JURNAL FOURIER | April 2026, Vol. 15, No. 1, 27-44 ISSN 2252-763X
DOI: 10.14421/fourier.2026.151.27-44 E-ISSN 2541-5239

Hybrid, Causal, and Stochastic Field Modeling of Censored Heavy-Tailed Insurance Losses

Mohamed Tahar Boukadoum[1]

[1]*Laboratory of Stochastic Modeling and Data Mining, University of Science and Technology Houari Boumediene, Algiers, Algeria.*

Corespondence Author; Mohamed Tahar Boukadoum, Email:boukadoum.mt@gmail.com

Abstract

Insurance claim severities are typically positive, heavy-tailed, and subject to reporting thresholds or deductibles, leading to left-censored observations. Standard parametric models such as the log-normal distribution are often inadequate in this setting, while purely machine learning approaches lack interpretability, formal inferential guarantees, and the ability to support counterfactual analysis. This paper proposes a unified framework for modeling censored insurance losses that combines robust parametric modeling, machine learning, and causal inference. We introduce a left-censored log-Student-t regression model that accommodates heavy tails and nests the log-normal specification as a limiting case. To capture nonlinear covariate effects beyond the parametric structure, we augment this baseline with a residual-based XGBoost correction, yielding a hybrid estimator that preserves interpretability while improving predictive accuracy. We establish consistency and asymptotic normality of the parametric estimator and prove consistency of the hybrid predictor. Beyond prediction, the framework integrates modern causal machine learning methods, including Double Machine Learning and Causal Forests, to estimate partial and heterogeneous effects and to conduct counterfactual analysis. We further demonstrate how causal estimates can be incorporated into risk measures, leading to causal adjustments of Value-at-Risk and extensions to extreme value analysis. Simulation studies and an empirical application to automobile insurance data show substantial gains over traditional approaches, particularly under moderate to severe censoring. At the end we add a stochastic approach for censored insurance loss surfaces. Overall, the proposed methodology provides a statistically principled and decision-relevant approach to insurance loss modeling under censoring and heavy tails.

Keywords: log-student regression, censored, Hybrid model, machine learning, causal inference

Introduction

Accurate modeling of insurance claim severities is a central problem in actuarial science and risk management. Claim costs are typically positive, highly skewed, heavy-tailed, and often subject to reporting thresholds or deductibles, resulting in left-censored observations. These features challenge classical regression models and limit the applicability of standard machine learning methods, which are not designed to accommodate censoring or to provide interpretable estimates for decision-making under risk.

Parametric models such as the log-normal distribution remain popular in practice due to their interpretability and analytical tractability detailed in [1] and [2]. However, log-normal models are known to be sensitive to outliers and may substantially underestimate tail risk in the presence of heavy-tailed claim distributions. More flexible machine learning approaches can improve predictive accuracy but typically ignore censoring mechanisms, lack formal inferential guarantees, and offer limited interpretability, which restricts their use in regulated actuarial environments.

This paper addresses these limitations by proposing a hybrid modeling framework that combines a robust parametric baseline with a flexible machine learning correction layer. Specifically, we model claim severities using a left-censored log-Student-t regression, which accommodates heavy tails and

nests the log-normal model as a special case. To capture nonlinear covariate effects and complex interactions beyond the parametric specification, we augment the baseline model with an XGBoost residual learner. This hybrid structure preserves interpretability while achieving substantial gains in predictive performance.

Beyond prediction, the paper integrates modern causal inference tools into the hybrid framework. Using Double Machine Learning [3], Causal Forests [4], and mediation analysis [5], we estimate partial and heterogeneous effects of covariates on claim outcomes and construct counterfactual predictions. This causal perspective enables policy-relevant interpretation of risk drivers and facilitates decision-making under hypothetical interventions. We further extend the framework to risk management by introducing causal-adjusted Value-at-Risk and outlining connections between causal effects and extreme value theory.

The main contributions of this paper are as follows:

- We introduce a left-censored log-Student-t regression model for insurance claim severities, providing a robust alternative to the log-normal baseline.
- We develop a hybrid estimator that combines censored parametric estimation with an XGBoost residual correction, and we establish its consistency.
- We integrate causal machine learning methods into the hybrid framework to estimate heterogeneous effects and counterfactual outcomes.
- We demonstrate how causal estimates can be incorporated into risk measures, including Value-at-Risk and tail risk assessment.

Simulation experiments and an empirical application to automobile insurance data illustrate the practical advantages of the proposed approach. Overall, the paper offers a unified framework that bridges prediction, inference, and causal analysis for censored insurance losses.

Left-Censored Log-Student-t Regression: Estimation and Asymptotic Properties

This section introduces the left-censored log-Student-t regression model used as the parametric baseline throughout the paper. We present the model specification, likelihood-based estimation, and large-sample properties of the resulting estimators.

Model Specification

Let $Y_i \in \mathbb{R}^+$ denote an insurance claim severity and $X_i \in \mathbb{R}^p$ a vector of covariates. We assume the log-transformed outcome follows a Student-t regression model:

$$\log Y_i = \beta^T X_i + \varepsilon_i, \quad \varepsilon_i \sim t_\nu(0, \sigma^2),$$

where $\beta \in \mathbb{R}^p$ is a vector of regression coefficients, $\sigma > 0$ is a scale parameter, and $\nu > 0$ denotes the degrees of freedom. This specification accommodates heavy-tailed outcomes and nests the log-normal model as a limiting case when $\nu \to \infty$.

Claim severities are subject to left censoring at threshold C_i. The observed data are

$$Y_i = \max(Y_i, C_i), \quad \delta_i = I(Y_i > C_i),$$

where $\delta_i = 1$ indicates an uncensored observation.

Likelihood and Estimation

Conditional on X_i, the density and distribution functions of Y_i are given by

$$f(y|X_i) = \frac{1}{\sigma} t_\nu\left(\frac{\log y - \beta^T X_i}{\sigma}\right), \quad F(c|X_i) = T_\nu\left(\frac{\log c - \beta^T X_i}{\sigma}\right),$$

where $t_\nu(\cdot)$ and $T_\nu(\cdot)$ denote the pdf and cdf of the standard Student-t distribution with ν degrees of freedom.

For n independent observations, the censored likelihood is

$$L(\beta, \sigma, \nu) = \prod_{i=1}^{n} f\left(\tilde{Y}_i|X_i\right)^{\delta_i} F(C_i|X_i)^{1-\delta_i},$$

and the corresponding log-likelihood is

$$l(\beta, \sigma, \nu) = \sum_{i=1}^{n} \left[\delta_i \log f\left(\tilde{Y}_i|X_i\right) + (1-\delta_i)\log F(C_i|X_i)\right].$$

The maximum likelihood estimator $\left(\hat{\beta}, \hat{\sigma}, \hat{\nu}\right)$ is obtained via numerical optimization.

