ISSN: 2658-8455
Volume 7, Issue 09 (2026), pp. 361-375.

When Learning Reproduces the Quantile: A Theoretical Analysis of Logistic-Regression Threshold Selection and the Role of Bounded Truncation in Extreme Insurance Loss Modelling

Amina NADIM, ***(PhD candidate)***
Laboratory of Applied Modelling in Economics and Management
Faculty of Law, Economics, and Social Sciences Ain Sebaâ
Hassan II University of Casablanca, Morocco.

Tarek ZARI, ***(Full Professor)***
Laboratory of Applied Modelling in Economics and Management
Faculty of Law, Economics, and Social Sciences Ain Sebaâ,
Hassan II University of Casablanca, Morocco.

Correspondence address :	Laboratory of Applied Modelling in Economics and Management Faculty of Law, Economics, and Social Sciences Ain Sebaâ, Hassan II University of Casablanca, Morocco.
Disclosure Statement :	The authors declare that they have not received any financial support that could have influenced the objectivity of this study. They take full responsibility for any potential plagiarism, the use of artificial intelligence in the writing process, as well as for the results presented in this article.
Conflict of Interest :	The authors report no conflicts of interest.
Cite this article :	NADIM, A., & ZARI, T. (2026). When Learning Reproduces the Quantile: A Theoretical Analysis of Logistic-Regression Threshold Selection and the Role of Bounded Truncation in Extreme Insurance Loss Modelling. *International Journal of Accounting, Finance, Auditing, Management and Economics*, *7*(9), 361–375. https://doi.org/10.5281/zenodo.21709754
License	**This is an open access article under the CC BY-NC-ND license**

Received: 25/06/2026 *Accepted: 03/08/2026*

When Learning Reproduces the Quantile: A Theoretical Analysis of Logistic-Regression Threshold Selection and the Role of Bounded Truncation in Extreme Insurance Loss Modelling

Abstract:

The choice of the threshold above which a loss is treated as extreme remains the central open problem in applying peaks-over-threshold methods to non-life insurance, governing a bias-variance trade-off with direct consequences for reserves, premium loading, and solvency capital. Since the foundational results of Pickands (1975), Balkema and de Haan (1974), and Hill (1975), threshold selection has relied mainly on graphical diagnostics and bias-correction methods, none of which resolve the choice of threshold itself; recent work has proposed automating this choice through supervised classifiers, most notably logistic regression, on the claim that this yields an objective, reproducible threshold. This paper makes two contributions that reposition that claim. First, working within the standard semi-parametric framework of extreme value theory, we establish the paper's central theoretical result (Proposition 1): when the sole covariate is the loss amount and labels are generated by an empirical quantile rule, the threshold returned by logistic-regression selection converges asymptotically, under stated regularity conditions, to the labelling quantile itself, so that the procedure recovers, rather than discovers, a threshold fixed implicitly at the labelling stage. Second, we extend Hill's (1975) estimator, via the bounded-domain correction of Nuyts (2010), to contractually right-truncated insurance losses, a structural feature of insurance data left unaddressed by prevailing threshold-selection methods. The analysis culminates in a unified decision framework situating heuristic, statistical, graphical, and learning-based methods within a common bias-variance logic. The contribution is conceptual and theoretical.

Keywords: Extreme value theory; threshold selection; peaks over threshold; Hill estimator; logistic regression; generalized Pareto distribution; non-life insurance; Value-at-Risk; Expected Shortfall.
Classification JEL: C13; C24; G22
Paper type: Theoretical Research

ISSN: 2658-8455
Volume 7, Issue 09 (2026), pp. 361-375.

1. Introduction :

In non-life insurance, a small number of large claims frequently accounts for a disproportionate share of a portfolio's total cost. The behaviour of these few observations dominates the quantities that matter most to an insurer: the technical reserve, the loading of premiums against adverse deviation, and the structure of excess-of-loss reinsurance treaties. Because the events of interest occur in the extreme right tail of the loss distribution, ordinary descriptive statistics offer little guidance; what is needed is a theoretically grounded description of tail behaviour that does not depend on the bulk of the data.

Two complementary extreme value theory (EVT) constructions are available: the block-maxima approach and the peaks-over-threshold (POT) approach. In insurance loss data, POT is generally preferred because it uses the data more efficiently. The difficulty is intrinsic and is usually framed as a bias-variance trade-off: a threshold set too low admits observations outside the asymptotic regime, while a threshold set too high leaves too few exceedances for reliable inference. The optimal threshold lies between these extremes, but its location depends on the unknown underlying distribution.

Faced with this sensitivity, practitioners have historically relied on graphical diagnostics, selecting a threshold by visual inspection of a region of approximate stability. Such diagnostics are informative but depend heavily on expert judgement, a genuine weakness in an actuarial setting. Recent research has therefore sought to automate threshold selection through supervised machine learning, most prominently logistic regression trained on quantile-based extremeness labels.

This paper subjects that proposal to theoretical scrutiny. Two structural conditions underlie the problem we identify. First, in the standard implementation of this approach, the classifier is trained on a single covariate, the loss amount itself, which is monotonically related to the label by construction. Second, the label is not observed independently of the covariate: it is assigned by thresholding the same loss amount at an empirical quantile. Under these two conditions jointly, univariate feature and quantile-based labelling, the two classes become perfectly separable along the very dimension the classifier is meant to search over, which is what drives the logistic-regression boundary back toward the labelling rule itself rather than toward any independent signal in the data. This mechanism has gone unremarked in the literature proposing such automated procedures, which evaluates their reproducibility without examining the labelling design that produces it.

We formalise this problem as two research hypotheses. **H1**: under a univariate loss-amount covariate and empirical-quantile labelling, logistic-regression threshold selection does not identify a threshold independent of the labelling process; asymptotically, it recovers the very quantile used to construct its own training labels (formalised as Proposition 1, Section 6). **H2**: the classical Hill estimator is structurally biased for insurance losses subject to contractual policy limits, and a bounded-domain correction in the spirit of Nuyts (2010) restores its validity under right truncation (Section 5).

