_T\}$. Using these notations, we focus on the least-squares cross-validation (LSCV) criterion. It is defined as

$$CV_{LS}(b;\alpha) = CV_{LS}^{-}(b;\alpha) + CV_{LS}^{+}(b;\alpha), \tag{16}$$

where

$$CV_{LS}^{\pm}(b;\alpha) := \int_{I_T} \left\{\hat{f}_{Y,b}^{\pm}(y;\alpha)\right\}^2 dx - \frac{2}{n_0} \sum_{i:Y_i \in I_T} \hat{f}_{Y,b,-i}^{\pm}(Y_i;\alpha). \tag{17}$$

Corresponding splicing point estimates are labelled "SB-LS" for each of transformations $T_1 - T_4$, where "SB" and "LS" abbreviate "shifted beta" and "least squares", respectively.

Alternatively we may rely on likelihood-based criteria. The likelihood cross-validation (LCV) criterion, which is analogous to $\hat{L}_2(h)$ of Marron (1985) and equation (2.1) of Van Es (1991), is given by

$$CV_L(b;\alpha) = CV_L^{-}(b;\alpha) + CV_L^{+}(b;\alpha),$$

where

$$CV_L^{\pm}(b;\alpha) := -\sum_{i:Y_i \in I_T} \ln\left\{\hat{f}^{\pm}_{Y,b,-i}(Y_i;\alpha)\right\}$$

is the negative log-likelihood. Moreover, it is possible to turn to the modified LCV (MLCV) criterion, which corresponds to $\hat{L}_5(h)$ of Marron (1985) and equation (2.2) of Van Es (1991). It takes the form of

$$CV_{ML}(b;\alpha) = CV_{ML}^{-}(b;\alpha) + CV_{ML}^{+}(b;\alpha),$$

where

$$\begin{aligned} CV_{ML}^{\pm}(b;\alpha) &:= -\left[\sum_{i:Y_i \in I_T} \ln\left\{\hat{f}^{\pm}_{Y,b,-i}(Y_i;\alpha)\right\} - \sum_{i=1}^{n} \int_{I_T} K_{Y_i}^{\pm}(u)\,du\right] \\ &= -\left[\sum_{i:Y_i \in I_T} \ln\left\{\hat{f}^{\pm}_{Y,b,-i}(X_i;\alpha)\right\} - \sum_{i=1}^{n}\left\{B\left(\frac{Y_i \pm \Delta}{b}+1, \frac{1-(Y_i \pm \Delta)}{b}+1; \bar{t}_T\right)\right.\right. \\ &\left.\left. -B\left(\frac{Y_i \pm \Delta}{b}+1, \frac{1-(Y_i \pm \Delta)}{b}+1; \underline{t}_T\right)\right\}\right], \end{aligned}$$

and the second term is intended to eliminate the endpoint effect of the interval $I_T = [\underline{t}_T, \bar{t}_T]$. However, LCV and MLCV tend to produce very similar and quite small smoothing parameter values. As a result, splicing point estimates implemented by these CV criteria become unstable and their performance measures are inferior to those implemented by LSCV. Therefore, we abstain from reporting the results from LCV or MLCV.

On the other hand, $\hat{t}_G$ is implemented by MLCV with the exponent α fixed at 0.70. This particular implementation method is based on simulation results in Funke and Hirukawa (2025), and the MLCV criterion for $\hat{t}_G$ can be also found therein. Candidates of b are taken from 100 equally-spaced grids over the interval $[0.005, 0.500]$. Finally, the bias-corrected estimator $\tilde{t}_G = \hat{t}_G + b$ can be obtained by simply adding the smoothing parameter value chosen via MLCV to $\hat{t}_G$. Estimators $\hat{t}_G$ and $\tilde{t}_G$ are labelled "SG-ML" and "SG-ML-BC", where "SG", "ML" and "BC" abbreviate "shifted gamma", "modified likelihood" and "bias corrected," respectively.

TABLE 3 ABOUT HERE

4.3 Results

Table 3 presents several performance measures of splicing point or threshold estimators. These include the bias ("Bias"), standard deviation ("SD") and root mean squared error ("RMSE") of each threshold estimator over 1000 Monte Carlo samples. In addition, for SG and SB estimators, Monte Carlo averages and standard deviations (in parentheses) of CV smoothing parameters ("$\hat{b}$") are reported for reference.

We start from examining the results from automated threshold detection methods that are immune to the choice of I_0. It can be found that Q-MAD generates the smallest RMSE among five automated methods for each model and sample size. Although Model B is thought to be more favorable than Model A for these automated methods, RMSEs of all these methods become worse in the former case. Rather, results from Model C are comparable with those from Model A. Furthermore, there are general tendencies of underestimation by KS, Q-MAD and ADST and overestimation by Q-SUP and AEB. In particular, the degree of overestimation by AEB is often substantial.

Now we look into kernel-based splicing point estimators. It can be immediately seen that two SG estimators are inferior to SB estimators in terms of SD and RMSE, regardless of the choice of I_0. A large SD in SG indicates that SG tends to be less efficient in finite samples than SB. It is not surprising that the splicing point estimate by SG becomes more volatile. Invoke that in our Monte Carlo design, the discontinuous point, which is located over the right-tail region in the original scale, has a tiny jump size. Under such circumstances, SG methods have difficulty in spotting a vertical gap over the sparse and fairly flat region. It turns out that data transformation is a better option than bias correction in the original scale.

While RMSEs from SG do not decrease with the sample size, those from SB do so in many cases. The latter numerically confirms consistency of SB. A closer look also clarifies that results from SB are largely affected by choices of the interval I_0, the transformation T and the exponent α. First, as regards the interval, Case (ii), i.e., $I_0 = [3.5, 5.5]$, produces the smallest RMSE for each SB estimator. This implies that SB method can estimate the splicing point more precisely if it lies in the left-hand side of the prespecified interval. In this view, it could be ideal to set the interval longer on the side of the right tail. Second, as regards the transformation, T_3 has a clear advantage in terms of RMSE. Superior performance of T_3 can be attributed to the largest magnification factor among four transformations. Therefore, it is advisable to choose a transformation that yields a large magnification factor around the midpoint of I_0. Third, as regards the exponent, $\alpha = 0.70$ outperforms other cases. These general tendencies hold regardless of whether the local structure (2) holds in the neighborhood of the true splicing point (Model A) or it is violated completely (Models B and C).

Before proceeding, it is worth giving readers a few words of warning about the interpretation of results from Model C. SB-LS-T_3 continues to dominate in terms of RMSE over other competing methods. RMSEs of all five automated threshold detection methods become smaller than those from Model B and comparable with those from Model A. However, Model C has a normal-type thin tail. Accordingly, such small estimation errors should be interpreted with caution. SB (and SG) methods are designed to find a vertical gap in the density, regardless of the degree of its tail thickness. Improvement of RMSEs by five automated methods from Model B (= their most favorable scenario) can be also regarded as paradoxical. These results underpin the assertion by Clauset et al. (2009) that conducting a goodness-of-fit test for the tail part is highly recommended after threshold detection; see Remark 5 for reference.

At any rate, Monte Carlo results indicate that SB-LS-T_3 with $\alpha = 0.70$ consistently outperforms automated threshold detection and other kernel-based splicing point estimation methods in terms of RMSE. It is confirmed that this estimator has most practical relevance and importance, and thus it will be applied to real data examples in the next section.

5 Real Data Examples

5.1 Data Description

We evaluate our proposed methodology using four openly available datasets of non-life insurance losses. The datasets include: (A) Danish fire insurance losses; (B) Norwegian fire insurance losses; (C) Belgian motor insurance losses; and (D) French motor insurance losses. The dataset (A), which consists of $2,492$ observations of fire-related losses recorded between 1980 and 1990, has achieved widespread recognition in non-life actuarial mathematics following the influential analysis by McNeal (1997). This dataset has played a significant role in the development of statistical methods for insurance mathematics, particularly in the modeling of loss distributions. Its extensive adoption in the actuarial literature has made it a de facto benchmark for evaluating new methodological approaches in non-life insurance modeling. The dataset is available under the name "`danish`" in the R-package "`SMPracticals`".

Datasets (B)-(D) correspond to "`norfire`", "`beMTPL97`" and "`freMTPL2sev`" in the R-package "`CASdatasets`", respectively. For (B), the variable "`Loss2012`" (= total loss amount denominated in 2012 constant Norwegian kroner) is chosen. Each of these datasets has more observations than (A). To expedite computation, we select

1/3, 1/6 and 1/10 subsamples randomly from full samples of (B)-(D), respectively, so that the number of observations of each downsized sample is roughly the same as that of (A).

The summary statistics presented in Table 4 confirm that all four datasets exhibit the characteristic features of loss distributions, including substantial right-skewness and heavy tails. It can be also found that the downsized samples for (B)-(D) well mimic distributional characteristics of their original samples.

TABLES 4-5 ABOUT HERE

5.2 Estimation Details

Judging from Monte Carlo results, we adopt SB-LS-T$_3$ with $\alpha = 0.70$ and compare it with other competing methods, namely, five automated threshold detection methods KS, Q-MAD, Q-SUP, AEB, ADST, and kernel-based splicing point estimation methods SG-ML(-BC) in the original scale. As in Section 4, CV smoothing parameters are chosen over 50 equally-spaced grids over the interval $[0.005, 0.250]$ for SB estimators and 100 equally-spaced points over $[0.005, 0.500]$ for SG estimators.

The selection of I_0 builds on both theoretical considerations and previous empirical findings. For (A), our choice incorporates insights from earlier studies, including the parametric composite model estimates of Cooray and Ananda (2005) and Scollnik and Sun (2012), who identify thresholds between 1 and 3 million kroner, as well as the graphical analysis of Reynkens et al. (2017), who suggest a higher threshold around 17 million kroner. Moreover, we prespecify the intervals for (B)-(D) to verify threshold estimates from all automated methods but ADST.

ADST is also implemented as follows. The 5% level of significance is adopted, and p-values for multiple tests are adjusted by the ForwardStop procedure. Candidates of thresholds are: 54 empirical percentiles from 20.0% until 99.5% with an increment of 1.5%, i.e., $\{20.0\%, 21.5\%, \ldots, 98.0\%, 99.5\%\}$, for (A) and (B); 14 empirical percentiles from 80.0% until 99.5% with an increment of 1.5%, i.e., $\{80.0\%, 81.5\%, \ldots, 98.0\%, 99.5\%\}$, for (C); and 34 empirical percentiles from 50.0% until 99.5% with an increment of 1.5%, i.e., $\{50.0\%, 51.5\%, \ldots, 98.0\%, 99.5\%\}$, for (D). The range of percentiles for each dataset roughly coincides with the corresponding interval I_0. Finally, if all threshold candidates are rejected, then the largest (= 99.5%) empirical percentile is chosen as the threshold estimate, as described on p.323 of Bader et al. (2018).