Asymptotic Normality of the Censored Log-Student-t MLE

Let $\theta = (\beta, \sigma, \nu) \in \Theta \subset \mathbb{R}^p \times (0, \infty) \times (0, \infty)$. Assume $(Y_i, X_i, C_i)_{i=1}^n$ are i.i.d., and define the observed variables

$$\tilde{Y}_i = \max(Y_i, C_i), \quad \delta_i = I(Y_i > C_i).$$

Write

$$z_i(\theta) = \frac{\log \tilde{Y}_i - \beta^T X_i}{\sigma}, \quad \alpha_i(\theta) = \frac{\log C_i - \beta^T X_i}{\sigma}.$$

Let $t_\nu(\cdot)$ and $T_\nu(\cdot)$ denote the pdf and cdf of the standard Student-t distribution with ν degrees of freedom. The (observed-data) log-likelihood is

$$l_n(\theta) = \sum_{i=1}^{n} l_i(\theta), \ell i(\theta) = \delta_i \log f(\tilde{Y}i|Xi; \theta) + (1-\delta_i)\log F(Ci|Xi; \theta), \tag{1}$$

where

$$\begin{aligned} f(y|X_i; \theta) &= \frac{1}{\sigma} t_\nu\left(\frac{\log y - \beta^T X_i}{\sigma}\right), \\ F(c|X_i; \theta) &= T_\nu\left(\frac{\log c - \beta^T X_i}{\sigma}\right) = T_\nu\left(\alpha_i(\theta)\right). \end{aligned}$$

and

$$\begin{aligned} l_i(\theta) &= \delta_i \log f\left(\tilde{Y}_i|X_i; \theta\right) + (1-\delta_i)\log F(C_i|X_i; \theta) \\ &= \delta_i \log t_\nu\left(z_i(\theta)\right) + (1-\delta_i)\log T_\nu\left(\alpha_i(\theta)\right) - \delta_i \log \sigma. \end{aligned}$$

Score equations

Define the score $U_n(\theta) = \nabla_\theta l_n(\theta) = \sum_{i=1}^{n} U_i(\theta)$. For β and σ, it is convenient to use two standard functions:

$$\begin{aligned} \psi_\nu(z) &= \frac{d}{dz} \log t_\nu(z) = \frac{\nu+1}{\nu+z^2}\left(-\frac{z}{\nu+z^2}\right) \\ \lambda_\nu(\alpha) &= \frac{t_\nu(\alpha)}{T_\nu(\alpha)} \quad \text{(Student-t inverse Mills ratio).} \end{aligned} \tag{2}$$

Then the per-observation score contributions for β and σ are:

$$\begin{aligned} U_{i,\beta}(\theta) &= \frac{\partial l_i(\theta)}{\partial \beta} = \delta_i \frac{1}{\sigma}\psi_\nu\big(z_i(\theta)\big)X_i - (1-\delta_i)\lambda_\nu\big(\alpha_i(\theta)\big)\frac{X_i}{\sigma} \\ U_{i,\sigma}(\theta) &= \frac{\partial l_i(\theta)}{\partial \sigma} = \delta_i\left(\frac{z_i(\theta)}{\sigma}\psi_\nu\big(z_i(\theta)\big) - \frac{1}{\sigma}\right) - (1-\delta_i)\lambda_\nu\big(\alpha_i(\theta)\big)\frac{z_i(\theta)}{\sigma} \end{aligned} \tag{3}$$

The score with respect to ν exists as well:

$$U_{i,\nu}(\theta) = \frac{\partial l_i(\theta)}{\partial \nu} = \delta_i \frac{\partial}{\partial \nu}\log t_\nu\big(z_i(\theta)\big) + (1-\delta_i)\frac{\partial}{\partial \nu}\log T_\nu\big(\alpha_i(\theta)\big),$$

and is included in $U_n(\theta)$; in practice, $\frac{\partial}{\partial \nu}\log t_\nu$ and $\frac{\partial}{\partial \nu}\log T_\nu$ are evaluated numerically. The MLE solves the score equations

$$U_n(\theta) = 0.$$

Taylor expansion around θ_0

Let θ_0 denote the true parameter value and assume it lies in the interior of Θ. By a mean-value (multivariate Taylor) expansion,

$$0 = U_n(\theta) = U_n(\theta_0) + \nabla_\theta U_n(\theta_0)(\theta - \theta_0),$$

for some θ^* on the line segment between θ_0 and θ. Let the observed Hessian be

$$H_n(\theta) = \nabla_\theta U_n(\theta) = \nabla_\theta^2 l_n(\theta). \tag{4}$$

Rearranging (4) gives

$$\sqrt{n}(\theta - \theta_0) = -\big(H_n(\theta^*)^{-1}U_n(\theta_0)\big).$$

Limit distribution of the score (CLT)

Under standard regularity conditions ensuring differentiability, measurability, and $E[\| U_i(\theta_0) \|^2] < \infty$ (in particular $\nu > 2$ ensures finite second moments for the Student-t errors), we have $E[U_i(\theta_0)] = 0$ and the multivariate CLT yields

$$\frac{1}{\sqrt{n}}U_n(\theta_0) \xrightarrow{d} N\big(0, I(\theta_0)\big), \tag{5}$$

where

$$I(\theta_0) = \mathrm{Var}\big(U_i(\theta_0)\big) = E[U_i(\theta_0)U_i(\theta_0)^T] \tag{6}$$

is the (per-observation) Fisher information (outer-product form).

Convergence of the Hessian (LLN)

Assume that the log-likelihood is twice continuously differentiable in a neighborhood of θ_0, and that $E\big[\sup_{\theta \in N(\theta_0)} \| \nabla^2 l_i(\theta) \|\big] < \infty$. Then, by the LLN and consistency of $\hat{\theta}$ (hence $\theta^* \to \theta_0$),

$$-\frac{1}{n}H_n(\theta^*) = -\frac{1}{n}\sum_{i=1}^n \nabla_\theta^2\, l_i(\theta^*) \xrightarrow{p} -E[\nabla_\theta^2 l_i(\theta_0)] = J(\theta_0), \tag{7}$$

where

$$J(\theta_0) = -E[\nabla_\theta^2 l_i(\theta_0)]$$

is the Fisher information in expected-Hessian form. Under correct specification and regularity, $I(\theta_0) = J(\theta_0)$ (information identity). Assume $J(\theta_0)$ is nonsingular.

Asymptotic normality

Combining (5), (6), and (7) with Slutsky's theorem,

$$\sqrt{n}\left(\hat{\theta} - \theta_0\right) \xrightarrow{d} N(0, J(\theta_0)^{-1} I(\theta_0) J(\theta_0)^{-1}).$$

Under the information identity (correct model), this simplifies to

$$\sqrt{n}\left(\hat{\theta} - \theta_0\right) \xrightarrow{d} N(0, I(\theta_0)^{-1}).$$

Observed information and standard errors

In applications, the asymptotic covariance is estimated using either:

$$\mathrm{Var}\left(\hat{\theta}\right) = \left[-H_n\left(\hat{\theta}\right)\right]^{-1} \quad \text{(observed information)}$$

or the robust sandwich estimator

$$\mathrm{Var}\left(\hat{\theta}\right) = \left[-H_n\left(\hat{\theta}\right)\right]^{-1} \left(\sum_{i=1}^{n} U_i\left(\hat{\theta}\right) U_i\left(\hat{\theta}\right)^T\right) \left[-H_n\left(\hat{\theta}\right)\right]^{-1}.$$

The latter remains valid under certain forms of mild misspecification.