Our second contribution addresses a structural feature of insurance data that prevailing selection methods ignore: real policies carry limits and caps, so losses are often bounded above by contractual sums insured. Data therefore live on a finite interval, bounded below, where the asymptotic regime begins, and above, where truncation bites. The standard Hill estimator accounts only for the lower bound and misbehaves under right truncation. Nuyts (2010) proposed a correction that treats both bounds symmetrically and remains well behaved under truncation, yet it has circulated almost entirely outside the actuarial literature, published in a general mathematics journal rather than in the applied EVT or actuarial outlets where right-truncated data are routinely encountered. We integrate it into the POT pipeline and argue that it is the natural remedy for capped insurance losses.

Together, these two contributions carry a single methodological message: threshold selection cannot be reduced to visual habit, nor to the mechanical application of a classifier whose answer is foreordained by its labels. A defensible procedure must engage with the architecture of the data, its tail heaviness, truncation, and sample size, within the bias-variance logic underlying all competing methods. We close by assembling the heuristic, statistical, graphical, and learning-based approaches into a unified framework.

The paper proceeds in four thematic blocks. Sections 2-3 review the literature and position the contribution. Sections 4-6 develop the theoretical framework, spanning the POT methodology, the truncation problem, and the formal analysis of logistic-regression threshold selection. Section 7 presents the unified decision framework, and Sections 8-11 discuss practical implications, limitations, and conclude.

2. Literature Review :

The literature relevant to this paper can be organised into four currents that, although they developed in dialogue, address distinct aspects of the same underlying problem. We treat each critically rather than descriptively, drawing out the limitations that motivate the next, and close the section with a synoptic comparison (Table 1).

2.1. Tail-index estimation and its pathologies

The semi-parametric estimation of a heavy tail begins with Hill (1975), who proposed estimating the tail index of a positive heavy-tailed distribution from the logarithms of the largest order statistics. The estimator is consistent across the whole max-domain of attraction of a Fréchet-type law, is trivial to compute, and performs well when the underlying distribution follows exactly a Pareto law. Its defect is equally well known: the estimate relies on the number k of upper order statistics retained, and the choice of k is itself a threshold-selection problem in disguise. Small k yields high variance; large k admits non-tail observations and induces bias; the Hill plot frequently fails to exhibit a stable region (Resnick and Stărică, 1997; Drees, de Haan and Resnick, 2000).

Beirlant et al. (2004) survey this family of semi-parametric estimators comprehensively, situating the Hill estimator alongside moment-based and probability-weighted-moment alternatives, and formalise the bias-variance trade-off underlying k-selection in terms of second-order regular variation. Their treatment makes explicit a point that recurs throughout this review: virtually every refinement of the Hill estimator, however constructed, ultimately reduces the threshold problem to the estimation of an auxiliary second-order parameter, substituting one unknown for another rather than eliminating the underlying indeterminacy.

Two broad responses to the Hill estimator's instability emerged. The first smooths or averages the estimator to dampen its volatility, either by integrating it over a moving window of values of k (Resnick and Stărică, 1997) or by kernel-weighting the order statistics (Csörgő, Deheuvels and Mason, 1985). The second attacks the bias directly, yielding the reduced-bias estimators that now represent the state of the art (Caeiro, Gomes and Pestana, 2005; Gomes and Martins, 2002; Gomes and Guillou, 2014). Both responses presuppose that the only obstacle to stability is statistical noise or asymptotic bias. Neither addresses the possibility that the data are confined to a finite interval, which is the situation Nuyts (2010) singles out.

Nuyts (2010) occupies a distinctive position in this current. Rather than smoothing the Hill plot or correcting its asymptotic bias, he reconsiders the construction of the estimator itself. The classical Hill statistic uses the smallest retained order statistic as a lower bound and implicitly sends the upper bound to infinity. Nuyts observes that in many empirical settings the data are bounded above, by physical limits, administrative cut-offs, or, in the case that concerns us, contractual caps and policy limits, and derives an estimator incorporating a finite upper bound

ISSN: 2658-8455
Volume 7, Issue 09 (2026), pp. 361-375.

IJAFAME
International Journal of Accounting, Finance, Auditing, Management & Economics

R symmetrically with the lower bound L. He demonstrates that the modified estimator matches the classical one when the tail is a pure power law extending far to the right, but markedly outperforms it when the distribution decreases slowly or the data are truncated. Despite this, the correction has found essentially no uptake in the actuarial literature specifically: a search of the three leading actuarial venues in which right-truncated severity data are routinely encountered, the ASTIN Bulletin, the Scandinavian Actuarial Journal, and Insurance: Mathematics and Economics, returns no article applying or extending the Nuyts correction to capped or policy-limited losses, despite each journal publishing extensively on Hill-type tail estimation and threshold selection more broadly (see, e.g., Scollnik, 2007, in the Scandinavian Actuarial Journal, and Wang, Hobæk Haff and Huseby, 2020, in Insurance: Mathematics and Economics, neither of which references the bounded-domain correction). The contribution has instead remained confined to the general mathematics venue in which it was originally published.

2.2. Automated threshold selection

The second current aims to provide an algorithmic, judgement-free choice of k or u. Caeiro and Gomes (2014) provide the authoritative survey. They organise the field around the asymptotic mean squared error (AMSE) of the Hill estimator, whose minimiser k0 can be expressed in terms of second-order parameters and estimated in turn, following the programme initiated by Hall (1982) and Hall and Welsh (1985). They assess bias-based diagnostics (Guillou and Hall, 2001; Drees and Kaufmann, 1998), single- and double-bootstrap approaches for estimating k0 without explicit second-order modelling (Hall, 1990; Draisma et al., 1999; Danielsson et al., 2001; Gomes and Oliveira, 2001), and sample-path-stability heuristics. Their central conclusion is that no single method dominates: relative performance depends on sample size, tail heaviness, and the estimator in use.