5.3 Estimation Results

Table 5 presents the estimation results. Superscripts "†" on ADST indicate that the 99.5% empirical percentile is chosen because all threshold candidates are rejected. In addition, superscripts "‡" on SG-ML mean that an estimation failure occurs as the diagnostic function is maximized at a boundary of I_0.

No estimation problems arise for (A) and (B). In these cases, estimates from SB-LS-T$_3$ are close to those from KS, ADST and SG-ML(-BC). On the other hand, for (C) and (D), both ADST and SG-ML exhibit estimation problems, whereas SB-LS-T$_3$ still yields splicing point estimates. For (C), the SB-LS-T$_3$ estimate is close to those of KS and Q-MAD, and for (D), it lies between those of Q-SUP and AEB.

The true location of the threshold or splicing point is unknown in each application. Nonetheless, our approach can provide an estimate even when some other competing methods fail to do so. Therefore, it represents a valuable complement to existing methods for threshold detection, offering enhanced capabilities for model validation.

6 Conclusion

This paper has introduced a novel nonparametric approach for estimating splicing points in actuarial loss distributions, a critical task for accurate risk modeling, pricing, and solvency capital determination. Recognizing the difficulties in detecting subtle discontinuities in the tail of loss distributions, particularly when data are sparse, our method leverages a data transformation technique. By mapping original loss observations onto the unit interval $[0,1]$, we effectively magnify the jump size at the splicing point and enhance the proximity of adjacent data points. The subsequent estimation in this transformed scale is performed using the asymmetric beta kernel, which adeptly handles boundary effects inherent to the unit interval and offers flexibility in capturing various distributional shapes without imposing strong parametric assumptions.

We have established the key theoretical properties of our splicing point estimator, including its strong consistency and asymptotic normality. Notably, the estimator is super-consistent by achieving a faster convergence rate than the standard parametric $\sqrt{n}$-rate. This is particularly advantageous in actuarial tasks, where such an estimator can be reliably used in multi-step estimation procedures or for subsequent model validation without dominating the uncertainty of other parameter estimates. Monte Carlo simulations have confirmed the attractive finite-sample performance of our method, and its practical applicability has been demonstrated through real-world insurance loss datasets, showcasing its potential as a valuable and robust addition to the actuary's toolkit for threshold identification.

Several avenues for future research emerge from this work. A natural extension would be the development of a framework to detect and estimate multiple splicing points, as complex loss dynamics may exhibit several distinct changes in their distributional characteristics. Incorporating covariate information into the estimation of splicing points could also provide deeper insights, allowing thresholds to vary based on specific risk factors or policy characteristics. Further investigation into data-driven methods for selecting the optimal transformation function T could enhance the method's objectivity and performance across diverse datasets.

A Appendix

The Appendix provides technical proofs of theorems and propositions. To save space, we defer proofs of lemmata to the Supplemental Material. Before proceeding, the following shorthand notation is adopted: for $a > 0$, $\Gamma(a) = \int_0^\infty t^{a-1} \exp(-t)\, dt$ is the gamma function; $\Psi(a) = d \ln \Gamma(a)/da = \Gamma^{(1)}(a)/\Gamma(a)$ and $\Psi^{(m)}(a) = d^m \Psi(a)/da^m$ signify the digamma and polygamma functions, respectively; $\dot{K}_c^{\pm}(u) = \partial K_y^{\pm}(u)/\partial y\big|_{y=c}$; $\ddot{K}_c^{\pm}(u) = \partial^2 K_y^{\pm}(u)/\partial y^2\big|_{y=c}$; $\dddot{K}_c^{\pm}(u) = \partial^3 K_y^{\pm}(u)/\partial y^3\big|_{y=c}$; and $H_i = \dot{K}_{t_T}^{-}(Y_i) - \dot{K}_{t_T}^{+}(Y_i)$. Finally, $Z_{b,y,m}^{\pm}$ is the beta random variable so that $Z_{b,y,m}^{\pm} \overset{d}{=} Beta\left(p_{b,y,m}^{\pm}, q_{b,y}^{\pm}\right) := Beta\left\{(y \pm \Delta)/b + m + 1, (1 - (y \pm \Delta))/b + 1\right\}$ for $m \in \{0, 1, 2, \ldots\}$.

A.1 Proof of Proposition 1

This proposition can be established by a minor modification of the proof of Theorem 2 in Hirukawa et al. (2022), and thus details are omitted. ■

A.2 Proof of Proposition 2

The proof requires the following lemmata. While the case with $m = 0$ in Lemma A2 is relevant for this proof, the cases with $m \geq 1$ will be employed for the proof of Theorem 2. Moreover, Lemma A3 is known as the Komatsu inequality.

Lemma A1. *Let $Z \overset{d}{=} Beta(p, q)$, where $p, q \to \infty$ and $p \asymp q$ so that both $q = O(p)$ and $p = O(q)$ hold. Also suppose that the argument z in the pdf of the beta random variable Z admits the location-scale transformation $z := \mu + \sigma v$, where*

$$\mu = E(Z) = \frac{p}{p+q},$$

and

$$\sigma^2 = Var\left(Z\right) = \frac{pq}{\left(p+q\right)^2\left(p+q+1\right)}.$$

Then, the pdf of Z can be approximated by

$$\begin{aligned} f_Z\left(z\right) &= f_Z\left(\mu+\sigma v\right) \\ &= \frac{\phi\left(v\right)}{\sigma}\left[1+\frac{v}{\sqrt{p+q+1}}\left(\sqrt{\frac{p}{q}}-\sqrt{\frac{q}{p}}\right)\right. \\ &\left. +\frac{v^3}{3}\left\{\left(\frac{q}{p+q}\right)^{3/2}\frac{1}{\sqrt{p}}-\left(\frac{p}{p+q}\right)^{3/2}\frac{1}{\sqrt{q}}\right\}+O\left(p^{-1}\right)\right]. \end{aligned}$$

Lemma A2. *Let $f_{Z^{\pm}_{b,y,m}}\left(z\right)$, $\mu^{\pm}_{b,y,m}$* and *$\left(\sigma^{\pm}_{b,y,m}\right)^2$ be the pdf, the mean and the variance of the beta random variable $Z^{\pm}_{b,y,m}$, respectively, where*

$$\mu^{\pm}_{b,y,m} = E\left(Z^{\pm}_{b,y,m}\right) = \frac{y\pm\Delta+\left(m+1\right)b}{1+\left(m+2\right)b}$$

and

$$\left(\sigma^{\pm}_{b,y,m}\right)^2 = Var\left(Z^{\pm}_{b,y,m}\right) = \frac{b\left\{y\pm\Delta+\left(m+1\right)b\right\}\left\{1-\left(y\pm\Delta\right)+b\right\}}{\left\{1+\left(m+2\right)b\right\}^2\left\{1+\left(m+3\right)b\right\}}.$$

Then, as $n\to\infty$, the incomplete beta function ratio $B\left(p^{\pm}_{b,y,m},q^{\pm}_{b,y};t_T\right)=\int_0^{t_T} f_{Z^{\pm}_{b,y,m}}\left(z\right)dz$ can be approximated by

$$\int_0^{t_T} f_{Z^{\pm}_{b,y,m}}\left(z\right)dz := C^{\pm}_{b,y,m}\left(0\right)+b^{1/2}\left\{\xi^{\pm}_{b,y,m}C^{\pm}_{b,y,m}\left(1\right)+\zeta^{\pm}_{b,y,m}C^{\pm}_{b,y,m}\left(3\right)\right\}+O\left(b\right), \quad \text{(A1)}$$

where

$$\begin{aligned} C^{\pm}_{b,y,m}\left(k\right) &= \int_{B^{\pm}_{b,y,m}}^{A^{\pm}_{b,y,m}} v^k\phi\left(v\right)dv \textit{ for } k=0,1,3, \\ A^{\pm}_{b,y,m} &= \frac{t_T-\mu^{\pm}_{b,y,m}}{\sigma^{\pm}_{b,y,m}} = \frac{\left[t_T-y\mp\Delta+\left\{\left(m+2\right)t_T-\left(m+1\right)\right\}b\right]\sqrt{1+\left(m+3\right)b}}{\sqrt{b\left\{\left(y\pm\Delta\right)+\left(m+1\right)b\right\}\left\{1-\left(y\pm\Delta\right)+b\right\}}}, \\ B^{\pm}_{b,y,m} &= -\frac{\mu^{\pm}_{b,y,m}}{\sigma^{\pm}_{b,y,m}} = \frac{\left\{-y\mp\Delta-\left(m+1\right)b\right\}\sqrt{1+\left(m+3\right)b}}{\sqrt{b\left\{\left(y\pm\Delta\right)+\left(m+1\right)b\right\}\left\{1-\left(y\pm\Delta\right)+b\right\}}}, \\ \xi^{\pm}_{b,y,m} &= \frac{b^{-1/2}}{\sqrt{p^{\pm}_{b,y,m}+q^{\pm}_{b,y}+1}}\left(\sqrt{\frac{p^{\pm}_{b,y,m}}{q^{\pm}_{b,y}}}-\sqrt{\frac{q^{\pm}_{b,y}}{p^{\pm}_{b,y,m}}}\right), \\ \zeta^{\pm}_{b,y,m} &= \frac{b^{-1/2}}{3}\left\{\left(\frac{q^{\pm}_{b,y}}{p^{\pm}_{b,y,m}+q^{\pm}_{b,y}}\right)^{3/2}\frac{1}{\sqrt{p^{\pm}_{b,y,m}}}-\left(\frac{p^{\pm}_{b,y,m}}{p^{\pm}_{b,y,m}+q^{\pm}_{b,y}}\right)^{3/2}\frac{1}{\sqrt{q^{\pm}_{b,y}}}\right\}, \end{aligned}$$

and the $O(b)$ rate in (A1) is uniform on $y \in I_T$. Furthermore, as $n \to \infty$,

$$\xi^{\pm}_{b,y,m} = \frac{2y-1}{\sqrt{y(1-y)}} \pm \frac{\Delta}{2\{y(1-y)\}^{3/2}} + O(b),$$

and

$$\zeta^{\pm}_{b,y,m} = \frac{1-2y}{3\sqrt{y(1-y)}} \mp \frac{\Delta}{6\{y(1-y)\}^{3/2}} + O(b),$$

where $O(b)$ rates are again uniform on $y \in I_T$.