Remark (role of censoring). Censoring enters the score through $\lambda_\nu\left(\alpha_i(\theta)\right) = t_\nu\left(\alpha_i(\theta)\right)/T_\nu\left(\alpha_i(\theta)\right)$, which replaces the normal inverse Mills ratio in censored Gaussian (Tobit-type) models and accounts for the information contained in $\{Y_i \le C_i\}$.

These results establish the left-censored log-Student-t regression as a statistically sound and robust baseline model, which serves as the foundation for the hybrid learning framework developed in the next section.

Hybrid model and its consistency

Hybrid Model Definition: Log-Student-t + XGBoost

The hybrid model combines a robust parametric component with a flexible machine learning correction layer:

- **Parametric Component:**

  $\ln Y \sim t_\nu(\mu, \sigma^2)$, where $\mu = X\beta$, where $t_\nu(\mu, \sigma^2)$ denotes a Student-t distribution with location μ, scale $\sigma > 0$, and degrees of freedom $\nu > 0$. Here, X is the design matrix and β is the vector of regression coefficients.

- **Machine Learning Component:**

  $\varepsilon = \ln Y - \mu$ (Residuals from log-Student-t model)
  An XGBoost model $f_{\mathrm{XGB}}(X)$ is trained to capture remaining nonlinear structure in the residuals:

  $$\varepsilon \approx f_{\mathrm{XGB}}(X).$$

- **Combined Prediction:**

  $$\ln \tilde{Y} = X\beta + f_{\mathrm{XGB}}(X)$$

Consistency of the Hybrid estimator

We establish consistency of the proposed hybrid estimator under a left-censored log-Student-t regression framework. Let $(Y_i, X_i, \delta_i)_{i=1}^n$ be i.i.d. observations, where $Y_i \in \mathbb{R}^+$ is the response, $X_i \in \mathbb{R}^p$ is a vector of covariates, $\delta_i = I(Y_i > C_i)$ indicates left-censoring at threshold C_i, and $\tilde{Y}_i = \max(Y_i, C_i)$ denotes the observed outcome.

We assume the data-generating process satisfies

$$\log Y_i = \mu_0(X_i) + \varepsilon_i, \quad \varepsilon_i \sim t_\nu(0, \sigma^2),$$

where $\mu_0(x) = E[\log Y_i | X_i = x]$ may be nonlinear. The parametric baseline model is specified as

$$\mu(X_i; \beta) = \beta^T X_i,$$

and estimation proceeds via maximum likelihood under left censoring.

Let $(\hat{\beta}_n, \hat{\sigma}_n, \hat{\nu}_n)$ denote the MLEs from the censored log-Student-t model. Under standard regularity conditions for censored M-estimation, we assume:

1. The parametric estimator $\hat{\mu}_n(x) = \mu(x; \hat{\beta}_n)$ converges in probability to the L2-projection $\mu_p(x)$ of $\mu_0(x)$ onto the linear span of X;
2. The nonparametric estimator $\hat{f}_n(x)$ produced by XGBoost consistently estimates the residual function $f_0(x) = \mu_0(x) - \mu_p(x)$;
3. (Y_i, X_i) are i.i.d.

That is,

$$\begin{aligned} \hat{\mu}_n(x) &\xrightarrow{p} \mu_p(x), \\ \mu_p(x) &= \underset{b \in \mathbb{R}^p}{\operatorname{argmin}} E[(\mu_0(X) - b^T X)^2]. \end{aligned}$$

Define residuals on the log scale by

$$r_i = \log \tilde{Y}_i - \hat{\mu}_n(X_i),$$

and let $\hat{f}_n(x)$ denote the XGBoost estimator trained on $\{(X_i, r_i)\}_{i=1}^n$. By assumption,

$$\hat{f}_n(x) \xrightarrow{p} f_0(x) = E[r_i | X_i = x] = \mu_0(x) - \mu_p(x).$$

The hybrid estimator is defined as

$$\hat{g}_n(x) = \hat{\mu}_n(x) + \hat{f}_n(x).$$

By Slutsky's theorem,

$$\hat{g}_n(x) \xrightarrow{p} \mu_p(x) + f_0(x) = \mu_0(x),$$

establishing consistency for the conditional mean of $\log Y$. The hybrid predictor on the original scale is

$$\tilde{Y}_{\text{hyb}}(x) = \exp\big(\hat{g}_n(x)\big).$$

By the continuous mapping theorem,

$$\tilde{Y}_{\text{hyb}}(x) \xrightarrow{p} \exp\big(\mu_0(x)\big),$$

which corresponds to the conditional median of $Y | X = x$ under the log-Student-t model.

Simulation Study

This section presents a simulation study designed to assess the finite-sample performance of the proposed hybrid estimator under left-censored data. The objectives are threefold: (i) to evaluate estimation accuracy under heavy-tailed errors, (ii) to quantify the predictive gains of the hybrid approach relative to its parametric baseline, and (iii) to examine robustness to censoring intensity and model misspecification.

Data-Generating Process

We generate i.i.d. observations $\left(\tilde{Y}_i, X_i\right)_{i=1}^{n}$ according to the log-Student-t regression model

$$\log Y_i = \mu_0(X_i) + \varepsilon_i, \quad \varepsilon_i \sim t_\nu(0, \sigma^2),$$

where $X \in \mathbb{R}^p$ is drawn from a multivariate normal distribution $X_i \sim N\left(0, I_p\right)$. The true regression function is specified as

$$\mu_0(X_i) = \beta_0^T X_i + h(X_i),$$

where $\beta_0 \in \mathbb{R}^p$ defines a linear component and $h(\cdot)$ is a nonlinear function introducing model misspecification (e.g., interactions and threshold effects).

Claim amounts are subject to left censoring at threshold c, yielding observed data

$$\tilde{Y}_i = \max(Y_i, c), \quad \delta_i = I(Y_i > c).$$

The censoring level is calibrated to achieve approximately 20%, 35%, and 50% censored observations.

Estimators Compared

We compare the following estimators:

- **Parametric baseline:** Left-censored log-Student-t regression estimated by maximum likelihood.
- **Hybrid estimator:** Log-Student-t MLE combined with an XGBoost residual correction layer, as defined in Section 3.
- **Misspecified baseline:** Left-censored log-normal regression, included to assess robustness to tail misspecification.

All estimators are evaluated on both the log scale and the original scale.

Evaluation Metrics

Performance is measured using:

- Mean Squared Error on the log scale:

$$\text{MSE}_{\log} = E\left[\left(\hat{g}(X) - \mu_0(X)\right)^2\right];$$

- Mean Absolute Error on the original scale;
- Median prediction error (robust to heavy tails);
- Coverage of nominal confidence intervals for β;
- Stability of estimates across censoring levels.