Wadsworth (2016) approaches the same problem from a different angle, exploiting the structure of the maximum-likelihood estimator of the generalised Pareto distribution rather than the Hill statistic. The proposal formalises threshold choice as a model-selection exercise conducted directly on the likelihood surface across candidate thresholds, replacing visual inspection of estimator stability with a testable, likelihood-based stopping rule. This constitutes a genuine advance in reproducibility over graphical diagnostics, since the decision rule is computable and repeatable rather than read off a plot. It nonetheless inherits a limitation common to the entire AMSE-based and likelihood-based family: the procedure characterises where the estimator's own sampling behaviour stabilises, not where the underlying data-generating process changes regime, so that a well-behaved likelihood surface is a necessary but not sufficient condition for having identified a threshold that is meaningful in the applied sense pursued in Section 4 of this paper.

Wang, Hobæk Haff and Huseby (2020) translate this comparative question into the actuarial target that ultimately matters: the reserve. They embed threshold selection within composite models, in which a flexible bulk distribution (gamma, log-normal, Weibull, or log-gamma) is joined to a Pareto-type tail above the threshold, and evaluate selection methods by the bias and root mean squared error of the resulting reserve rather than by the fit of the tail alone. Their simulation study finds that, when data are plentiful, the square-root rule performs best overall, with the exponentiality test close behind; when data are scarce, simultaneous estimation of the threshold and the model parameters is preferable. The practical lesson is that the merit of a selection method cannot be judged in isolation from the downstream quantity it is meant to serve, a lesson we carry into our unified framework in Section 4.3.

2.3. Graphical diagnostics

The third current comprises the graphical tools that remain the default in applied work. Alaswed

and Mohamed (2019) compare the mean residual life plot (MRLP), in which approximate linearity above a threshold signals GPD behaviour, with the threshold-choice plot (TCP), in which one seeks the level above which the GPD parameter estimates remain stable. Applying both to Danish fire-insurance losses and arbitrating by deviance and the Akaike information criterion, they find the TCP threshold to be the better-supported decision. The study demonstrates that graphical diagnostics can be disciplined by goodness-of-fit testing, but it also lays bare the limitation common to the whole current: the plots must be read, and reading them is a matter of judgement that resists full reproducibility.

2.4. Learning-based threshold selection

The fourth current applies supervised learning to the selection challenge. Iroko, Tukur and Adeyanju (2025) propose labelling each claim as extreme or non-extreme according to an initial empirical quantile, fitting a logistic regression predicting the probability of extremeness, and defining the operational threshold as the smallest claim whose predicted probability exceeds a fixed cut-off. They report that the final threshold agrees with the linear region of the MRLP, interpreted as external validation.

This proposal is the immediate occasion for the present paper. We accept that the approach is reproducible and integrates conveniently into a pipeline. We dispute, however, that it constitutes an objective discovery of the threshold: the classifier is given a single monotone feature and trained on labels that are themselves a deterministic monotone function of that feature, so the fitted probability is necessarily increasing in the loss, and the level at which it crosses any fixed cut-off is pinned to a quantile close to the one used for labelling. Section 4.2 turns this observation into a precise statement.

Beyond the univariate case, the wider machine-learning literature on extremes escapes this critique, but the conditions under which it does so are worth making explicit rather than merely asserted, since they delimit precisely where the present paper's argument stops applying. High-dimensional extreme quantile regression (Tang, Wang and Li, 2024) bring genuinely new covariate information to bear on the classification, information not deterministically reducible to the loss amount itself, such as claim cause, exposure characteristics, or policyholder attributes, so that the fitted decision surface reflects structural features of the risk rather than merely reconstructing the label from the variable that generated it. The epistemological condition for supervised learning to contribute genuine discovery in threshold selection is therefore not the sophistication of the classifier but the informational content of its covariates relative to its labels: a highly flexible model trained on the same single monotone feature used to construct the labels remains exposed to the circularity formalised in Section 4.2, regardless of its capacity, whereas even a simple model trained on independent structural covariates can, in principle, recover information the univariate quantile rule cannot. This distinction between model complexity and covariate informativeness is what motivates situating learning-based methods, alongside heuristic, AMSE-based, and graphical approaches, within the unified decision framework of Section 4.3, rather than treating automation as a self-sufficient remedy for the limitations identified in the first three currents.

ISSN: 2658-8455
Volume 7, Issue 09 (2026), pp. 361-375.

Table 1. Synoptic comparison of threshold-selection approaches

Approach	Selection criterion	Implicit assumptions	Advantages	Limitations	Suited data type
Hill-type / reduced-bias (Hill, 1975; Caeiro, Gomes & Pestana, 2005)	Minimise AMSE of tail-index estimate over k	Regular variation; second-order condition estimable	Simple, well studied, asymptotically efficient	k-selection circular; unstable in small samples	Unbounded, moderately heavy-tailed data
Bounded-domain correction (Nuyts, 2010)	Symmetric lower/upper bound estimating equation	Data confined to finite interval [L, R]	Robust under truncation and slow variation	Almost unused in actuarial practice; R must be identified	Contractually capped losses
AMSE / bootstrap methods (Caeiro & Gomes, 2014)	Minimise estimated MSE via bootstrap or second-order fit	Second-order regular variation	Judgement-free, reproducible	Sensitive to sample size and tail heaviness; no dominant method	Large, homogeneous samples
Likelihood-based (Wadsworth, 2016)	Stability of GPD likelihood surface across candidate thresholds	Correct GPD specification above threshold	Testable, reproducible decision rule	Characterises estimator behaviour, not regime change	Data with well-identified GPD tail
Graphical diagnostics (Alaswed & Mohamed, 2019)	Visual stability (MRLP, TCP)	Approximate linearity above threshold	Intuitive, widely used, can be disciplined by AIC/deviance	Not fully reproducible; judgement-dependent	Exploratory analysis, moderate samples
Univariate logistic-regression selection (Iroko, Tukur & Adeyanju, 2025)	Predicted probability of extremeness crosses cut-off	Single monotone covariate; quantile-based labels	Reproducible, automatable	Circular: recovers labelling quantile (this paper, Section 4.2)	Not recommended as sole covariate device

Source: authors' compilation, this paper.