Lemma A3 (Komatsu, 1955). *For $x > 0$,*

$$\frac{2}{\sqrt{x^2+4}+x} < e^{\frac{x^2}{2}} \int_x^{\infty} e^{-\frac{t^2}{2}} dt < \frac{2}{\sqrt{x^2+2}+x}.$$

A.2.1 Proof of Proposition 2

Notice that it is easiest and fastest to verify all the calculations in this proof with the aid of Maple™ or Mathematica®.

Proof of (i). Put $m = 0$ in (A1). In this case, $\int_0^{t_T} K_y^{\pm}(u)\,du = B\left(p^{\pm}_{b,y,0}, q^{\pm}_{b,y}; t_T\right) = \int_0^{t_T} f_{Z^{\pm}_{b,y,0}}(z)\,dz$. Lemma A2 implies that each of $\xi^{\pm}_{b,y,0}$, $\varsigma^{\pm}_{b,y,0}$, $C^{\pm}_{b,y,0}(1)$, and $C^{\pm}_{b,y,0}(3)$ is at most $O(1)$ uniformly on $y \in I_T$. Then,

$$\int_0^{t_T} K_y^{\pm}(u)\,du = C^{\pm}_{b,y,0}(0) + O\left(b^{1/2}\right) = \Phi\left(A^{\pm}_{b,y,0}\right) - \Phi\left(B^{\pm}_{b,y,0}\right) + O\left(b^{1/2}\right),$$

where $\Phi(\cdot)$ is the cdf of $N(0,1)$ and the $O\left(b^{1/2}\right)$ rate is uniform on $y \in I_T$.

We start from working on $\Phi\left(A^{\pm}_{b,y,0}\right)$, where

$$A^{\pm}_{b,y,0} = \frac{\{t_T - y \mp \Delta + (2t_T - 1)b\}\sqrt{1+3b}}{\sqrt{b\{(y \pm \Delta) + b\}\{1-(y \pm \Delta)+b\}}}.$$

By a third-order Taylor expansion of $\Phi\left(A^{\pm}_{b,y,0}\right)$ around $\Delta = 0$,

$$\Phi\left(A^{\pm}_{b,y,0}\right) = \Phi\left(A^{\pm}_{b,y,0}\right)\Big|_{\Delta=0} + \phi\left(A^{\pm}_{b,y,0}\right) \frac{\partial A^{\pm}_{b,y,0}}{\partial \Delta}\Bigg|_{\Delta=0} \Delta + O\left(\frac{\Delta^2}{b}\right) + O\left(\frac{\Delta^3}{b^{3/2}}\right),$$

where the coefficient on the $O(\Delta^2/b)$ term is shown to be common across $\Phi\left(A^{+}_{b,y,0}\right)$ and $\Phi\left(A^{-}_{b,y,0}\right)$ (although it is not specified explicitly), and the $O\left(\Delta^3/b^{3/2}\right)$ rate is uniform on $y \in I_T$. By straightforward calculations,

$$A^{\pm}_{b,y,0}\Big|_{\Delta=0} = \sqrt{\frac{1+3b}{b(y+b)(1-y+b)}}\{t_T - y + (2t_T-1)b\} := A_{b,y,0},$$

and

$$\left.\frac{\partial A^{\pm}_{b,y,0}}{\partial \Delta}\right|_{\Delta=0} = \mp \frac{(1+2b)\sqrt{1+3b}\,\{(1-2y)\,t_T + (y+b)\}}{2b^{1/2}\,\{(y+b)\,(1-y+b)\}^{3/2}}.$$

It follows that

$$\begin{aligned}\Phi\left(A^{\pm}_{b,y,0}\right) &= \Phi\left(A_{b,y,0}\right) \mp \phi\left(A_{b,y,0}\right) \frac{(1+2b)\sqrt{1+3b}\,\{(1-2y)\,t_T + (y+b)\}}{2b^{1/2}\,\{(y+b)\,(1-y+b)\}^{3/2}} \left(\frac{\Delta}{b^{1/2}}\right) \\ &\quad + O\left(\frac{\Delta^2}{b}\right) + O\left(\frac{\Delta^3}{b^{3/2}}\right).\end{aligned}$$

Next, we demonstrate that for each fixed m, $\Phi\left(B^{\pm}_{b,y,m}\right) \to 0$ at an exponential rate as $n \to \infty$. It can be immediately found that $B^{\pm}_{b,y,m} = -b^{-1/2}\sqrt{y/\left(1-y\right)}\,\{1+o\left(1\right)\}$ for each fixed m, where the $o\left(1\right)$ term is uniform on $y \in I_T$. Hence, on $y \in I_T$, there are constants $0 < \underline{C}_B < \overline{C}_B < \infty$ so that $\underline{C}_B b^{-1/2} \le \left|B^{\pm}_{b,y,m}\right| \le \overline{C}_B b^{-1/2}$. In addition, Lemma A3 implies that

$$\frac{2e^{-x^2/2}}{\sqrt{2\pi}\left(\sqrt{x^2+4}+x\right)} < \int_x^{\infty} \frac{e^{-t^2/2}}{\sqrt{2\pi}} dt < \frac{2e^{-x^2/2}}{\sqrt{2\pi}\left(\sqrt{x^2+2}+x\right)}.$$

When $x = \left|B^{\pm}_{b,y,m}\right|$, lower and upper bounds of this double inequality are bounded respectively by

$$\frac{2e^{-x^2/2}}{\sqrt{2\pi}\left(\sqrt{x^2+4}+x\right)} \ge \frac{2e^{-\overline{C}_B^2/(2b)}}{\sqrt{2\pi}\left(\sqrt{\overline{C}_B^2/b+4}+\overline{C}_B/b^{1/2}\right)} = O\left\{b^{1/2}\exp\left(-\frac{\overline{C}_B^2}{2b}\right)\right\},$$

and

$$\frac{2e^{-x^2/2}}{\sqrt{2\pi}\left(\sqrt{x^2+2}+x\right)} \le \frac{2e^{-\underline{C}_B^2/(2b)}}{\sqrt{2\pi}\left(\sqrt{\underline{C}_B^2/b+2}+\underline{C}_B/b^{1/2}\right)} = O\left\{b^{1/2}\exp\left(-\frac{\underline{C}_B^2}{2b}\right)\right\}.$$

These converge to zero at an exponential rate as $n \to \infty$, so does $\Phi\left(B^{\pm}_{b,y,m}\right) = \int_{\left|B^{\pm}_{b,y,m}\right|}^{\infty}\left(e^{-t^2/2}/\sqrt{2\pi}\right)dt$.

Combining above results, we conclude that

$$\begin{aligned}&\int_0^{t_T} K^{\pm}_y\left(u\right)du \\ &= \Phi\left(A_{b,y,0}\right) \mp \phi\left(A_{b,y,0}\right) \frac{(1+2b)\sqrt{1+3b}\,\{(1-2y)\,t_T + (y+b)\}}{2b^{1/2}\,\{(y+b)\,(1-y+b)\}^{3/2}} \left(\frac{\Delta}{b^{1/2}}\right) \\ &\quad + O\left(\frac{\Delta^2}{b}\right) + O\left(\frac{\Delta^3}{b^{3/2}}\right).\end{aligned}$$

Part (i) can be established because $J\left(y\right) = \int_0^{t_T} K^{-}_y\left(u\right)du - \int_0^{t_T} K^{+}_y\left(u\right)du$ holds, the $O\left(\Delta/b^{1/2}\right)$ terms can be rewritten as $\mp\left(1/2\right)Q\left(y\right)\left(\Delta/b^{1/2}\right)$, and the $O\left(\Delta^2/b\right)$ terms are cancelled out.

Proof of (ii). $Q(y)$ can be further rewritten as

$$Q(y) := \frac{(1+2b)\sqrt{1+3b}}{\sqrt{2\pi}} Q_0(y),$$

where

$$Q_0(y) = \frac{(1-2y)t_T + (y+b)}{\{(y+b)(1-y+b)\}^{3/2}} \exp\left[-\frac{(1+3b)\{t_T - y + (2t_T - 1)b\}^2}{2b(y+b)(1-y+b)}\right].$$

It follows that

$$Q_0^{(1)}(y) := \frac{\psi(y)}{2b\{(y+b)(1-y+b)\}^{7/2}} \exp\left[-\frac{(1+3b)\{t_T - y + (2t_T - 1)b\}^2}{2b(y+b)(1-y+b)}\right],$$

where

$$\begin{aligned}
\psi(y) &= \psi_0(y) + \psi_1(y)b + O(b^2), \\
\psi_0(y) &= -(y - t_T)\{(2t_T - 1)y - t_T\}^2, \\
\psi_1(y) &= 7(2y-1)^2 t_T^3 + \left(-20y^3 - 12y^2 + 9y + 1\right) t_T^2 \\
&\quad + \left(8y^4 + 4y^3 + 16y^2 - 5y\right) t_T - 4y^4 - 4y^2,
\end{aligned}$$

and the $O(b^2)$ rate in $\psi(y)$ is uniform on $y \in I_T$.

Heuristically, $\psi(y) \approx \psi_0(y) = -(y - t_T)\{(2t_T - 1)y - t_T\}^2$ for $b \approx 0$, and it is straightforward to see that $Q(y)$ is maximized at $y \approx t_T$. Let t_T^* solve $\psi(y) = 0$. Our argument so far suggests that $t_T^* \approx t_T$ holds for a sufficiently small $b > 0$. Then, it is reasonably conjectured that t_T^* can be expanded up to the $O(b^2)$ term, taking the form of $t_T^* = t_T + cb + O(b^2)$ for some $|c| < \infty$. The remaining task is to determine c. Substituting t_T^* into $\psi(y)$ gives

$$\psi(t_T^*) = \psi\left\{t_T + cb + O\left(b^2\right)\right\} = 4t_T^2 (t_T - 1)^2 \{-c + 2(2t_T - 1)\} b + O\left(b^2\right).$$

It suffices to find c that makes the right-hand side at most $O(b^2)$, and this occurs when $c = 2(2t_T - 1)$. Then, we have $t_T^* = t_T + 2(2t_T - 1)b + O(b^2)$, which completes the proof. ■

A.3 Proof of Theorem 1

It follows from Assumption 3(ii) and the definitions of $\hat{t}_B$ and t_T that $\left|\hat{t}_B - t_0\right| = \left|T^{-1}\left(\hat{t}_T\right) - T^{-1}(t_T)\right| \le \left|\hat{t}_T - t_T\right| / \underline{T}^{(1)}$. Because $\left|\hat{t}_T - t_T\right| \le \left|\hat{t}_T - t_T^*\right| + |t_T^* - t_T|$ and Proposition 2(ii) implies that $|t_T^* - t_T| = O(b)$, we only need to demonstrate that $\left|\hat{t}_T - t_T^*\right| = O(c_n)$ *a.s.* However, this statement can be established by a similar argument to the proof of Theorem 1 in Funke and Hirukawa (2025), and thus details are omitted. ■

A.4 Proof of Theorem 2

The proof requires the following lemmata.