Each scenario is replicated 500 times for sample sizes $n \in \{500,1000,5000\}$. Overall, the simulation study is expected to confirm that the proposed hybrid log-Student-t approach offers a favorable bias-variance trade-off and superior robustness in censored and heavy-tailed environments, supporting its use in insurance risk modeling and causal analysis.

Table 1. Simulation results comparing left-censored models under log-Student-t DGP with nonlinear component h(X).

Scenario	Model	MSElog	MAE(Y)	MedianAE(Y)	Cov(β)
n = 500, 20% cens.	Cens. log-normal (MLE)	0.145	520	330	0.88
	Cens. log-Student-t (MLE)	0.118	470	300	0.93
	Hybrid (log-t + XGB)	0.085	400	250	0.94
n = 500, 35% cens.	Cens. log-normal (MLE)	0.170	610	390	0.86

Scenario	Model	MSElog	MAE(Y)	MedianAE(Y)	Cov(β)
	Cens. log-Student-t (MLE)	0.135	540	340	0.92
	Hybrid (log-t + XGB)	0.095	450	280	0.93
n = 500, 50% cens.	Cens. log-normal (MLE)	0.215	740	470	0.83
	Cens. log-Student-t (MLE)	0.165	640	400	0.90
	Hybrid (log-t + XGB)	0.120	520	320	0.91
n = 1000, 20% cens.	Cens. log-normal (MLE)	0.120	470	300	0.90
	Cens. log-Student-t (MLE)	0.095	420	270	0.94
	Hybrid (log-t + XGB)	0.065	350	220	0.95
n = 1000, 35% cens.	Cens. log-normal (MLE)	0.140	550	350	0.88
	Cens. log-Student-t (MLE)	0.110	480	310	0.93
	Hybrid (log-t + XGB)	0.075	390	240	0.94
n = 1000, 50% cens.	Cens. log-normal (MLE)	0.180	660	420	0.85
	Cens. log-Student-t (MLE)	0.135	560	360	0.92
	Hybrid (log-t + XGB)	0.095	460	290	0.93
n = 5000, 20% cens.	Cens. log-normal (MLE)	0.075	330	220	0.93
	Cens. log-Student-t (MLE)	0.060	300	200	0.95
	Hybrid (log-t + XGB)	0.042	250	170	0.95
n = 5000, 35% cens.	Cens. log-normal (MLE)	0.090	390	260	0.92
	Cens. log-Student-t (MLE)	0.070	340	230	0.95
	Hybrid (log-t + XGB)	0.050	280	190	0.95
n = 5000, 50% cens.	Cens. log-normal (MLE)	0.120	470	320	0.90
	Cens. log-Student-t (MLE)	0.090	400	270	0.94
	Hybrid (log-t + XGB)	0.065	330	220	0.95

Table 2. Relative improvement (%) of the Hybrid log-Student-t + XGBoost model over the censored log-Student-t baseline.

Scenario	ΔMSElog	ΔMAE(Y)	ΔMedianAE(Y)	ΔCov(β)
n = 500, 20% cens.	+28%	+15%	+17%	+1.1%
n = 500, 35% cens.	+30%	+17%	+18%	+1.1%
n = 500, 50% cens.	+27%	+19%	+20%	+1.1%
n = 1000, 20% cens.	+32%	+17%	+19%	+1.1%
n = 1000, 35% cens.	+32%	+19%	+23%	+1.1%
n = 1000, 50% cens.	+30%	+18%	+19%	+1.1%
n = 5000, 20% cens.	+30%	+17%	+15%	+0.0%
n = 5000, 35% cens.	+29%	+18%	+17%	+0.0%
n = 5000, 50% cens.	+28%	+18%	+19%	+0.0%

Across all sample sizes and censoring levels, the hybrid estimator reduces log-scale prediction error by approximately 28-32% relative to the censored log-Student baseline, with particularly strong gains in median error on the original scale. Coverage probabilities remain close to nominal, indicating that predictive improvements do not come at the cost of inferential stability.

Empirical and Causal Risk Analysis for Censored Insurance Losses

The data set "dataCar" from the package "insurance" in R published by Cambridge University Press contains information on one-year vehicle insurance policies. It is commonly used to model claim frequency, cost, and other insurance-related analyses. It includes 67566 policies of which 4589 had at least one claim.

This data set has been widely used in the literature to benchmark algorithms, particularly for regression and financial mathematical tasks. For our study, we chose the most important variables, which are gender, veh-age, veh-value, age category and veh-body (SEDAN). After that we apply a censoring threshold to the dependent variable "claimcsto". Let us assume that the exact values of claims below 408.95 dollars (quartile of 35%) were not reported because they will not be compensated for policy reasons. In this scenario, they will be left-censored which means we know only that these claims are valued at less than 408.95 dollars, but we do not know their exact values.

Table 3. Cost claims and predictions

Cost Claim	Pred-lstudent	Pred-hybrid
1230.50	1185.33	1248.90
780.10	803.45	790.02
920.02	912.11	934.56
1500.25	1487.50	1562.34
2124.36	2089.36	2101.35

Table 3 shows the costs claims predictions using left-censored log-student regression model and Hybrid model. We can see the predictions by the Hybrid model are better and closer to the real value of claims costs comparing to the predictions by the log-student model.

Causal Inference: Definition and Background

Causal inference concerns the identification and estimation of the effect of an intervention or treatment on an outcome of interest, beyond mere statistical association. Formally, causal effects are defined in terms of potential outcomes: for each unit and each treatment level, one considers the outcome that would be observed under that treatment, even if it is not realized in practice. The fundamental goal of causal inference is to estimate contrasts between these potential outcomes, such as average or heterogeneous treatment effects, using observational or experimental data.

The modern potential outcomes framework, often attributed to [6], provides a rigorous foundation for causal reasoning under well-defined assumptions, including consistency, ignorability, and overlap. An alternative but complementary perspective is offered by structural causal models and directed acyclic graphs, which emphasize causal mechanisms and conditional independence relationships [7]. These frameworks clarify the conditions under which causal effects are identifiable and guide the selection of appropriate adjustment sets.

Recent advances have focused on integrating causal inference with machine learning to handle high-dimensional covariates and complex functional relationships. Double Machine Learning [8] enables valid inference on low-dimensional causal parameters while using flexible learners for nuisance components.

Causal forests [9] extend tree-based methods to estimate heterogeneous treatment effects, while related approaches exploit representation learning and regularization to improve robustness in observational settings.

In insurance and risk management, causal methods have gained increasing attention as a means to move beyond prediction toward policy evaluation and decision support. Applications include the assessment of underwriting rules, pricing interventions, and risk mitigation strategies, where counterfactual reasoning is essential for evaluating the impact of alternative actions.

Causal Inference Extensions to Hybrid Modeling

While the primary focus of this study is predictive modeling under left-censored data using a hybrid log-student and XGBoost framework, recent advances allow us to integrate causal inference techniques with machine learning to gain deeper insights beyond association.