3. Research methodology :

We collect here the constructions used in the rest of the paper and fix notation.

3.1. The peaks-over-threshold construction

Let X be a positive random variable representing an individual loss, with distribution function F and right endpoint $x^{\mathrm{F}} = \sup\{x : F(x) < 1\}$. For a threshold $u < x^{\mathrm{F}}$, the distribution of the excess $X - u$ given $X > u$ is

$$F_{\mathrm{u}}(y) = \mathrm{P}(X - u \le y \mid X > u), \quad y \ge 0. \qquad (1)$$

The Pickands–Balkema–de Haan theorem (Balkema and de Haan, 1974; Pickands, 1975) states that, for a wide class of distributions in the max-domain of attraction of an extreme value law, there exists a positive function $\sigma(u)$ such that

$$\lim_{u \to x^{\mathrm{f}}} \sup_{0 \le y < x^{\mathrm{f}} - u} | F_{\mathrm{u}}(y) - G_{\xi,\sigma(u)}(y) | = 0, \qquad (2)$$

where $G_{\xi,\sigma}$ is the generalized Pareto distribution. The theorem is the formal justification for fitting a GPD to exceedances above a sufficiently high threshold, and it makes transparent why the choice of u is unavoidable: the approximation is exact only in the limit, so u must be high enough for the bias to be small yet low enough for enough exceedances to remain.

3.2. The generalized Pareto distribution

The GPD has distribution function

$$G_{\xi,\sigma}(y) = 1 - (1 + \xi\, y/\sigma)^{-1/\xi} \quad \text{for } \xi \neq 0, \quad \text{and} \quad G_{0,\sigma}(y) = 1 - \exp(-y/\sigma) \quad \text{for } \xi = 0, \qquad (3)$$

with scale $\sigma > 0$ and shape $\xi \in \mathbb{R}$. The support is $y \ge 0$ when $\xi \ge 0$ and $0 \le y \le -\sigma/\xi$ when $\xi < 0$. The shape parameter governs tail heaviness: $\xi > 0$ gives a heavy, Pareto-type tail with infinite right endpoint; $\xi = 0$ gives an exponential tail; and, decisively for our purposes, $\xi < 0$ gives a short tail with a *finite* upper endpoint at $u - \sigma/\xi$. A negative shape estimate is thus the statistical fingerprint of an upper bound in the data the very situation produced by a policy cap.

Given exceedances $y_1, \ldots, y_{n_u}$ above u, the parameters are estimated by maximising the log-likelihood

$$\ell(\sigma, \xi) = -n_{\mathrm{u}} \log \sigma - (1 + 1/\xi) \Sigma_{\mathrm{i}} \log(1 + \xi\, y_{\mathrm{i}}/\sigma), \quad \xi \neq 0, \qquad (4)$$

which is maximised numerically; the maximum-likelihood estimator is well behaved for $\xi > -1/2$ and exists for $\xi \le 1$ (del Castillo and Daoudi, 2009).

3.3. The Hill estimator

For a heavy right tail ($\xi > 0$, equivalently tail index $\alpha = 1/\xi$), the Hill estimator based on the top k order statistics $X_{n-k+1:n} \le \ldots \le X_{n:n}$ is

$$\hat{\xi}^{\mathrm{H}}_{k,n} = (1/k) \Sigma_{i=1}^{k} \{ \log X_{n-i+1:n} - \log X_{n-k:n} \}. \qquad (5)$$

Consistency requires $k = k_n \to \infty$ and $k/n \to 0$; asymptotic normality requires in addition a second-order condition controlling the rate at which the tail approaches its Pareto limit. Writing U for the tail quantile function and A for the second-order auxiliary function, the standard representation is

$$\hat{\xi}^{\mathrm{H}}_{k,n} \approx \xi + \xi\, Z_k/\sqrt{k} + A(n/k)/(1 - \rho), \qquad (6)$$

with Z_k asymptotically standard normal and $\rho \le 0$ the second-order parameter. The two terms on the right make the bias–variance trade-off explicit: the stochastic term $\xi\, Z_k/\sqrt{k}$ shrinks as k

ISSN: 2658-8455
Volume 7, Issue 09 (2026), pp. 361-375.

grows, while the bias term $A(n/k)/(1-\rho)$ grows. Minimising the asymptotic mean squared error yields the optimal k_0, whose estimation underlies the automatic methods of Section 2.2.

3.4. Downstream risk measures

The actuarial purpose of the whole exercise is the estimation of tail risk. Given a fitted GPD above u with N_u exceedances out of n observations, the tail estimator of Smith yields, for a confidence level α,

$$\mathrm{VaR}_\alpha = u + (\sigma/\xi)\,\{\,[\,(n/N_u)(1-\alpha)\,]^{-\xi} - 1\,\}, \qquad (7)$$

$$\mathrm{ES}_\alpha = \mathrm{VaR}_\alpha/(1-\xi) + (\sigma - \xi u)/(1-\xi), \qquad (8)$$

valid for $\xi < 1$. Both measures inherit the threshold dependence of the GPD parameters; a change in u propagates through σ and ξ into VaR and ES, and ultimately into the reserve and the solvency capital. This propagation is the reason threshold selection is not a technical afterthought but a determinant of the numbers an insurer reports.