Lemma A4. *Put $y = t_T$ in $C^{\pm}_{b,y,m}(0)$, $C^{\pm}_{b,y,m}(1)$ and $C^{\pm}_{b,y,m}(3)$ defined in Lemma A2. Then, as $n \to \infty$,*

$$\begin{aligned} C^{\pm}_{b,t_T,m}(0) &= \frac{1}{2} \mp \frac{1}{\sqrt{2\pi}\sqrt{t_T(1-t_T)}}\left(\frac{\Delta}{b^{1/2}}\right) + \frac{(2t_T-1)-m(1-t_T)}{\sqrt{2\pi}\sqrt{t_T(1-t_T)}}b^{1/2} \\ &\mp \left[\frac{m}{\sqrt{2\pi}\sqrt{t_T(1-t_T)}} + \frac{1-m^2(1-t_T)^2}{2\sqrt{2\pi}\{t_T(1-t_T)\}^{3/2}} + \right]\Delta b^{1/2} \\ &+ O\left\{\max\left(b^{3/2}, \frac{\Delta^2}{b}\right)\right\}, \qquad \text{(A2)} \\ C^{\pm}_{b,t_T,m}(1) &= -\frac{1}{\sqrt{2\pi}} \mp \frac{(2t_T-1)-m(1-t_T)}{\sqrt{2\pi}t_T(1-t_T)}\Delta + O(b), \qquad \text{(A3)} \end{aligned}$$

and

$$C^{\pm}_{b,t_T,m}(3) = -\frac{2}{\sqrt{2\pi}} + O(b). \qquad \text{(A4)}$$

Lemma A5. *As $n \to \infty$,*

$$E\left\{\left(\frac{b^{1/2}}{\Delta}\right)\hat{J}^{(1)}(t_T)\right\} \to \frac{d_T(2-t_T)}{\sqrt{2\pi}\{t_T(1-t_T)\}^{3/2}}.$$

Lemma A6. *As $n \to \infty$,*

$$Var\left\{\sqrt{\frac{nb^{5/2}}{\Delta^2}}\hat{J}^{(1)}(t_T)\right\} \to \frac{3}{2\sqrt{\pi}\{t_T(1-t_T)\}^{5/2}}\left\{\frac{f_Y(t_T^-)+f_Y(t_T^+)}{2}\right\}.$$

Lemma A7. *If $\left|\hat{t}_T - t_T\right| = o_p\left(b^{1/2}\right)$, then, as $n \to \infty$,*

$$\left(\frac{b^{3/2}}{\Delta}\right)\hat{J}^{(2)}(\varsigma) \xrightarrow{p} -\sqrt{\frac{2}{\pi}}\frac{d_T}{\{t_T(1-t_T)\}^{3/2}}$$

for any ς on the line segment joining $\hat{t}_T$ and t_T.

Lemma A8. *As $n \to \infty$, $E|H_i|^3 = O(\Delta^3/b^4)$.*

A.4.1 Proof of Theorem 2

In this proof, the limiting distribution of $\hat{t}_T$ in the transformed scale is first derived, and then the result is converted to the one in the original scale. Taking a mean-value expansion for the left-hand side of the first-order condition $\hat{J}^{(1)}\left(\hat{t}_T\right) = 0$, we have

$$\begin{aligned} 0 &= \hat{J}^{(1)}\left(t_T\right) + \hat{J}^{(2)}\left(\bar{t}_T\right)\left(\hat{t}_T - t_T\right) \\ &= E\left\{\hat{J}^{(1)}\left(t_T\right)\right\} + \left[\hat{J}^{(1)}\left(t_T\right) - E\left\{\hat{J}^{(1)}\left(t_T\right)\right\}\right] + \hat{J}^{(2)}\left(\bar{t}_T\right)\left(\hat{t}_T - t_T\right) \qquad \text{(A5)} \end{aligned}$$

for some $\bar{t}_T$ on the line segment joining $\hat{t}_T$ and t_T.

Rearranging (A5), we obtain

$$\begin{aligned} &\sqrt{\frac{n}{b^{1/2}}}\left[\hat{t}_T - t_T - \left\{-\frac{E\left(\hat{J}^{(1)}\left(t_T\right)\right)}{\hat{J}^{(2)}\left(\bar{t}_T\right)}\right\}\right] \\ &= -\sqrt{\frac{n}{b^{1/2}}}\left[\frac{\hat{J}^{(1)}\left(t_T\right) - E\left\{\hat{J}^{(1)}\left(t_T\right)\right\}}{\hat{J}^{(2)}\left(\bar{t}_T\right)}\right]. \qquad \text{(A6)} \end{aligned}$$

Theorem 1 allows us to use Lemma A7, which is satisfied. Then, in conjunction with Lemma A5, the bias term can be simplified as

$$-\frac{E\left\{\hat{J}^{(1)}\left(t_T\right)\right\}}{\hat{J}^{(2)}\left(\bar{t}_T\right)} = \left(1 - \frac{t_T}{2}\right) b\left\{1 + o_p(1)\right\}.$$

To demonstrate asymptotic normality of $\sqrt{n/b^{1/2}}\left(b^{3/2}/\Delta\right)\left[\hat{J}^{(1)}\left(t_T\right) - E\left\{\hat{J}^{(1)}\left(t_T\right)\right\}\right]$, we also check Lyapunov's condition. This quantity can be expressed as

$$\sum_{i=1}^{n} R_i := \sum_{i=1}^{n}\sqrt{\frac{b^{5/2}}{n\Delta^2}}\left\{H_i - E\left(H_i\right)\right\}.$$

Then, by C_r-inequality, Jensen's inequality (due to the convexity of z^3 for $z \geq 0$) and Lemma A8,

$$E\left|R_i\right|^3 \leq 8\left(\frac{b^{5/2}}{n\Delta^2}\right)^{3/2} E\left|H_i\right|^3 = O\left(n^{-3/2}b^{-1/4}\right).$$

It also follows from Lemma A6 that $Var\left(R_i\right) = O\left(n^{-1}\right)$. Therefore,

$$\frac{\sum_{i=1}^{n} E\left|R_i\right|^3}{\left\{\sum_{i=1}^{n} Var\left(R_i\right)\right\}^{3/2}} = O\left(\frac{1}{\sqrt{nb^{1/2}}}\right) \to 0,$$

and Lyapunov's condition is established.

By a central limit theorem, in conjunction with Lemma A6,

$$\sqrt{\frac{n}{b^{1/2}}}\left(\frac{b^{3/2}}{\Delta}\right)\left[\hat{J}^{(1)}\left(t_T\right)-E\left\{\hat{J}^{(1)}\left(t_T\right)\right\}\right]$$
$$\xrightarrow{d} N\left(0,\frac{3}{2\sqrt{\pi}\left\{t_T\left(1-t_T\right)\right\}^{5/2}}\left\{\frac{f_Y(t_T^-)+f_Y(t_T^+)}{2}\right\}\right).$$

Finally, using Lemma A7 for the right-hand side of (A6) yields

$$-\sqrt{\frac{n}{b^{1/2}}}\left[\frac{\left(b^{3/2}/\Delta\right)\left\{\hat{J}^{(1)}\left(t_T\right)-E\left\{\hat{J}^{(1)}\left(t_T\right)\right\}\right\}}{\left(b^{3/2}/\Delta\right)\hat{J}^{(2)}\left(\bar{t}_T\right)}\right]$$
$$\xrightarrow{d} N\left(0,\frac{3\sqrt{\pi}\sqrt{t_T\left(1-t_T\right)}}{4d_T^2}\left\{\frac{f_Y(t_T^-)+f_Y(t_T^+)}{2}\right\}\right).$$

So far we have obtained

$$\sqrt{\frac{n}{b^{1/2}}}\left[\hat{t}_T-t_T-\left(1-\frac{t_T}{2}\right)b\left\{1+o_p\left(1\right)\right\}\right]$$
$$\xrightarrow{d} N\left(0,\frac{3\sqrt{\pi}\sqrt{t_T\left(1-t_T\right)}}{4d_T^2}\left\{\frac{f_Y(t_T^-)+f_Y(t_T^+)}{2}\right\}\right). \tag{A7}$$

By the definitions of $\hat{t}_B$ and t_T, the left-hand side of (A7) can be rewritten as

$$\sqrt{\frac{n}{b^{1/2}}}\left[T\left(\hat{t}_B\right)-T\left(t_0\right)-\left\{1-\frac{T\left(t_0\right)}{2}\right\}b\left\{1+o_p\left(1\right)\right\}\right].$$

Now, by a mean-value expansion, $T\left(\hat{t}_B\right)-T\left(t_0\right)=T^{(1)}\left(\check{t}\right)\left(\hat{t}_B-t_0\right)$ for some $\check{t}$ on the line segment joining $\hat{t}_B$ and t_0. By Assumption 3(ii), $T^{(1)}$ is Lipschitz continuous on $\mathbb{R}_+$, and thus we may also take $M>0$ as the Lipschitz constant. Because $\check{t}$ lies between $\hat{t}_B$ and t_0 and $\left|\hat{t}_B-t_0\right|=O_p\left(c_n\right)=o_p\left(b^{1/2}\right)$ by Theorem 1, it holds that $\left|\check{t}-t_0\right|\leq\left|\hat{t}_B-t_0\right|=o_p\left(b^{1/2}\right)$. Therefore,

$$\begin{aligned}&\left|\left\{T\left(\hat{t}_B\right)-T\left(t_0\right)\right\}-T^{(1)}\left(t_0\right)\left(\hat{t}_B-t_0\right)\right|\\&=\left|T^{(1)}\left(\check{t}\right)-T^{(1)}\left(t_0\right)\right|\left|\hat{t}_B-t_0\right|\\&\leq M\left|\check{t}-t_0\right|\left|\hat{t}_B-t_0\right|\leq M\left|\hat{t}_B-t_0\right|^2=o_p\left(b\right).\end{aligned}$$

In the end,

$$\begin{aligned}&\sqrt{\frac{n}{b^{1/2}}}\left[T\left(\hat{t}_B\right)-T\left(t_0\right)-\left\{1-\frac{T\left(t_0\right)}{2}\right\}b\left\{1+o_p\left(1\right)\right\}\right]\\&=\sqrt{\frac{n}{b^{1/2}}}\left[T^{(1)}\left(t_0\right)\left(\hat{t}_B-t_0\right)+o_p\left(b\right)-\left\{1-\frac{T\left(t_0\right)}{2}\right\}b\left\{1+o_p\left(1\right)\right\}\right]\\&=\sqrt{\frac{n}{b^{1/2}}}T^{(1)}\left(t_0\right)\left[\hat{t}_B-t_0-\left\{\frac{1-T\left(t_0\right)/2}{T^{(1)}\left(t_0\right)}\right\}b\left\{1+o_p\left(1\right)\right\}\right].\end{aligned} \tag{A8}$$