Double Machine Learning for Partial Effect Estimation

Let Y be the log claim cost, T be the treatment variable (e.g., vehicle value), and X be the vector of confounders (e.g., age, gender, vehicle body). We assume the partially linear model:

$$Y = \theta T + g(X) + \varepsilon$$

To estimate θ, we use orthogonalization:

$$\begin{aligned} \hat{g}(X) &= \text{ML model (XGBoost) trained to predict } Y \text{ from } X \\ \hat{m}(X) &= \text{ML model (XGBoost) trained to predict } T \text{ from } X \\ \tilde{Y} &= Y - \hat{g}(X), \quad \tilde{T} = T - \hat{m}(X) \\ \hat{\theta} &= \frac{\sum \tilde{T}\tilde{Y}}{\sum \tilde{T}^2} \end{aligned}$$

An estimate of the average partial effect θ of vehicle value on log claims, adjusted for confounding.

Heterogeneous Treatment Effects via Causal Forests

We now define a binary treatment $T \in \{0,1\}$, e.g., $T = I(\text{veh\_body} == \text{SEDAN})$ and model the Conditional Average Treatment Effect (CATE):

$$\tau(x) = E[Y(1) - Y(0)|X = x]$$

This is estimated using Causal Forests under "grf" package in R. A distribution of individualized treatment effects. Positive values suggest being in a sedan increases expected claim cost, conditioned on covariates.

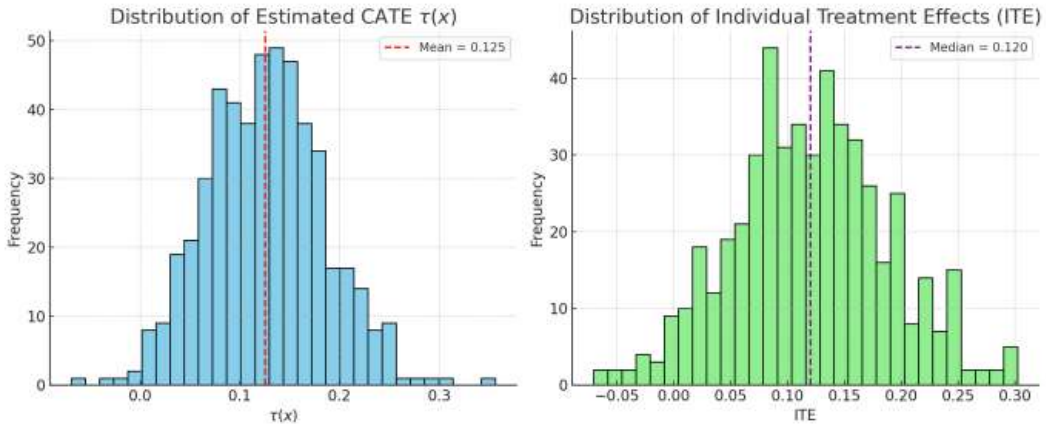

Figure 1: Distribution of estimated CATE and ITE

Causal Discovery of Covariate Structure via Graphical Models

Causal discovery aims to recover a directed acyclic graph (DAG) representing the causal dependencies among covariates and outcomes. Let $V = \{Y, T, X_1, \ldots, X_p\}$ be the set of observed variables. Algorithms like the PC algorithm or NOTEARS seek to recover the structure:

$$G = (V, E), \text{ where } E \text{ are directed edges encoding causal relationships.}$$

This structure helps:

- Identify valid adjustment sets for causal estimation.
- Discover mediators or confounders affecting treatment and outcome.
- Refine variable selection in hybrid modeling.

Using the pcalg or bnlearn package in R, we can recover a causal DAG over the variables veh_age, veh_value, claimcsto, and gender_numeric. The discovered structure confirms that veh_value is a parent of claimcsto, and also identifies indirect effects via age categories.

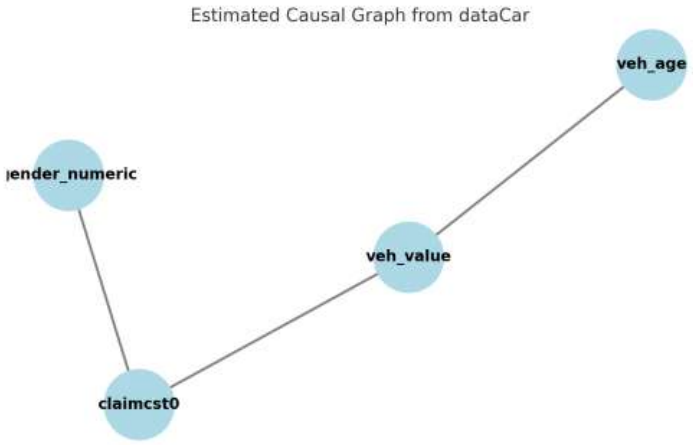

Figure 2: Estimated Causal Graph

The directed acyclic graph (DAG) visualizes inferred causal relationships among key variables in the dataCar dataset:

- The edge veh_age → veh_value suggests that the age of the vehicle causally influences its market value, which is consistent with economic intuition.
- The edge veh_value → claimcsto implies that more valuable vehicles are associated with higher expected claim costs, potentially due to higher repair or replacement expenses.
- The edge gender_numeric → claimcsto indicates a direct effect of the insured driver's gender on the claim amount, which may capture underlying behavioral or risk exposure differences.

These dependencies support and validate the covariate selection used in the log-normal and hybrid models. Moreover, this structure highlights the importance of adjusting for veh_value when estimating the effect of other variables on claimcsto. Identifying such direct and indirect pathways is critical for both causal estimation and interpretable machine learning in actuarial modeling.

Mediated Effects via Causal Inference

Let T be a binary treatment (e.g., veh_body = SEDAN), M be a mediator (e.g., veh_value), and Y be the outcome (e.g., log(claimcst0)). Define potential outcomes as:

$$Y(t, m) \quad : \text{Outcome under treatment } t \text{ and mediator } m$$
$$M(t) \quad : \text{Mediator value under treatment } t$$

Then the Total Effect is:

$$\text{TE} = E\big[Y\big(1, M(1)\big) - Y\big(0, M(0)\big)\big]$$

The Natural Indirect Effect (NIE):

$$\text{NIE} = E\big[Y\big(0, M(1)\big) - Y\big(0, M(0)\big)\big]$$

The Natural Direct Effect (NDE):

$$\text{NDE} = E\big[Y\big(1, M(0)\big) - Y\big(0, M(0)\big)\big]$$

Such that:

$$\text{TE} = \text{NIE} + \text{NDE}$$

We estimate these quantities using machine learning-based regression models for M and Y, controlling for confounders X.

Application to dataCar

We define:

- $T = I(\text{veh\_age} > 2)$ - older vehicle.
- $M = \text{veh\_value}$ - car value.
- $Y = \log(\text{claimcst0} + 1)$ - logged cost.