4. Results of theoretical analysis:

4.1. The Truncation Problem and the Bounded-Domain Correction

We now develop the structural argument. Insurance contracts are not unbounded: a motor policy has a sum insured, a medical policy has annual and lifetime caps, and reinsurance layers have explicit limits. The losses recorded in a portfolio are therefore observations of a random variable that is, by construction, confined to a finite interval. Even where a hard cap is absent, administrative practice and the finite size of any real portfolio impose an effective upper bound. The data live on [DL, DR], not on [DL, ∞).

The classical Hill estimator is, in the formulation of Nuyts (2010), the special case of a more general construction in which the upper bound is sent to infinity. Suppose the tail density behaves as a power law, $f(x) \approx \lambda\, x^{-\mu}$, over a range in which both a lower bound L and an upper bound R are operative. Writing $\alpha = \mu - 1$, the conditional mean of log x over [L, R] satisfies an exact identity that, when $R \to \infty$ and $\mu > 1$, collapses to the familiar relation underlying the Hill statistic. Retaining a finite R instead gives the estimating equation

$$(1/(l-r+1))\,\Sigma_{j=r}^{l} \log X_j = 1/\alpha + [\,\log X_l \cdot X_l^{-\alpha} - \log X_r \cdot X_r^{-\alpha}\,] \,/\, (\,X_l^{-\alpha} - X_r^{-\alpha}\,), \qquad (9)$$

where Xl and Xr are the order statistics serving as the lower and upper bounds of the retained sample. Equation (9) is transcendental in α; Nuyts shows it is solved either by direct numerical root-finding or by a contraction iteration that, empirically, converges to high accuracy within a handful of steps. The estimator of the tail exponent is then $\hat{\mu} = \hat{\alpha} + 1$.

Three features of this construction matter for insurance. First, it is symmetric in the two bounds: the upper endpoint enters on the same footing as the lower endpoint, so the estimator uses the information that the data stop at R rather than pretending they continue indefinitely. When R is genuinely large the correction term is negligible and the estimator agrees with Hill; when R is close to the bulk of the data precisely the capped-portfolio case the correction is substantial and, as Nuyts demonstrates on controlled examples, it removes the gross bias that the classical estimator exhibits. Second, the construction remains well behaved when the tail is not a pure power law but is modulated by a slowly varying factor, such as a logarithmic term, which is common in empirical loss data. Third, the symmetry permits the estimator to handle increasing densities, corresponding to a negative exponent, which the classical Hill estimator cannot represent at all.

The relevance to actuarial practice is direct. A negative GPD shape estimate, of the kind reported in the logistic-regression study of Iroko, Tukur and Adeyanju (2025), indicates a finite upper endpoint and is the maximum-likelihood echo of the same truncation that equation (9)

accommodates explicitly. Yet the standard POT pipeline first selects a threshold using methods that assume an unbounded tail, and only then discovers, through a negative ξ, that the tail was bounded after all. We propose instead to acknowledge the bound at the estimation stage. Concretely, when the portfolio carries a known policy limit, or when diagnostic evidence (a Hill plot that bends downward at the largest order statistics, a negative shape estimate) points to truncation, the bounded-domain estimator should be used in place of, or alongside, the classical Hill estimator when assessing the stability of the tail index across candidate thresholds. The upper bound R is then not a nuisance to be assumed away but a known contractual quantity to be exploited.

We stress the limits of this proposal. We do not claim, in the absence of data or simulation, that the bounded-domain estimator improves reserve estimates in any particular portfolio; that is an empirical question we leave open. What we claim is narrower and, we believe, secure: that the bounded-domain construction is the theoretically appropriate response to a structural feature of insurance data that the prevailing methods ignore, and that integrating it into the POT pipeline removes a known source of bias whose presence the existing literature inadvertently confirms.

4.2. A Formal Analysis of Logistic-Regression Threshold Selection

This section contains the paper's central theoretical result. We formalise the logistic-regression selection procedure exactly as proposed, and show that in the univariate setting it cannot discover a threshold: it reproduces the quantile used to generate its labels.

❖ ***The procedure:***

Let X1, …, Xn be independent losses with continuous, strictly increasing distribution function F. Fix a labelling probability $q \in (0, 1)$ (in the original proposal $q = 0.90$) and let $u_{init} = F_n^{-1}(q)$ be the corresponding empirical quantile, with F_n the empirical distribution function. Each observation receives the binary label

$$Y_i = 1\{ X_i > u_{init} \}. \qquad (10)$$

A logistic regression of Y on the single covariate X is fitted, positing

$$\pi(x) = P(Y = 1 \mid X = x) = 1 / (1 + \exp(-\beta_0 - \beta_1 x)), \qquad (11)$$

with β_0, β_1 estimated by maximum likelihood, giving fitted probabilities $\hat{p}_i = \hat{\pi}(X_i)$. Finally, for a fixed cut-off $\tau \in (0, 1)$ (in the original proposal $\tau = 0.90$) the operational threshold is

$$u^{ML} = \min\{ X_i : \hat{p}_i \geq \tau \}. \qquad (12)$$

❖ ***The main result:***

Proposition 1. Let the labels be generated by (10) with labelling probability q, and let the operational threshold uML be defined by (11)–(12) from a logistic regression on the single covariate X. Suppose F is continuous and strictly increasing on its support. Then:

(i) the labels are a deterministic monotone step function of the covariate, $Y_i = 1\{ X_i > u_{init} \}$, so the data are perfectly separated at uinit;

(ii) the maximum-likelihood logistic fit is degenerate, with $\beta_1 \to +\infty$ and decision boundary $x = u_{init}$, so the fitted probability $\hat{p}(x)$ converges to the step function $1\{ x > u_{init} \}$ for every $x \neq u_{init}$;

(iii) consequently, for any fixed cut-off $\tau \in (0, 1)$, the operational threshold satisfies $u_{ML} = u_{init} + o_P(1)$, and hence $u_{ML} \to F^{-1}(q)$ almost surely as $n \to \infty$.