It also follows from (6) and (7) that

$$
\begin{aligned}
\frac{f_Y(t_T^-)+f_Y(t_T^+)}{2} &= g_Y\left(t_T\right)+\frac{d_T}{2} \\
&= \frac{1}{T^{(1)}\left(t_0\right)}\left\{g_Y\left\{T\left(t_0\right)\right\}T^{(1)}\left(t_0\right)+\frac{d_T T^{(1)}\left(t_0\right)}{2}\right\} \\
&= \frac{1}{T^{(1)}\left(t_0\right)}\left\{f_X(t_0^+)+\frac{d_0}{2}\right\} \\
&= \frac{1}{T^{(1)}\left(t_0\right)}\left\{\frac{f_X(t_0^-)+f_X(t_0^+)}{2}\right\}.
\end{aligned}
$$

Then, by (7) and the definition of t_T, the asymptotic variance in (A7) reduces to

$$
\begin{aligned}
&\frac{3\sqrt{\pi}\sqrt{t_T\left(1-t_T\right)}}{4d_T^2}\left\{\frac{f_Y(t_T^-)+f_Y(t_T^+)}{2}\right\} \\
&= \frac{3\sqrt{\pi}\sqrt{T\left(t_0\right)\left\{1-T\left(t_0\right)\right\}}}{4\left\{d_0/T^{(1)}\left(t_0\right)\right\}^2}\frac{1}{T^{(1)}\left(t_0\right)}\left\{\frac{f_X(t_0^-)+f_X(t_0^+)}{2}\right\} \\
&= \frac{3\sqrt{\pi}\sqrt{T\left(t_0\right)\left\{1-T\left(t_0\right)\right\}}}{4d_0^2}T^{(1)}\left(t_0\right)\left\{\frac{f_X(t_0^-)+f_X(t_0^+)}{2}\right\}. \qquad \text{(A9)}
\end{aligned}
$$

Substituting (A8) and (A9) into (A7) establishes the stated result. This completes the proof. ■

Declaration of Interest

The authors report that there are no competing interests to declare.

Data Statement

The datasets used in Section 5 are openly available; see Section 5.1 for more details.

References

[1] Bader, B., J. Yan, and X. Zhang (2018): "Automated Threshold Selection for Extreme Value Analysis via Ordered Goodness-of-Fit Tests with Adjustment for False Discovery Rate," *Annals of Applied Statistics*, 12, 310 - 329.

[2] Chen, S. X. (1999): "Beta Kernel Estimators for Density Functions," *Computational Statistics & Data Analysis*, 31, 131 - 145.

[3] Chen, S. X. (2000): "Probability Density Function Estimation Using Gamma Kernels," *Annals of the Institute of Statistical Mathematics*, 52, 471 - 480.

[4] Chu, C. K., and P. E. Cheng (1996): "Estimation of Jump Points and Jump Values of a Density Function," *Statistica Sinica*, 6, 79 - 95.

[5] Clauset, A., C. R. Shalizi, and M. E. J. Newman (2009): "Power-Law Distributions in Empirical Data," *SIAM Review*, 51, 661 - 703.

[6] Cooray, K., and M. M. A. Ananda (2005): "Modeling Actuarial Data with a Composite Lognormal-Pareto Model," *Scandinavian Actuarial Journal*, 5, 321 - 334.

[7] Couallier, V. (1999): "Estimation Non Paramétrique d'une Discontinuité dans une Densité," *Comptes Rendus de l'Académie des Sciences - Série I*, 329, 633 - 636.

[8] Danielsson, J., L. M. Ergun, L. de Haan, and C. G. de Vries (2019): "Tail Index Estimation: Quantile-Driven Threshold Selection," Bank of Canada Staff Working Paper 2019-28.

[9] Davison, A. C., and R. L. Smith (1990): "Models for Exceedances over High Thresholds," *Journal of the Royal Statistical Society, Series B*, 52, 393 - 442.

[10] Drees, H., A. Janßen, S. I. Resnick, and T. Wang (2020): "On a Minimum Distance Procedure for Threshold Selection in Tail Analysis," *SIAM Journal on Mathematics of Data Science*, 2, 75 - 102.

[11] DuMouchel, W. H. (1983): "Estimating the Stable Index α in Order to Measure Tail Thickness: A Critique," *Annals of Statistics*, 11, 1019 - 1031.

[12] Funke, B., and M. Hirukawa (2025): "Nonparametric Estimation of Splicing Points in Skewed Cost Distributions: A Kernel-Based Approach," *Journal of Nonparametric Statistics*, forthcoming.

[13] Hirukawa, M., I. Murtazashvili, and A. Prokhorov (2022): "Uniform Convergence Rates for Nonparametric Estimators Smoothed by the Beta Kernel," *Scandinavian Journal of Statistics*, 49, 1353 - 1382.

[14] Huh, J. (2002): "Nonparametric Discontinuity Point Estimation in Density or Density Derivatives," *Journal of the Korean Statistical Society*, 31, 261 - 276.

[15] Komatsu, Y. (1955): "Elementary Inequalities for Mills' Ratio," *Reports of Statistical Application Research, Union of Japanese Scientists and Engineers*, 4, 69 - 70.

[16] Klugman, S. A., H. H. Panjer, and G. E. Willmot (2019): *Loss Models: From Data to Decisions,* 5th Edition. Hoboken, NJ: John Wiley & Sons.

[17] Kratz, M. F., and S. I. Resnick (1996): "The QQ-estimator and Heavy Tails, " *Communications in Statistics. Stochastic Models*, 12, 699 - 724.

[18] Loretan, M., and P. C. B. Phillips (1994): "Testing the Covariance Stationarity of Heavy-Tailed Time Series: An Overview of the Theory with Applications to Several Financial Datasets," *Journal of Empirical Finance*, 1, 211 - 248.

[19] MacDonald, A., C. J. Scarrott, D. Lee, B. Darlow, M. Reale, and G. Russell (2011): "A Flexible Extreme Value Mixture Model," *Computational Statistics & Data Analysis*, 55, 2137 - 2157.

[20] Marron, J. S. (1985): "An Asymptotically Efficient Solution to the Bandwidth Problem of Kernel Density Estimation," *Annals of Statistics*, 13, 1011 - 1023.

[21] McNeal, A. J. (1997): "Estimating the Tails of Loss Severity Distributions Using Extreme Value Theory," *ASTIN Bulletin*, 27, 117 - 137.

[22] Moscovich, A., B. Nadler, and C. Spiegelman (2016): "On the Exact Berk-Jones Statistics and Their p-Value Calculation," *Electronic Journal of Statistics*, 10, 2329 - 2354.

[23] Northrop, P. J., and C. L. Coleman (2014): "Improved Threshold Diagnostic Plots for Extreme Value Analyses," *Extremes*, 17, 289 - 303.

[24] Reiss, R.-D., and M. Thomas (2007): *Statistical Analysis of Extreme Values: with Applications to Insurance, Finance, Hydrology and Other Fields.* Basel: Birkhäuser.

[25] Resnick, S. I. (1997): "Discussion of the Danish Data on Large Fire Insurance Losses," *ASTIN Bulletin*, 27, 139 - 151.

[26] Reynkens, T., R. Verbelen, J. Beirlant, and K. Antonio (2017): "Modelling Censored Losses Using Splicing: A Global Fit Strategy with Mixed Erlang and Extreme Value Distributions," *Insurance: Mathematics and Economics*, 77, 65 - 77.

[27] Scarrott, C., and A. MacDonald (2012): "A Review of Extreme Value Threshold Estimation and Uncertainty Quantification," *REVSTAT - Statistical Journal*, 10, 33 - 60.

[28] Scollnik, D. P. M., and C. Sun (2012): "Modeling with Weibull-Pareto Models," *North American Actuarial Journal*, 16, 260 - 272.

[29] Van Es, B. (1991): "Likelihood Cross-Validation Bandwidth Selection for Nonparametric Kernel Density Estimators," *Journal of Nonparametric Statistics*, 1, 83 - 110.

[30] Wadsworth, J. L. (2016): "Exploiting Structure of Maximum Likelihood Estimators for Extreme Value Threshold Selection," *Technometrics*, 58, 116 - 126.

[31] Wadsworth, J. L., and J. A. Tawn (2012): "Likelihood-Based Procedures for Threshold Diagnostics and Uncertainty in Extreme Value Modelling," *Journal of the Royal Statistical Society, Series B*, 74, 543 - 567.

Table 1: Examples of Transformations That Satisfy Assumption 3

	Transformation	Magnification Factor			
		$d_T/d_0 = 1/T^{(1)}(t_0)$	(i)	(ii)	(iii)
Arctangent:	$T_1(x) = \frac{2\arctan(x/t_M)}{\pi}$	$\frac{\pi(t_0^2+t_M^2)}{2t_M}$	12.5664	12.6536	12.6786
Exponential CDF:	$T_2(x) = 1 - \exp\left(-\frac{x\ln 2}{t_M}\right)$	$\frac{t_M 2^{t_0/t_M}}{\ln 2}$	11.5416	12.0218	11.1501
Rational Function:	$T_3(x) = \frac{x}{t_M(1+x/t_M)}$	$\frac{(t_0+t_M)^2}{t_M}$	16.0000	16.0556	16.0714
Hyperbolic Tangent:	$T_4(x) = \frac{\exp\{x\ln 3/(2t_M)\}-\exp\{-x\ln 3/(2t_M)\}}{\exp\{x\ln 3/(2t_M)\}+\exp\{-x\ln 3/(2t_M)\}}$	$\frac{t_M\left\{3^{t_0/(2t_M)}+3^{-t_0/(2t_M)}\right\}^2}{2\ln 3}$	9.7092	10.3055	9.2305

Table 2: Characteristic Numbers of Underlying Distributions

Model	Distribution	Mode	c_L	$f_X\left(t_0^-\right)$	$f_X\left(t_0^+\right)$	d_0
A	Log-Normal + Quadratic	1.4382	0.9498	0.1002	0.0502	0.0500
B	Splicing with Weibull & GPD	1.9252	0.9438	0.0910	0.0375	0.0535
C	Splicing with Weibull & Half-Normal	1.9252	0.9438	0.0910	0.0375	0.0535