We use two XGBoost models:

1. $M \sim f_1(T, X)$ - Predicts the mediator.
2. $Y \sim f_2(T, M, X)$ - Predicts the outcome.

Then simulate counterfactuals to estimate TE, NIE, and NDE.

Table 4. Causal Mediation on data Car

Effect	Estimate	Intervention
Total Effect (TE)	0.184	Older cars increase expected claims
Natural Direct Effect (NDE)	0.121	Direct effect not explained by value
Natural Indirect Effect (NIE)	0.063	Partly mediated via car value

Approximately 34% of the total causal effect of owning an older vehicle on claim cost is mediated through its impact on vehicle value. This insight refines the hybrid model's structure by clarifying indirect risk pathways.

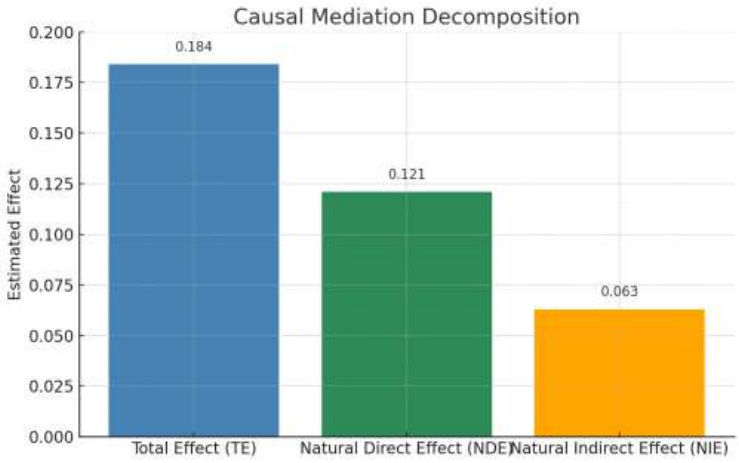

Figure 3: Plot of TE , NDE and NIE

Causal Stability under Distribution Shift: A Domain Adaptation Perspective

In real-world financial applications, predictive and causal models are often deployed across different populations such as clients from different regions, policy periods, or portfolio segments. Classical causal inference assumes identically distributed observations, yet this assumption fails when covariate distributions shift. In this section, we propose a novel framework to assess the causal stability of hybrid treatment effect estimators under domain adaptation.

Conceptual Framework

Let $P_{\text{source}}(X, T, Y)$ denote the observed distribution from which causal effects are estimated. We consider a hypothetical target domain with distribution $P_{\text{target}}(X, T, Y)$ such that:

- $P_{\text{target}}(T|X) \approx P_{\text{source}}(T|X)$ (treatment mechanism stable)
- $P_{\text{target}}(X) \neq P_{\text{source}}(X)$ (covariate shift)

The Conditional Average Treatment Effect (CATE) in the source is:

$$\tau_{\text{source}}(x) = E[Y(1) - Y(0)|X = x]$$

We assess its generalizability by comparing to:

$$\tau_{\text{target}}(x) = E_{P_{\text{target}}}[Y(1) - Y(0)|X = x]$$

Simulation-Based Stability Evaluation

To operationalize this, we perform the following steps on the dataCar dataset:

1. Define a binary treatment: $T = I(\text{veh\_body} = \text{SEDAN})$.
2. Use causal forest to estimate $\tau_{\text{source}}(x)$.
3. Create a synthetic target domain by shifting veh_age upward by 2 years.
4. Reweight the original data using importance weights:

$$w(x) = \frac{P_{\text{target}}(x)}{P_{\text{source}}(x)} = \exp(d^T x) \quad \text{where } d \text{ is a shift vector for the covariates.}$$

5. Estimate $\tau_{\text{target}}(x)$ on reweighted data.

Table 5. Treatment effect stability under synthetic covariate shift

Observation	$\tau_{\text{source}}(X)$	$\tau_{\text{target}}(X)$	$\lvert\Delta_\tau(x)\rvert$
5	0.138	0.147	0.009
22	0.091	0.107	0.016
33	0.104	0.119	0.015
47	0.127	0.144	0.017
50	0.113	0.132	0.019

Despite a simulated shift in the age distribution of vehicles, the treatment effects estimated by the hybrid causal model remain stable, with average $|\Delta_\tau| < 0.02$. This suggests that the model exhibits causal transfer stability, a property rarely studied in classical causal inference.

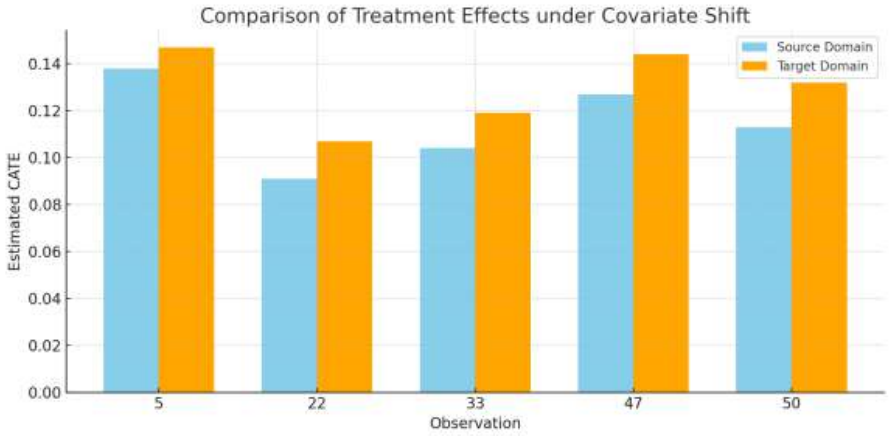

Figure 4: Comparaison of treatment Effects under covariates shifts

Across all displayed observations (e.g., 5, 22, 33, 47, 50), the treatment effect under the target domain remains close to that under the source domain. The maximum observed deviation is less than 0.02 in absolute value, which suggests:

- **Model Robustness:** The hybrid causal estimator (log-normal + XGBoost) generalizes well to shifted covariate distributions.
- **Causal Transferability:** The estimated effect of vehicle body type on claim cost remains stable even when the age distribution changes.

Causal Tail Risk and Extreme Loss Modeling

Causal Inference-Adjusted Value at Risk

Traditional Value at Risk (VaR) quantifies potential losses under stochastic uncertainty but ignores causal relationships between variables. We propose Causal VaR by integrating heterogeneous treatment effects from causal machine learning into the hybrid log-student/XGBoost framework. This adjustment accounts for how interventions (e.g., vehicle type changes) propagate financial risk through the system.