Proof. (i) is immediate from (10): the label is by definition the indicator that the loss exceeds uinit, a strictly increasing transformation of X. There is thus a value of the covariate, namely uinit, such that all observations below it carry label 0 and all observations above it carry label 1. The two label classes are linearly separable in the covariate.

ISSN: 2658-8455
Volume 7, Issue 09 (2026), pp. 361-375.

IJAFAME
International Journal of Accounting, Finance, Auditing, Management & Economics

(ii) Under complete separation the logistic log-likelihood has no finite maximiser. Writing the log-likelihood as

$$\ell(\beta_0, \beta_1) = \Sigma_i \, [\, Y_i \log \pi(X_i) + (1 - Y_i) \log(1 - \pi(X_i)) \,],$$

consider the one-parameter family β0 = −β1uinit, β1 > 0, whose decision boundary is fixed at x = uinit. For this family π(Xi) = 1/(1 + exp(−β1(Xi − uinit)))). For every observation with Yi = 1 we have Xi > uinit, so π(Xi) → 1 as β1 → ∞; for every observation with Yi = 0 we have Xi < uinit, so π(Xi) → 0. Each term of the log-likelihood therefore tends to 0 from below, the log-likelihood increases monotonically along the family, and its supremum (the value 0) is approached only as β1 → ∞. Hence the maximum-likelihood estimate diverges, β̂1 → +∞ with β̂0/β̂1 → −uinit, and the fitted probability converges pointwise to the Heaviside step 1{x > uinit} at every continuity point x ≠ uinit. (In finite samples the same logic applies up to the usual numerical regularisation: solvers report a very large β̂1 and a boundary indistinguishable from uinit.)

(iii) Fix τ ∈ (0, 1). By (ii), p̂(x) → 1{x > uinit}, so for all sufficiently large β̂1 the set {x : p̂(x) ≥ τ} equals {x : x ≥ cn)} for a crossing point cn → uinit. The smallest observed loss in this set is the smallest Xi exceeding cn; since cn → uinit and the order statistics around the q-quantile are dense, uML = uinit + oP(1). Finally the Glivenko–Cantelli theorem gives uinit = Fn−1(q) → F−1(q) almost surely, and the conclusion follows.

❖ ***Interpretation:***

Proposition 1 makes precise the sense in which the procedure is circular. The cut-off τ does not select a threshold in any substantive way: because the fitted probability degenerates to a step at uinit, almost any τ strictly between 0 and 1 returns the same point. The operational threshold is governed not by τ but by q, the labelling probability fixed before the regression is ever run. The logistic regression contributes no information beyond the quantile rule; it is an elaborate identity map from the labelling quantile to the operational threshold.

This explains, without appeal to data, two features reported in the empirical literature. The agreement between the logistic threshold and the linear region of the mean residual life plot is not independent validation: it holds because the labelling quantile was chosen to sit in a region that the analyst already regarded as the tail, and the proposition guarantees the logistic threshold will land there too. The sensitivity of the selected threshold to the labelling quantile, which the original study documents, is not a minor caveat but the direct empirical manifestation of part (iii): move q, and uML moves with it, by construction.

Two qualifications delimit the result. First, it is specific to the univariate, monotone-feature construction. If the classifier is given covariates other than the loss amount policyholder characteristics, claim cause, exposure the labels cease to be a deterministic function of the features, separation no longer holds, the logistic fit is non-degenerate, and the predicted probability can encode genuine structure. In that richer setting logistic regression, or a more flexible learner, may legitimately inform threshold selection, and our critique does not apply. Second, the result concerns what the procedure estimates, not whether the resulting threshold is a bad threshold. If the labelling quantile happens to be well chosen, the operational threshold inherits that quality. The point is solely that the merit, such as it is, resides entirely in the choice of q, and that dressing a quantile rule in the apparatus of supervised learning confers an appearance of objectivity that the underlying mechanics do not support.

4.3. A Unified Threshold-Selection Framework for Truncated Insurance Losses

The preceding sections argue that no method escapes the bias–variance trade-off, that the learning-based shortcut is, in the univariate case, a quantile rule in disguise, and that truncation is a structural feature the prevailing methods ignore. We now assemble these strands into a

single decision framework. The framework is organised not by the historical lineage of methods but by the two questions that actually determine which method is defensible: what is the structure of the data, and what is the downstream objective?

❖ ***Common logic:***

Every method in the literature can be read as a particular way of locating the point at which squared bias and variance balance. Heuristic rules fix the number of exceedances as a function of sample size alone $k = \varepsilon n$ for the fixed-quantile rule, $k = \sqrt{n}$ for the square-root rule (Ferreira, de Haan and Peng, 2003) and so place the balance point by a rule of thumb that is blind to tail heaviness. AMSE-based and bootstrap methods estimate the balance point directly from second-order structure, at the cost of estimating nuisance parameters. Graphical methods invite the analyst to locate the balance point visually, trading formality for transparency. Learning-based methods, in the univariate case, locate it at a pre-specified quantile. Seeing the methods this way clarifies that they differ not in aim but in the information they use and the burden they impose.

❖ ***A decision framework:***

On this basis we propose the following ordering of considerations.