Table 3: Monte Carlo Results

		$n=250$					$n=500$				
Estimator	α	Bias	SD	RMSE	$\hat{b}$		Bias	SD	RMSE	$\hat{b}$	
					Model A: Log-Normal + Quadratic						
KS	–	−1.0402	0.5501	1.1767	–	(–)	−0.7721	0.4399	0.8886	–	(–)
Q-MAD	–	−0.2391	0.3979	0.4642	–	(–)	−0.1964	0.3963	0.4423	–	(–)
Q-SUP	–	0.1116	0.6432	0.6528	–	(–)	0.4523	0.8826	0.9918	–	(–)
AEB	–	0.5491	0.5357	0.7671	–	(–)	0.6200	0.3490	0.7115	–	(–)
ADST	–	−1.4147	0.7803	1.6156	–	(–)	−1.2834	0.7266	1.4748	–	(–)
					Case (i): $I_0=[3,5]$						
SG-ML	0.70	−0.5966	0.4652	0.7565	0.0608	(0.0234)	−0.7394	0.4212	0.8510	0.0689	(0.0187)
SG-ML-BC	0.70	−0.5358	0.4520	0.7010	–	(–)	−0.6705	0.4104	0.7862	–	(–)
SB-LS-T$_1$	0.55	−0.5535	0.1371	0.5702	0.0397	(0.0064)	−0.5598	0.0923	0.5674	0.0398	(0.0046)
	0.60	−0.5276	0.1391	0.5456	0.0492	(0.0076)	−0.5319	0.0944	0.5402	0.0494	(0.0052)
	0.65	−0.5020	0.1400	0.5212	0.0590	(0.0086)	−0.5070	0.0951	0.5158	0.0592	(0.0061)
	0.70	−0.4794	0.1397	0.4994	0.0686	(0.0099)	−0.4837	0.0950	0.4929	0.0689	(0.0069)
SB-LS-T$_2$	0.55	−0.5954	0.1178	0.6070	0.0391	(0.0059)	−0.6013	0.0797	0.6066	0.0392	(0.0043)
	0.60	−0.5711	0.1208	0.5838	0.0487	(0.0069)	−0.5777	0.0816	0.5834	0.0487	(0.0049)
	0.65	−0.5489	0.1214	0.5622	0.0584	(0.0081)	−0.5537	0.0827	0.5598	0.0584	(0.0058)
	0.70	−0.5274	0.1227	0.5415	0.0680	(0.0093)	−0.5319	0.0832	0.5384	0.0681	(0.0066)
SB-LS-T$_3$	0.55	−0.4172	0.1671	0.4494	0.0233	(0.0049)	−0.4187	0.1176	0.4349	0.0233	(0.0036)
	0.60	−0.3857	0.1652	0.4196	0.0295	(0.0056)	−0.3859	0.1159	0.4029	0.0295	(0.0041)
	0.65	−0.3565	0.1631	0.3920	0.0358	(0.0064)	−0.3586	0.1153	0.3767	0.0358	(0.0046)
	0.70	−0.3323	0.1612	0.3693	0.0418	(0.0073)	−0.3315	0.1122	0.3499	0.0420	(0.0053)
SB-LS-T$_4$	0.55	−0.6742	0.0953	0.6809	0.0502	(0.0069)	−0.6811	0.0652	0.6842	0.0501	(0.0048)
	0.60	−0.6579	0.0991	0.6653	0.0620	(0.0080)	−0.6642	0.0667	0.6675	0.0620	(0.0057)
	0.65	−0.6411	0.1012	0.6491	0.0742	(0.0094)	−0.6466	0.0695	0.6503	0.0742	(0.0066)
	0.70	−0.6237	0.1037	0.6322	0.0864	(0.0108)	−0.6281	0.0702	0.6320	0.0865	(0.0076)
					Case (ii): $I_0=[3.5,5.5]$						
SG-ML	0.70	−0.2460	0.3369	0.4172	0.0474	(0.0162)	−0.2921	0.3271	0.4385	0.0505	(0.0145)
SG-ML-BC	0.70	−0.1985	0.3302	0.3853	–	(–)	−0.2416	0.3186	0.3998	–	(–)
SB-LS-T$_1$	0.55	−0.3477	0.0981	0.3613	0.0524	(0.0072)	−0.3573	0.0715	0.3643	0.0524	(0.0051)
	0.60	−0.3364	0.1005	0.3511	0.0644	(0.0085)	−0.3442	0.0715	0.3516	0.0644	(0.0059)
	0.65	−0.3227	0.1029	0.3388	0.0768	(0.0099)	−0.3300	0.0740	0.3382	0.0767	(0.0068)
	0.70	−0.3076	0.1059	0.3253	0.0890	(0.0113)	−0.3143	0.0745	0.3230	0.0890	(0.0077)
SB-LS-T$_2$	0.55	−0.3620	0.0929	0.3737	0.0502	(0.0069)	−0.3714	0.0695	0.3779	0.0501	(0.0048)
	0.60	−0.3529	0.0955	0.3656	0.0619	(0.0081)	−0.3613	0.0695	0.3679	0.0618	(0.0056)
	0.65	−0.3405	0.0979	0.3542	0.0739	(0.0095)	−0.3478	0.0707	0.3549	0.0739	(0.0065)
	0.70	−0.3249	0.1006	0.3401	0.0861	(0.0110)	−0.3333	0.0719	0.3409	0.0859	(0.0076)
SB-LS-T$_3$	0.55	−0.2582	0.1220	0.2856	0.0320	(0.0057)	−0.2637	0.0862	0.2774	0.0321	(0.0038)
	0.60	−0.2371	0.1230	0.2671	0.0400	(0.0064)	−0.2428	0.0874	0.2581	0.0400	(0.0045)
	0.65	−0.2127	0.1214	0.2449	0.0482	(0.0075)	−0.2196	0.0856	0.2357	0.0481	(0.0051)
	0.70	−0.1907	0.1224	0.2266	0.0561	(0.0085)	−0.1944	0.0847	0.2121	0.0562	(0.0059)
SB-LS-T$_4$	0.55	−0.3995	0.0854	0.4085	0.0623	(0.0080)	−0.4115	0.0615	0.4161	0.0620	(0.0055)
	0.60	−0.3961	0.0854	0.4052	0.0767	(0.0095)	−0.4090	0.0631	0.4139	0.0763	(0.0065)
	0.65	−0.3910	0.0875	0.4007	0.0914	(0.0112)	−0.4030	0.0651	0.4082	0.0911	(0.0076)
	0.70	−0.3825	0.0910	0.3931	0.1064	(0.0129)	−0.3931	0.0670	0.3987	0.1061	(0.0090)
					Case (iii): $I_0=[2.4,4.5]$						
SG-ML	0.70	−1.1608	0.4941	1.2616	0.1161	(0.1162)	−1.3476	0.3672	1.3967	0.1137	(0.0837)
SG-ML-BC	0.70	−1.0447	0.4801	1.1497	–	(–)	−1.2338	0.3618	1.2858	–	(–)
SB-LS-T$_1$	0.55	−0.6200	0.2251	0.6596	0.0293	(0.0053)	−0.6190	0.1597	0.6393	0.0294	(0.0039)
	0.60	−0.5842	0.2252	0.6261	0.0367	(0.0062)	−0.5821	0.1597	0.6036	0.0369	(0.0046)
	0.65	−0.5533	0.2231	0.5966	0.0443	(0.0073)	−0.5504	0.1577	0.5726	0.0446	(0.0052)
	0.70	−0.5269	0.2211	0.5714	0.0519	(0.0084)	−0.5243	0.1563	0.5471	0.0522	(0.0060)
SB-LS-T$_2$	0.55	−0.7289	0.1877	0.7526	0.0299	(0.0052)	−0.7310	0.1308	0.7426	0.0300	(0.0039)
	0.60	−0.6977	0.1885	0.7227	0.0374	(0.0062)	−0.6977	0.1316	0.7100	0.0376	(0.0044)
	0.65	−0.6683	0.1882	0.6943	0.0452	(0.0072)	−0.6691	0.1328	0.6822	0.0454	(0.0052)
	0.70	−0.6428	0.1877	0.6696	0.0529	(0.0083)	−0.6424	0.1316	0.6558	0.0532	(0.0059)
SB-LS-T$_3$	0.55	−0.4683	0.2400	0.5262	0.0171	(0.0038)	−0.4559	0.1724	0.4874	0.0171	(0.0029)
	0.60	−0.4327	0.2350	0.4923	0.0218	(0.0045)	−0.4206	0.1688	0.4532	0.0219	(0.0032)
	0.65	−0.4043	0.2311	0.4657	0.0267	(0.0051)	−0.3939	0.1657	0.4274	0.0268	(0.0037)
	0.70	−0.3807	0.2269	0.4432	0.0313	(0.0058)	−0.3706	0.1627	0.4047	0.0315	(0.0043)
SB-LS-T$_4$	0.55	−0.8635	0.1527	0.8769	0.0393	(0.0063)	−0.8707	0.1036	0.8768	0.0393	(0.0044)
	0.60	−0.8367	0.1551	0.8510	0.0489	(0.0075)	−0.8430	0.1067	0.8498	0.0490	(0.0052)
	0.65	−0.8114	0.1585	0.8268	0.0588	(0.0087)	−0.8158	0.1088	0.8230	0.0590	(0.0062)
	0.70	−0.7857	0.1590	0.8016	0.0687	(0.0100)	−0.7897	0.1094	0.7973	0.0690	(0.0071)