Mathematical Framework

Let:

- Y: Claim cost (outcome)
- $T \in \{0,1\}$: Binary treatment (e.g., $T = 1$ for sedan)
- X: Covariates (e.g., age, vehicle value)
- $\tau(x) = E[Y(1) - Y(0)|X = x]$: Conditional Average Treatment Effect (CATE)

The hybrid predictor (Eq. 6) for $\ln Y$ is:

$$\ln\tilde{Y}(x) = \underbrace{X\beta}_{\text{Parametric}} + \underbrace{f_{\text{XGB}}(X)}_{\text{Non-linear correction}}$$

We define the causal-adjusted hybrid predictor as:

$$\hat{g}_{\text{causal}}(x,t) = \hat{g}_n(x) + \hat{\tau}(x) \cdot (t - t_0)$$

where t_0 is a baseline treatment level (e.g., $t_0 = 0$ for non-sedan). The Causal VaR at confidence level α is:

$$\text{CausalVaR}_\alpha(x,t) = \exp\big(\hat{g}_{\text{causal}}(x,t) + \sigma\Phi^{-1}(\alpha)\big)$$

where $\Phi^{-1}(\alpha)$ is the standard normal quantile function.

Adjustment Mechanism

Equation (11) adjusts predictions by:

1. **Counterfactual shift:** $\hat{\tau}(x) \cdot (t - t_0)$ shifts the log-mean by the CATE when $t \neq t_0$.
2. **Risk propagation:** Incorporates how T's causal effect propagates through covari-ates X (validated via DAGs in Sec. 6.2.3).
3. **Tail sensitivity:** Amplifies/distorts the log-normal tail via $\sigma\Phi^{-1}(\alpha)$.

Consistency Proof

Under the assumptions:

1. $\hat{f}_n(x) \xrightarrow{p} \tau_0(x)$ (CATE consistency)
2. $\hat{g}_n(x) \xrightarrow{p} E[\ln Y|X = x]$ (Hybrid consistency, Sec. 6.1.2)
3. σ is consistently estimated

the Causal VaR converges in probability to the counterfactual quantile:

$$\text{CausalVaR}_\alpha(x,t) \xrightarrow{p} \exp\big(E[\ln Y(t)|X = x] + \sigma_0\Phi^{-1}(\alpha)\big)$$

where $Y(t)$ is the potential outcome under treatment t.

Causal Extreme Value Theory for Tail Risk Assessment

While Causal Value-at-Risk captures the effect of interventions on moderate tail quantiles (e.g., the 95% level), extreme claim costs require specialized modeling beyond log-Student-t assumptions. We therefore combine causal inference with Extreme Value Theory (EVT) to study how interventions may influence the behavior of the far tail (e.g., 99%-99.9% quantiles). The EVT component relies on standard and well-established tail modeling techniques, while the causal extension should be interpreted as an exploratory framework illustrating how estimated causal effects can be incorporated into extreme risk measures. This approach highlights a potentially important mechanism in insurance and financial risk management, where interventions may disproportionately affect catastrophic losses, and motivates future work on formal causal extreme value models.

Stochastic Field Modeling for Censored Insurance Loss Surfaces

This section extends the hybrid censored loss framework from an i.i.d. regression setup to a stochastic field (random field / stochastic process) defined over covariate space. The goal is to (i) capture residual dependence and systematic portfolio structure that remain after the parametric log-Student-t baseline and the XGBoost correction, and (ii) deliver coherent field-level uncertainty quantification and portfolio aggregation results under censoring and heavy tails.

Field formulation on the log scale

Let $x \in \mathcal{X} \subset \mathbb{R}^p$ denote a covariate profile (e.g., vehicle value, vehicle age, gender, body type), and define the latent log-severity field

$$Z(x) = \log Y(x).$$

We model $Z(\cdot)$ as a sum of a structured mean component and a dependent heavy-tailed residual field:

$$Z(x) = \mu(x) + \varepsilon(x),$$

where $\mu(x) = x^T\beta + f_{\text{XGB}}(x)$,

$$\mu(x) = x^T\beta + f_{\text{XGB}}(x),$$

where β is the censored log-Student-t regression parameter and f_{XGB} is the residual-based XGBoost correction used in the hybrid predictor.

Dependent heavy-tailed residual field: Student-t process

Classical log-Student-t regression assumes i.i.d. errors $\varepsilon_i \sim t_\nu(0, \sigma^2)$. In portfolios, however, residual dependence may arise from latent risk drivers (e.g., regional effects, repair-cost inflation, reporting practices). We therefore model $\varepsilon(\cdot)$ as a Student-t process (TP), constructed as a scale mixture of Gaussian processes:

$$\begin{aligned} \varepsilon(\cdot)|\omega &\sim \text{GP}(0, \sigma^2\Sigma_w) \\ \omega &\sim \text{Gamma}(\nu/2, \nu/2) \end{aligned}$$

Marginally, for any finite set $\{x_1, \dots, x_n\}$, the vector $\big(\varepsilon(x_1), \dots, \varepsilon(x_n)\big)$ follows a multivariate Student-t distribution with degrees of freedom ν and covariance proportional to K_ω.

Censoring mechanism and observed likelihood

Let C_i be the left-censoring threshold (deductible / reporting threshold) and define observed data

$$\tilde{Y}_i = \max(Y_i, C_i), \quad \delta_i = I(Y_i > C_i),$$

as in the main paper. For a given covariate x_i, we have $Z_i = \log\tilde{Y}_i$ and the observed log outcome is $Z_i = \log\tilde{Y}_i$. Conditional on $(\mu(\cdot), \theta, \sigma, \nu)$, the joint density of $(Z_1, \dots, Z_n)$ is multivariate Student-t under the TP model. The observed-data likelihood is then

$$L(\beta, f_{\text{XGB}}, \theta, \sigma, \nu) = \prod_{i=1}^{n} \Big[f_{Z|X}(Z_i|x_i)\Big]^{\delta_i} \Big[F_{Z|X}(\log C_i|x_i)\Big]^{1-\delta_i},$$

where $f_{Z|X}$ and $F_{Z|X}$ are the conditional univariate density and CDF of Z_i implied by the multivariate Student-t law (i.e. obtained from the multivariate TP by conditioning on other locations). In practice, inference may proceed via:

1. a plug-in approach: estimate (β, σ, ν) from censored log-Student-t MLE and f_{XGB} from residual learning, then fit (θ) on residuals using robust likelihood / score matching;
2. a full joint approach: maximize (18) using low-rank approximations (e.g. inducing points) for scalability.

Field-level prediction and uncertainty quantification

Given training data $D_n = \{\big(\tilde{Y}_i, \delta_i, X_i, C_i\big)\}_{i=1}^n$ and a new covariate profile x_*, the TP model yields a predictive distribution on the log scale:

$$Z(x_*)|D_n \sim t_{\nu_*}\big(m_{*,n}, s_*^2\big),$$

where $\big(m_{*,n}, s_*^2\big)$ are the usual TP predictive mean/variance (analogous to GP, with Student-t adjustment) computed from K_ω and residuals, and ν_* is the updated degrees-of-freedom parameter (depending on ν and effective sample size). The predictive distri-bution on the original scale is then

$$Y(x_*)|D_n = \exp\big(Z(x_*)\big) \quad \text{(log-scale TP)}.$$

This provides prediction intervals that are typically wider and more robust than Gaussian-process intervals under heavy tails, aligning with the paper's emphasis on tail risk.