- **Diagnose truncation first**. Before selecting a threshold, examine whether the portfolio carries known policy limits and whether the upper order statistics show the downward bend or negative shape signalling a finite endpoint. If truncation is present, the tail-index assessment should use the bounded-domain estimator of equation (9) rather than the classical Hill estimator, with the contractual cap supplying the upper bound R. This step is logically prior to selection because it determines which estimator the selection diagnostics should be built on.
- **Match the method to the sample size**. Following the evidence assembled by Wang, Hobæk Haff and Huseby (2020), when exceedances are plentiful the square-root rule and the exponentiality test are reliable and cheap, with the latter preferable for very heavy tails; when data are scarce, simultaneous estimation of threshold and parameters is more stable. AMSE-based and double-bootstrap methods are appropriate when second-order structure can be estimated with confidence.
- **Use graphical diagnostics as corroboration, not adjudication**. The mean residual life and threshold-choice plots remain valuable as checks on a threshold chosen by a reproducible rule, and goodness-of-fit statistics such as deviance and the Akaike information criterion (Alaswed and Mohamed, 2019) can arbitrate between candidates, but they should not be the sole basis of an auditable decision.
- **Treat univariate logistic regression as a quantile rule**. In light of Proposition 1, if a logistic-regression threshold is reported it should be understood and documented as equivalent to the labelling quantile q. If genuine learning is intended, the classifier must be supplied with covariates beyond the loss amount; otherwise the simpler and more honest course is to state the quantile rule directly.
- **Judge the threshold by the downstream quantity.** Because VaR, ES and the reserve are what the insurer reports, the stability of these quantities across a band of candidate thresholds is the most relevant diagnostic of all. A threshold that yields risk measures robust to small perturbations is preferable to one that optimises an intermediate criterion but leaves the reported figures fragile.

The framework does not prescribe a single method, because the review establishes that none is universally best. It prescribes instead an order of reasoning: diagnose structure, choose a reproducible rule suited to the sample, corroborate graphically, demystify any learning-based step, and validate against the downstream measure. This is, we contend, the appropriate response to a problem that is genuinely irreducible.

ISSN: 2658-8455
Volume 7, Issue 09 (2026), pp. 361-375.

5. Discussion :

The two contributions of this paper pull in the same direction. The formal analysis of logistic-regression selection is, on its face, a negative result: it withholds from a fashionable method the objectivity it claims. But its constructive content is the clarification it brings. By showing exactly what the univariate procedure estimates, it tells practitioners how to interpret a logistic threshold (as a quantile), how to report it honestly (by stating q), and how to make the method genuinely informative (by adding covariates). The bounded-domain proposal is constructive in a different way: it identifies a mismatch between the prevailing methods and the data, and supplies the existing remedy that the actuarial literature had overlooked.

Our findings sit comfortably with the broader literature. Caeiro and Gomes (2014) conclude that no selection method is universally optimal; Proposition 1 sharpens this by removing one candidate from contention as a source of new information in the univariate case. Wang, Hobæk Haff and Huseby (2020) show that the merit of a method depends on the downstream reserve and the sample size; our framework elevates that finding into an organising principle. The persistent message of three decades of work that threshold selection is irreducibly a matter of trading bias against variance under uncertainty about the data-generating process is reinforced, not overturned, by the appearance of machine-learning methods. Automation can make a choice reproducible, but it cannot make the underlying trade-off disappear, and it can disguise the point at which a human assumption entered.

We are candid about the price of a purely theoretical treatment. Proposition 1 is a limiting statement; in finite samples with numerical regularisation the degeneracy is approximate rather than exact, although the approximation is precisely what produces the near-deterministic behaviour observed in practice. The bounded-domain argument establishes appropriateness, not empirical superiority; whether the correction materially improves reserve or capital estimates in a given portfolio is a question only data or simulation can settle. We regard these as well-defined openings for subsequent empirical work rather than as weaknesses of the conceptual claims, which stand on their own terms.

6. Conclusion

Threshold selection is the hinge on which peaks-over-threshold modelling of extreme insurance losses turns, and it has resisted a definitive solution because it embodies an irreducible trade-off between bias and variance under uncertainty about the tail. This paper has examined two responses to that difficulty. The first, the recent proposal to automate selection by logistic regression, we have shown to be, in its univariate form, a quantile rule in disguise: under transparent conditions the procedure returns the very quantile used to label its training data, so that its reproducibility is genuine but its objectivity is illusory. The second, the long-overlooked bounded-domain correction of the Hill estimator, we have argued to be the structurally appropriate response to the truncation that policy limits impose on insurance data, and we have integrated it into the POT pipeline. Around these results we have built a unified framework that orders the competing methods by data structure and downstream objective rather than by historical lineage.

The principal limitation of this work is deliberate: it is theoretical, and it makes no empirical claim. Proposition 1 is an asymptotic statement, and although its finite-sample shadow is exactly the behaviour practitioners report, a simulation study quantifying the rate at which ML approaches the labelling quantile, and the residual influence of numerical regularisation, would usefully complement the proof. Likewise, the case for the bounded-domain estimator is an argument from structure; a controlled comparison of reserve and capital estimates obtained with the classical Hill estimator, with reduced-bias estimators, and with the bounded-domain

estimator, on portfolios with known caps, would establish whether the structural appropriateness translates into practical gain.

Several further directions follow naturally. The independence assumption underlying the POT construction is unrealistic for insurance data subject to catastrophe clustering, seasonal effects and policyholder-level dependence; integrating declustering or explicit dependence structures with the bounded-domain estimator is an open problem. The multivariate extension of the bounded-domain idea truncation in several dimensions, as arises with combined limits across coverages has not been studied. And the constructive half of our critique invites development: identifying the covariates that allow a classifier to inform threshold selection genuinely, and characterising when the resulting non-degenerate logistic fit improves on a quantile rule, would turn a negative result about the univariate case into a positive methodology for the multivariate one.

The overarching message is methodological humility of a productive kind. Neither visual habit nor algorithmic novelty dissolves the bias variance trade-off; what they can do is make a choice transparent or opaque, structurally informed or structurally naive. A defensible actuarial practice diagnoses the structure of its data, chooses a reproducible rule suited to the sample, and validates that rule against the quantities it must ultimately report, rather than inferring objectivity from the sophistication of the method used to reach it. The contributions here are conceptual, and they point clearly to the empirical work that should follow.