Table 3: *(continued)*

		$n = 250$					$n = 500$				
Estimator	α	Bias	SD	RMSE	$\hat{b}$		Bias	SD	RMSE	$\hat{b}$	
Model B: Splicing with Weibull & GPD											
KS	–	−1.1510	0.4794	1.2469	–	(–)	−1.0183	0.5832	1.1735	–	(–)
Q-MAD	–	−0.3241	0.3500	0.4771	–	(–)	−0.3154	0.3797	0.4936	–	(–)
Q-SUP	–	0.1315	0.9630	0.9719	–	(–)	0.5670	1.4624	1.5684	–	(–)
AEB	–	1.6324	1.1467	1.9949	–	(–)	1.7231	0.7848	1.8934	–	(–)
ADST	–	−1.0581	0.9258	1.4059	–	(–)	−0.7138	1.0491	1.2689	–	(–)
Case (i): $I_0 = [3, 5]$											
SG-ML	0.70	−0.5461	0.4892	0.7332	0.0484	(0.0202)	−0.5154	0.5629	0.7632	0.0499	(0.0197)
SG-ML-BC	0.70	−0.4976	0.4763	0.6888	–	(–)	−0.4655	0.5473	0.7185	–	(–)
SB-LS-T$_1$	0.55	−0.3760	0.1432	0.4023	0.0198	(0.0067)	−0.3604	0.1079	0.3762	0.0198	(0.0049)
	0.60	−0.3557	0.1378	0.3815	0.0256	(0.0080)	−0.3426	0.1036	0.3579	0.0257	(0.0057)
	0.65	−0.3392	0.1350	0.3651	0.0315	(0.0092)	−0.3268	0.1003	0.3418	0.0318	(0.0066)
	0.70	−0.3245	0.1340	0.3511	0.0373	(0.0105)	−0.3131	0.0990	0.3284	0.0376	(0.0076)
SB-LS-T$_2$	0.55	−0.4262	0.1281	0.4450	0.0207	(0.0064)	−0.4163	0.0960	0.4272	0.0206	(0.0047)
	0.60	−0.4089	0.1257	0.4278	0.0266	(0.0076)	−0.3990	0.0927	0.4097	0.0267	(0.0055)
	0.65	−0.3924	0.1255	0.4120	0.0328	(0.0089)	−0.3839	0.0900	0.3943	0.0330	(0.0064)
	0.70	−0.3794	0.1232	0.3989	0.0387	(0.0101)	−0.3709	0.0895	0.3816	0.0390	(0.0073)
SB-LS-T$_3$	0.55	−0.3114	0.1513	0.3462	0.0110	(0.0039)	−0.2948	0.1133	0.3158	0.0109	(0.0030)
	0.60	−0.2936	0.1431	0.3266	0.0142	(0.0047)	−0.2782	0.1091	0.2989	0.0143	(0.0035)
	0.65	−0.2759	0.1382	0.3086	0.0177	(0.0056)	−0.2623	0.1046	0.2824	0.0178	(0.0041)
	0.70	−0.2594	0.1370	0.2934	0.0210	(0.0065)	−0.2449	0.1007	0.2647	0.0212	(0.0047)
SB-LS-T$_4$	0.55	−0.5007	0.1194	0.5147	0.0294	(0.0080)	−0.4944	0.0861	0.5018	0.0294	(0.0060)
	0.60	−0.4862	0.1190	0.5005	0.0374	(0.0095)	−0.4795	0.0854	0.4871	0.0377	(0.0070)
	0.65	−0.4729	0.1178	0.4874	0.0457	(0.0111)	−0.4682	0.0840	0.4757	0.0459	(0.0080)
	0.70	−0.4606	0.1162	0.4750	0.0541	(0.0126)	−0.4561	0.0829	0.4636	0.0544	(0.0091)
Case (ii): $I_0 = [3.5, 5.5]$											
SG-ML	0.70	−0.2494	0.4820	0.5427	0.0374	(0.0122)	−0.2007	0.5973	0.6301	0.0367	(0.0132)
SG-ML-BC	0.70	−0.2120	0.4737	0.5190	–	(–)	−0.1639	0.5862	0.6087	–	(–)
SB-LS-T$_1$	0.55	−0.2452	0.1090	0.2684	0.0309	(0.0080)	−0.2419	0.0794	0.2546	0.0309	(0.0055)
	0.60	−0.2322	0.1058	0.2551	0.0398	(0.0096)	−0.2306	0.0771	0.2431	0.0397	(0.0066)
	0.65	−0.2181	0.1062	0.2426	0.0488	(0.0112)	−0.2154	0.0767	0.2286	0.0489	(0.0077)
	0.70	−0.2027	0.1057	0.2286	0.0578	(0.0129)	−0.2005	0.0764	0.2146	0.0579	(0.0088)
SB-LS-T$_2$	0.55	−0.2685	0.1021	0.2873	0.0311	(0.0076)	−0.2683	0.0748	0.2786	0.0310	(0.0053)
	0.60	−0.2572	0.1000	0.2760	0.0400	(0.0091)	−0.2564	0.0736	0.2667	0.0400	(0.0064)
	0.65	−0.2445	0.1004	0.2643	0.0490	(0.0108)	−0.2440	0.0718	0.2543	0.0490	(0.0074)
	0.70	−0.2299	0.1005	0.2509	0.0580	(0.0123)	−0.2290	0.0719	0.2400	0.0582	(0.0085)
SB-LS-T$_3$	0.55	−0.1759	0.1137	0.2095	0.0173	(0.0049)	−0.1716	0.0886	0.1931	0.0172	(0.0035)
	0.60	−0.1612	0.1111	0.1958	0.0226	(0.0059)	−0.1543	0.0839	0.1756	0.0226	(0.0042)
	0.65	−0.1404	0.1074	0.1768	0.0281	(0.0070)	−0.1338	0.0812	0.1565	0.0282	(0.0050)
	0.70	−0.1177	0.1082	0.1598	0.0335	(0.0082)	−0.1134	0.0784	0.1379	0.0336	(0.0057)
SB-LS-T$_4$	0.55	−0.3254	0.0954	0.3391	0.0422	(0.0094)	−0.3284	0.0687	0.3355	0.0422	(0.0065)
	0.60	−0.3191	0.0943	0.3327	0.0536	(0.0112)	−0.3202	0.0677	0.3272	0.0538	(0.0077)
	0.65	−0.3105	0.0947	0.3246	0.0655	(0.0132)	−0.3119	0.0680	0.3193	0.0656	(0.0091)
	0.70	−0.2999	0.0968	0.3152	0.0774	(0.0153)	−0.3017	0.0683	0.3094	0.0775	(0.0105)
Case (iii): $I_0 = [2.4, 4.5]$											
SG-ML	0.70	−0.7643	0.4375	0.8806	0.0982	(0.1093)	−0.7253	0.4562	0.8568	0.0878	(0.0745)
SG-ML-BC	0.70	−0.6661	0.4182	0.7865	–	(–)	−0.6375	0.4320	0.7701	–	(–)
SB-LS-T$_1$	0.55	−0.3657	0.2112	0.4223	0.0155	(0.0056)	−0.3459	0.1487	0.3766	0.0155	(0.0042)
	0.60	−0.3474	0.1995	0.4006	0.0199	(0.0067)	−0.3288	0.1439	0.3589	0.0200	(0.0050)
	0.65	−0.3343	0.1976	0.3883	0.0242	(0.0079)	−0.3134	0.1404	0.3433	0.0245	(0.0056)
	0.70	−0.3208	0.1942	0.3750	0.0285	(0.0089)	−0.2983	0.1375	0.3285	0.0291	(0.0063)
SB-LS-T$_2$	0.55	−0.4599	0.1915	0.4982	0.0161	(0.0058)	−0.4396	0.1350	0.4599	0.0161	(0.0043)
	0.60	−0.4425	0.1862	0.4801	0.0207	(0.0069)	−0.4220	0.1320	0.4422	0.0208	(0.0050)
	0.65	−0.4266	0.1823	0.4639	0.0254	(0.0081)	−0.4065	0.1298	0.4267	0.0257	(0.0059)
	0.70	−0.4132	0.1800	0.4507	0.0300	(0.0092)	−0.3924	0.1265	0.4123	0.0305	(0.0065)
SB-LS-T$_3$	0.55	−0.3197	0.2048	0.3796	0.0088	(0.0033)	−0.3064	0.1493	0.3408	0.0085	(0.0027)
	0.60	−0.3119	0.1977	0.3693	0.0110	(0.0040)	−0.2940	0.1450	0.3278	0.0111	(0.0032)
	0.65	−0.2998	0.1974	0.3590	0.0136	(0.0048)	−0.2802	0.1385	0.3126	0.0137	(0.0035)
	0.70	−0.2868	0.1918	0.3451	0.0160	(0.0056)	−0.2661	0.1380	0.2998	0.0163	(0.0040)
SB-LS-T$_4$	0.55	−0.5438	0.1827	0.5737	0.0232	(0.0077)	−0.5244	0.1280	0.5398	0.0232	(0.0057)
	0.60	−0.5251	0.1774	0.5543	0.0294	(0.0091)	−0.5057	0.1262	0.5212	0.0299	(0.0065)
	0.65	−0.5084	0.1733	0.5372	0.0361	(0.0104)	−0.4907	0.1242	0.5062	0.0365	(0.0074)
	0.70	−0.4933	0.1712	0.5221	0.0427	(0.0118)	−0.4768	0.1229	0.4924	0.0432	(0.0084)