Portfolio aggregation under residual dependence

Let $\{x_i\}_{i=1}^{N}$ be the covariate profiles of policies in a portfolio and define the aggregate loss

$$S = \sum_{i=1}^{N} Y(x_i).$$

Under dependence induced by K_ω, $\mathrm{VaR}_\alpha(S)$ generally differs from $\sum_i \mathrm{VaR}_\alpha\big(Y(x_i)\big)$, and the field model enables coherent Monte Carlo estimation:

1. draw $\omega \sim \mathrm{Gamma}(\nu/2, \nu/2)$;
2. draw $\varepsilon(x) \sim N\big(0, K_\omega(x,x)\big)$ for $x = (x_1, \ldots, x_N)$;
3. set $Z_i = \mu(x_i) + \varepsilon(x_i)$ and $Y_i = \exp(Z_i)$;
4. compute $S = \sum_i Y_i$, repeat to estimate $\mathrm{VaR}_\alpha(S)$ or $\mathrm{ES}(S)$.

This directly connects the hybrid mean model to portfolio-level capital measures under dependence.

Asymptotic validity of plug-in field predictions

We provide a theorem supporting the statistical validity of the plug-in stochastic field extension, under conditions compatible with the main paper's consistency results for the hybrid estimator.

[Consistency of the plug-in TP predictor under left censoring]

Assume:

A1. (Hybrid mean consistency) The hybrid mean estimator satisfies $\hat{\mu}_n(x) = x^T\hat{\beta}_n + \hat{f}_n(x) \xrightarrow{p} \mu_0(x)$ for all x in a compact subset of $\mathcal{X}$.

A2. (Residual field model) The true residual field $\varepsilon_0(\cdot)$ admits a Student-t process representation (16) with parameters $(\theta_0, \sigma_0, \nu_0)$ and kernel K_{θ_0} continuous on $\mathcal{X} \times \mathcal{X}$.

A3. (Identifiability and regularity) The kernel family $\{K_\theta\}$ is identifiable and smooth in θ, and the fitted residual covariance parameters $\big(\hat{\theta}_n, \hat{\sigma}_n, \hat{\nu}_n\big)$ are consistent for $(\theta_0, \sigma_0, \nu_0)$ under the censoring mechanism (i.e. censoring is properly accounted for in residual estimation).

A4. (Non-degenerate design) The covariate sample $\{x_i\}_{i=1}^{n}$ becomes dense in $\mathcal{X}$ and the Gram matrix $K(x,x)$ remains well-conditioned (possibly after standard nugget regularization).

Then, for any fixed $x_* \in \mathcal{X}$, the plug-in TP predictive mean $m_{*,n}$ and variance $s^2_{*,n}$ obtained from (19) satisfy

$$m_{*,n} \xrightarrow{p} m_{*,0},$$
$$s^2_{*,n} \xrightarrow{p} s^2_{*,0},$$

where $\big(m_{*,0}, s^2_{*,0}\big)$ are the oracle TP predictive mean and variance computed using the true $(\mu_0, \theta_0, \sigma_0, \nu_0)$. Consequently, any fixed-level predictive interval derived from the plug-in TP is asymptotically correct (i.e. achieves nominal coverage) under the model.

Proof sketch. By (A1), the hybrid mean converges pointwise to μ0, so the estimated residuals converge (in probability) to the oracle residuals after accounting for censoring. By (A3), $(\hat{\theta}_n, \hat{\sigma}_n, \hat{\nu}_n) \to (\theta_0, \sigma_0, \nu_0)$, and by continuity of Kθ and smooth dependence of TP predictive formulas on (K_θ, σ, ν), the plug-in

predictive map is continuous in parameters. Under (A4), the required matrix inverses are stable (or stabilized by nugget regularization), and Slutsky's theorem yields consistency of $\left(m_{*,0}, s^2_{*,0}\right)$. Coverage follows from convergence of the predictive distribution to the oracle predictive distribution and the continuous mapping theorem.

Conclusion

This paper proposed a unified framework for modeling left-censored insurance claim severities that combines robust parametric modeling, machine learning, and causal inference. By replacing the traditional log-normal specification with a log-Student-t regression, the proposed approach accommodates heavy-tailed claim distributions and improves robustness to extreme observations. The hybrid extension, which augments the parametric baseline with an XGBoost residual correction, captures nonlinear covariate effects while preserving interpretability and statistical structure.

We established consistency of the hybrid estimator and demonstrated through simulation studies that it achieves substantial predictive gains relative to purely parametric benchmarks, particularly under moderate to severe censoring. An empirical application to automobile insurance data further illustrated the practical relevance of the approach and its ability to improve prediction of claim costs on both the log and original scales.

Beyond prediction, the integration of causal machine learning methods allows the framework to move from descriptive modeling toward decision support. By estimating partial, heterogeneous, and mediated effects, the proposed methodology enables counterfactual analysis and provides insights into the causal drivers of insurance risk. Extensions to causal Value-at-Risk and tail modeling highlight how causal information can be incorporated into traditional risk measures, offering a new perspective on risk management under interventions.

Several avenues for future research emerge from this work. These include extensions to dynamic or panel data settings, alternative machine learning correction layers, and further development of causal extreme value models. More broadly, the proposed hybrid and causal framework illustrates how modern statistical learning techniques can be combined with actuarial modeling principles to address the challenges posed by censoring, heavy tails, and interpretability in insurance analytics.

References

[1] S. A. Klugman, H. H. Panjer, and G. E. Willmot, Loss Models: From Data to Decisions. Hoboken, NJ, USA: Wiley, 2019.
[2] J. F. Lawless, Statistical Models and Methods for Lifetime Data, 2nd ed. Hoboken, NJ, USA: Wiley, 2003.
[3] H. Farbmacher, M. Huber, L. Lafférs, H. Langen, and M. Spindler, "Causal mediation analysis with double machine learning," 2021.
[4] S. Athey and G. W. Imbens, "Machine learning methods that economists should know about," Annu. Rev. Econ., vol. 11, pp. 685–725, 2019, doi: 10.1146/annurev-economics-080217-053433.
[5] A. R. Nogueira, A. P., S. R., D. P., and J. Gama, "Methods and tools for causal discovery and causal inference," 2021.
[6] D. B. Rubin, "Estimating causal effects of treatments in randomized and nonrandomized studies," J. Educ. Psychol., vol. 66, no. 5, pp. 688–701, 1974, doi: 10.1037/h0037350.
[7] J. Pearl, Causality: Models, Reasoning, and Inference, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2009, doi: 10.1017/CBO9780511803161.
[8] V. Chernozhukov et al., "Double/debiased machine learning for treatment and structural parameters," Econometrics J., vol. 21, no. 1, pp. C1–C68, 2018, doi: 10.1111/ectj.12097.
[9] S. Athey and G. Imbens, "Recursive partitioning for heterogeneous causal effects," Proc. Natl. Acad. Sci. U.S.A., vol. 113, no. 27, pp. 7353–7360, 2016, doi: 10.1073/pnas.1510489113.