Références

(1). Alaswed, H. A., & Mohamed, M. A. (2019). Threshold selection using graphical methods. *Journal of Pure & Applied Sciences*, 17(1), 42-45.

(2). Balkema, A. A., & de Haan, L. (1974). Residual life time at great age. *Annals of Probability*, 2(5), 792-804.

(3). Beirlant, J., Goegebeur, Y., Segers, J., & Teugels, J. L. (2004). *Statistics of extremes: Theory and applications.* John Wiley & Sons.

(4). Caeiro, F., & Gomes, M. I. (2014). Threshold selection in extreme value analysis. In *Extreme value modeling and risk analysis: Methods and applications* (pp. 69-82). Wiley.

(5). Caeiro, F., Gomes, M. I., & Pestana, D. (2005). Direct reduction of bias of the classical Hill estimator. *REVSTAT Statistical Journal*, 3(2), 113-136.

(6). Coles, S. (2001). *An introduction to statistical modeling of extreme values.* Springer.

(7). Csörgő, S., Deheuvels, P., & Mason, D. (1985). Kernel estimates of the tail index of a distribution. *The Annals of Statistics*, 13(3), 1050-1077.

(8). Danielsson, J., de Haan, L., Peng, L., & de Vries, C. G. (2001). Using a bootstrap method to choose the sample fraction in tail index estimation. *Journal of Multivariate Analysis*, 76(2), 226-248.

(9). Davison, A. C., & Smith, R. L. (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society: Series B*, 52(3), 393-442.

(10). del Castillo, J., & Daoudi, J. (2009). Estimation of the generalized Pareto distribution. *Statistics & Probability Letters*, 79(5), 684-688.

(11). Draisma, G., de Haan, L., Peng, L., & Pereira, T. T. (1999). A bootstrap-based method to achieve optimality in estimating the extreme-value index. *Extremes*, 2(4), 367-404.

(12). Drees, H., & Kaufmann, E. (1998). Selecting the optimal sample fraction in univariate extreme value estimation. *Stochastic Processes and Their Applications*, 75(2), 149-172.

(13). Drees, H., de Haan, L., & Resnick, S. (2000). How to make a Hill plot. *The Annals of Statistics*, 28(1), 254-274.

ISSN: 2658-8455
Volume 7, Issue 09 (2026), pp. 361-375.

(14). Embrechts, P., Klüppelberg, C., & Mikosch, T. (1997). *Modelling extremal events for insurance and finance.* Springer.
(15). Ferreira, A., de Haan, L., & Peng, L. (2003). On optimising the estimation of high quantiles of a probability distribution. *Statistics*, 37(5), 401-434.
(16). Gomes, M. I., & Guillou, A. (2014). Extreme value theory and statistics of univariate extremes: A review. *International Statistical Review*, 83(2), 263-292.
(17). Gomes, M. I., & Martins, M. J. (2002). Asymptotically unbiased estimators of the tail index based on external estimation of the second order parameter. *Extremes*, 5(1), 5-31.
(18). Gomes, M. I., & Oliveira, O. (2001). The bootstrap methodology in statistics of extremes: Choice of the optimal sample fraction. *Extremes*, 4(4), 331-358.
(19). Guillou, A., & Hall, P. (2001). A diagnostic for selecting the threshold in extreme value analysis. *Journal of the Royal Statistical Society: Series B*, 63(2), 293-305.
(20). Hall, P. (1982). On some simple estimates of an exponent of regular variation. *Journal of the Royal Statistical Society: Series B*, 44(1), 37-42.
(21). Hall, P. (1990). Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems. *Journal of Multivariate Analysis*, 32(2), 177-203.
(22). Hall, P., & Welsh, A. H. (1985). Adaptive estimates of parameters of regular variation. *The Annals of Statistics*, 13(1), 331-341.
(23). Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, 3(5), 1163-1174.
(24). Iroko, T. C., Tukur, I., & Adeyanju, V. (2025). Threshold selection for peak over threshold models using logistic regression. *Earthline Journal of Mathematical Sciences*, 15(6), 1021-1036. https://doi.org/10.34198/ejms.15625.10211036
(25). Northrop, P. J., & Coleman, C. L. (2014). Improved threshold diagnostic plots for extreme value analyses. *Extremes*, 17(2), 289-303.
(26). Nuyts, J. (2010). Inference about the tail of a distribution: Improvement on the Hill estimator. *International Journal of Mathematics and Mathematical Sciences*, 2010, Article 924013. https://doi.org/10.1155/2010/924013
(27). Pickands, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics*, 3(1), 119-131.
(28). Reiss, R.-D., & Thomas, M. (2007). *Statistical analysis of extreme values with applications to insurance, finance, hydrology and other fields* (3rd ed.). Birkhäuser.
(29). Resnick, S., & Stărică, C. (1997). Smoothing the Hill estimator. *Advances in Applied Probability*, 29(1), 271-293.
(30). Scarrott, C., & MacDonald, A. (2012). A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT Statistical Journal*, 10(1), 33-60.
(31). Scollnik, D. P. M. (2007). On composite lognormal-Pareto models. *Scandinavian Actuarial Journal*, 2007(1), 20-33.
(32). Smith, R. L. (1987). Estimating tails of probability distributions. *The Annals of Statistics*, 15(3), 1174-1207.
(33). Tang, Y., Wang, H. J., & Li, D. (2024). High-dimensional extreme quantile regression. *arXiv*. https://arxiv.org/abs/2411.13822
(34). Wadsworth, J. L. (2016). Exploiting structure of maximum likelihood estimators for extreme value threshold selection. *Technometrics*, 58(1), 116-126.
(35). Wang, Y., Hobæk Haff, I., & Huseby, A. (2020). Modelling extreme claims via composite models and threshold selection methods. *Insurance: Mathematics and Economics*, 91, 257-268. https://doi.org/10.1016/j.insmatheco.2020.02.009.