Table 3: *(continued)*

Estimator	α	$n = 250$ Bias	SD	RMSE	$\hat{b}$		$n = 500$ Bias	SD	RMSE	$\hat{b}$	
Model C: Splicing with Weibull & Half-Normal											
KS	–	−0.9789	0.3837	1.0514	–	(–)	−0.8630	0.2805	0.9074	–	(–)
Q-MAD	–	−0.3250	0.3040	0.4450	–	(–)	−0.3429	0.3064	0.4599	–	(–)
Q-SUP	–	0.3436	0.6683	0.7514	–	(–)	0.8676	0.6395	1.0779	–	(–)
AEB	–	0.5004	0.4249	0.6565	–	(–)	0.5785	0.2777	0.6417	–	(–)
ADST	–	−1.4826	0.5377	1.5771	–	(–)	−1.3992	0.6323	1.5354	–	(–)
Case (i): $I_0 = [3, 5]$											
SG-ML	0.70	−0.6078	0.4844	0.7772	0.0562	(0.0243)	−0.5700	0.5669	0.8039	0.0577	(0.0234)
SG-ML-BC	0.70	−0.5515	0.4694	0.7242	–	(–)	−0.5123	0.5484	0.7504	–	(–)
SB-LS-T$_1$	0.55	−0.3317	0.1625	0.3694	0.0198	(0.0060)	−0.3158	0.1220	0.3385	0.0197	(0.0043)
	0.60	−0.3114	0.1540	0.3474	0.0257	(0.0071)	−0.2998	0.1148	0.3211	0.0256	(0.0052)
	0.65	−0.2942	0.1508	0.3306	0.0316	(0.0084)	−0.2827	0.1124	0.3042	0.0316	(0.0060)
	0.70	−0.2762	0.1494	0.3140	0.0374	(0.0098)	−0.2651	0.1091	0.2867	0.0376	(0.0069)
SB-LS-T$_2$	0.55	−0.3981	0.1409	0.4223	0.0212	(0.0058)	−0.3889	0.1050	0.4028	0.0211	(0.0043)
	0.60	−0.3805	0.1373	0.4045	0.0272	(0.0070)	−0.3719	0.0997	0.3851	0.0272	(0.0051)
	0.65	−0.3614	0.1343	0.3856	0.0335	(0.0082)	−0.3545	0.0974	0.3676	0.0336	(0.0059)
	0.70	−0.3442	0.1321	0.3687	0.0397	(0.0095)	−0.3379	0.0964	0.3514	0.0398	(0.0068)
SB-LS-T$_3$	0.55	−0.2661	0.1740	0.3180	0.0111	(0.0036)	−0.2531	0.1317	0.2853	0.0109	(0.0027)
	0.60	−0.2518	0.1625	0.2997	0.0143	(0.0043)	−0.2346	0.1212	0.2640	0.0144	(0.0032)
	0.65	−0.2302	0.1566	0.2784	0.0179	(0.0052)	−0.2169	0.1167	0.2463	0.0179	(0.0037)
	0.70	−0.2086	0.1554	0.2602	0.0214	(0.0061)	−0.1956	0.1136	0.2262	0.0214	(0.0043)
SB-LS-T$_4$	0.55	−0.4785	0.1279	0.4953	0.0302	(0.0075)	−0.4743	0.0919	0.4831	0.0302	(0.0054)
	0.60	−0.4637	0.1255	0.4804	0.0384	(0.0089)	−0.4594	0.0899	0.4682	0.0384	(0.0063)
	0.65	−0.4484	0.1242	0.4653	0.0469	(0.0102)	−0.4444	0.0877	0.4530	0.0470	(0.0074)
	0.70	−0.4343	0.1233	0.4514	0.0552	(0.0119)	−0.4305	0.0882	0.4394	0.0555	(0.0085)
Case (ii): $I_0 = [3.5, 5.5]$											
SG-ML	0.70	−0.2616	0.5004	0.5647	0.0418	(0.0159)	−0.2291	0.6028	0.6448	0.0426	(0.0167)
SG-ML-BC	0.70	−0.2197	0.4905	0.5375	–	(–)	−0.1865	0.5895	0.6183	–	(–)
SB-LS-T$_1$	0.55	−0.1578	0.1200	0.1982	0.0328	(0.0085)	−0.1573	0.0860	0.1792	0.0325	(0.0061)
	0.60	−0.1422	0.1171	0.1842	0.0424	(0.0101)	−0.1443	0.0824	0.1662	0.0420	(0.0070)
	0.65	−0.1270	0.1164	0.1723	0.0520	(0.0118)	−0.1269	0.0817	0.1509	0.0518	(0.0083)
	0.70	−0.1089	0.1162	0.1593	0.0615	(0.0136)	−0.1084	0.0819	0.1358	0.0615	(0.0095)
SB-LS-T$_2$	0.55	−0.1902	0.1151	0.2223	0.0331	(0.0079)	−0.1953	0.0816	0.2117	0.0327	(0.0056)
	0.60	−0.1781	0.1119	0.2103	0.0426	(0.0095)	−0.1818	0.0785	0.1980	0.0423	(0.0066)
	0.65	−0.1627	0.1120	0.1975	0.0523	(0.0112)	−0.1639	0.0775	0.1813	0.0521	(0.0078)
	0.70	−0.1440	0.1116	0.1822	0.0620	(0.0128)	−0.1457	0.0772	0.1649	0.0618	(0.0090)
SB-LS-T$_3$	0.55	−0.0846	0.1269	0.1525	0.0183	(0.0050)	−0.0833	0.0948	0.1262	0.0180	(0.0037)
	0.60	−0.0696	0.1222	0.1406	0.0240	(0.0061)	−0.0664	0.0873	0.1096	0.0238	(0.0044)
	0.65	−0.0473	0.1216	0.1305	0.0298	(0.0073)	−0.0433	0.0854	0.0958	0.0298	(0.0052)
	0.70	−0.0219	0.1205	0.1224	0.0356	(0.0085)	−0.0198	0.0860	0.0882	0.0355	(0.0060)
SB-LS-T$_4$	0.55	−0.2515	0.1087	0.2740	0.0452	(0.0098)	−0.2578	0.0767	0.2690	0.0449	(0.0068)
	0.60	−0.2440	0.1079	0.2668	0.0575	(0.0118)	−0.2503	0.0736	0.2609	0.0571	(0.0082)
	0.65	−0.2332	0.1078	0.2569	0.0701	(0.0138)	−0.2376	0.0745	0.2490	0.0698	(0.0097)
	0.70	−0.2198	0.1086	0.2451	0.0828	(0.0160)	−0.2246	0.0749	0.2368	0.0825	(0.0111)
Case (iii): $I_0 = [2.4, 4.5]$											
SG-ML	0.70	−0.8056	0.4221	0.9095	0.1275	(0.1394)	−0.7564	0.4502	0.8803	0.1080	(0.1005)
SG-ML-BC	0.70	−0.6781	0.4078	0.7913	–	(–)	−0.6484	0.4241	0.7748	–	(–)
SB-LS-T$_1$	0.55	−0.3191	0.2451	0.4024	0.0158	(0.0050)	−0.3003	0.1778	0.3490	0.0158	(0.0038)
	0.60	−0.3041	0.2323	0.3827	0.0201	(0.0059)	−0.2843	0.1668	0.3296	0.0204	(0.0043)
	0.65	−0.2905	0.2270	0.3686	0.0245	(0.0070)	−0.2676	0.1613	0.3124	0.0250	(0.0049)
	0.70	−0.2738	0.2234	0.3534	0.0288	(0.0080)	−0.2502	0.1588	0.2963	0.0295	(0.0056)
SB-LS-T$_2$	0.55	−0.4437	0.2115	0.4915	0.0169	(0.0054)	−0.4269	0.1517	0.4530	0.0170	(0.0041)
	0.60	−0.4260	0.2049	0.4728	0.0217	(0.0064)	−0.4080	0.1455	0.4331	0.0218	(0.0047)
	0.65	−0.4079	0.1996	0.4541	0.0264	(0.0074)	−0.3895	0.1422	0.4146	0.0268	(0.0053)
	0.70	−0.3916	0.1986	0.4391	0.0311	(0.0085)	−0.3729	0.1405	0.3985	0.0316	(0.0061)
SB-LS-T$_3$	0.55	−0.2820	0.2412	0.3711	0.0089	(0.0030)	−0.2646	0.1777	0.3187	0.0090	(0.0024)
	0.60	−0.2751	0.2299	0.3585	0.0114	(0.0037)	−0.2581	0.1684	0.3082	0.0115	(0.0029)
	0.65	−0.2591	0.2261	0.3439	0.0141	(0.0043)	−0.2391	0.1602	0.2878	0.0143	(0.0032)
	0.70	−0.2439	0.2214	0.3294	0.0165	(0.0050)	−0.2234	0.1601	0.2748	0.0169	(0.0037)
SB-LS-T$_4$	0.55	−0.5382	0.1971	0.5732	0.0242	(0.0072)	−0.5201	0.1395	0.5385	0.0245	(0.0054)
	0.60	−0.5173	0.1915	0.5516	0.0308	(0.0084)	−0.5002	0.1359	0.5184	0.0311	(0.0061)
	0.65	−0.4981	0.1878	0.5323	0.0376	(0.0097)	−0.4827	0.1340	0.5009	0.0380	(0.0070)
	0.70	−0.4807	0.1849	0.5150	0.0441	(0.0110)	−0.4657	0.1324	0.4841	0.0447	(0.0078)

Table 4: Descriptive Statistics of Datasets

Data	n	Mean	SD	SK	Min.	Q1	Q2	Q3	90%	95%	99%	Max.
(A) Danish Fire Insurance Losses (in millions of Danish kroner)												
Original	2,492	3.063	7.975	19.884	0.313	1.157	1.634	2.646	5.080	8.406	24.614	263.250
(B) Norwegian Fire Insurance Losses (in millions of Norwegian kroner)												
Original	9,181	5.235	18.277	26.287	0.741	1.526	2.354	4.096	8.762	15.414	50.783	881.448
Downsized	3,064	4.861	11.169	10.359	0.741	1.550	2.391	4.090	8.453	14.546	51.604	257.319
(C) Belgian Motor Insurance Losses (in thousands of Euros)												
Original	18,276	1.448	3.875	11.073	0.000	0.145	0.572	1.441	3.022	4.158	18.344	140.032
Downsized	3,151	1.451	3.833	9.009	0.001	0.145	0.529	1.426	3.021	3.957	20.451	81.946
(D) French Motor Insurance Losses (in thousands of French francs)												
Original	26,444	2.266	29.371	109.524	0.001	0.686	1.172	1.212	2.768	4.766	16.510	4,075.401
Downsized	2,638	2.608	27.168	42.476	0.001	0.739	1.172	1.304	2.912	5.369	18.551	1,301.173

Note: n = sample size; Mean = average; SD = standard deviation; SK = skewness; Min. = minimum value; Q1 = first quartile; Q2 = median (i.e., second quartile); Q3 = third quartile; 90% = 90% quantile; 95% = 95% quantile; 99% = 99% quantile; and Max. = maximum value.

Table 5: Results for Real Data Examples

Data	Estimator	Estimate of t_0	Estimator	α	I_0	Estimate of t_0	$\hat{b}$
(A) Danish Fire Insurance Losses (in millions of Danish kroner)							
Original	KS	1.375	SG-ML	0.70	$[1, 30]$	1.861	0.235
	Q-MAD	29.037	SG-ML-BC	–	–	2.096	–
	Q-SUP	11.123					
	AEB	25.288	SB-LS-T_3	0.70	$[1, 30]$	1.808	0.005
	ADST	1.406					
(B) Norwegian Fire Insurance Losses (in millions of Norwegian kroner)							
Downsized	KS	2.221	SG-ML	0.70	$[2, 50]$	2.756	0.350
	Q-MAD	35.794	SG-ML-BC	–	–	2.924	–
	Q-SUP	123.274					
	AEB	47.528	SB-LS-T_3	0.70	$[2, 50]$	2.702	0.005
	ADST	2.121					
(C) Belgian Motor Insurance Losses (in thousands of Euros)							
Downsized	KS	3.233	SG-ML	0.70	$[2, 40]$	$40.000^{\ddagger}$	0.005
	Q-MAD	3.022	SG-ML-BC	–	–	40.005	–
	Q-SUP	40.035					
	AEB	18.592	SB-LS-T_3	0.70	$[2, 40]$	2.435	0.060
	ADST	$29.451^{\dagger}$					
(D) French Motor Insurance Losses (in thousands of French francs)							
Downsized	KS	1.318	SG-ML	0.70	$[1, 20]$	$20.000^{\ddagger}$	0.005
	Q-MAD	3.204	SG-ML-BC	–	–	20.005	–
	Q-SUP	7.000					
	AEB	18.900	SB-LS-T_3	0.70	$[1, 20]$	11.247	0.005
	ADST	$37.149^{\dagger}$					

Note: Superscripts "†" on ADST indicate that the 99.5% empirical percentile is chosen, whereas superscripts "‡" on SG-ML mean that an estimation failure occurs.