Understanding the relation between extreme operational risk losses and economic categorical covariates to improve associated capital charge estimates

ERASMUS UNIVERSITY ROTTERDAM
Erasmus School of Economics
Master Thesis Econometrics and Management Science

Name student: Constantijn de Jonge
Student ID number: 431331

Supervisor: prof. dr. Chen Zhou
Second assessor: dr. Carlo Cavicchia

Date final version: July 31, 2022

Abstract

We investigate the relation between extreme operational risk (OpRisk) losses and economic categorized factors. Such a relation is recently investigated by Groll et al. (2019) using the *gamboostLSS* (Mayr et al. (2012)) gradient and log-likelihood boosting optimizer. Our simulation study shows that this optimizer, although the log-likelihood iteratively improves, associated in- and out-of-sample parameter estimates deteriorate after a bliss number of iterations. Therefore, stopping criteria must be investigated. Furthermore, we introduce a robust *gamboostLSS* alternative to stabilize model term updating schemes since the scale parameter of the Generalized Pareto Distribution (GPD) dominates classical log-likelihood-type updating procedures. We apply our methodology in a real-world application using UniCredit's (UC) OpRisk loss data and twenty economic factors. We identify high UC's Tier-I capital ratios, low Italian unemployment rate (UR IT), and high United States' financial market volatility as key drivers for the loss severity, whereas high Milano Italia Borsa returns, high UC's deposit growth rate, and low UR IT drive the loss frequency. We use the categorized risk drivers in a scenario analysis and find that these factors are able to explain differences in total losses among economic conditions of the business cycle. Associated capital charges are within a fixed set of 720 charges and hence banks could reduce liquidity risk by the use of our methodology.

Keywords: Basel IV, categorical covariates, GAMLSS, gradient boosting optimizers, L_1-norms, operational risk, requested capital charge.

Contents

1. Introduction.

Catastrophic losses in financial portfolios led to the default of many financial institutions during the 2007-2010 subprime mortgage crisis. A notorious example is the fourth-largest investment bank of the United States, Lehman Brothers. These extreme losses partially originated from Operational Risk (OpRisk). Under Basel IV, banks currently have to meet an OpRisk capital charge using a simple calculation because OpRisk is extremely difficult to model due to its undiversifiable and unexpected nature. Nevertheless, in this thesis we explore the possibility to model extreme OpRisk losses.

We investigate the relation between extreme OpRisk losses and economic categorized factors. Such a relation can help to identify OpRisk drivers and to improve capital charges. We add to the literature by examining capital charges for models that include economic categorized factors. Furthermore, we examine these charges for two conditions of the business cycle, namely crises and expansions. Moreover, we examine current gradient boosting optimizers for Generalized Additive Models for Location, Shape and Scale (GAMLSS; Rigby and Stasinopoulos (2005)), *GAMLSSBoostAlt*. We benchmark the best performing Groll et al. (2019) models, detailed in Section 4.2.4, since they did a similar study.

The Basel Committee for Banking Supervision (BCBS) defines OpRisk as 'The risk of loss resulting from inadequate or failed internal processes, people and systems or from external events.' (BCBS (2004)), and Basel II lists seven risk categories, also known as event types, that reflect all OpRisk causes. For future reference, we refer to the associated realized losses as OpRisk losses or simply OpRisks. We estimate total OpRisk losses over a time frame by the loss severities, which are the financial values of losses; and the loss frequencies, which are the probable number of losses over time. We focus on the loss severities since they drive total OpRisks (Hambuckers et al. (2018)). OpRisk loss data is scarce and hence non-parametric methods, such as historical simulation, almost surely lead to huge estimation uncertainty. We therefore use extreme value theory (EVT; Embrechts et al. (2013)) techniques to parametrically model the tail of a loss series.

We apply the peaks-over-threshold (POT) method, which only considers losses above a threshold. The resulting losses, loss severity exceedances (LSEs), are obtained by subtracting a threshold from the loss severities above this threshold. Loss frequency exceedances (LFEs) are the associated number of losses above a threshold. POT does not consider losses under a threshold since these losses do not reflect the tail of loss series. LSEs approximately follow a scaled generalized Pareto distribution GPD(ξ, σ), with a shape (ξ) and scale (σ) parameter, for a sufficiently high threshold (Balkema and de Haan (1974)). Both GPD parameters have their own linear predictor as we use linear relations between parameters and covariates. We use continuous (metric) and categorical covariates, such as risk categories and economic factors, to model OpRisk losses.

The use of categorical covariates has clear benefits for banks: since we estimate one parameter for each non-reference covariate group, the continuously changing capital charge under metric covariates is reduced to a fixed set of capital charges under the categorical covariate representation. The requested capital charges do not change when data in the next time period fall into identical groups as the current time period. If new economic data do fall in other groups, then these groups could actually be fused by fused L_1-norms (later introduced) and hence capital charges remain identical. Even if new economic data fall into completely different groups, for example when there is a shift in economic condition, then the new capital charge is still within the set of fixed and known capital charges. With regard to the economic factors that we use: the size of OpRisk losses largely depend on the availability of money, which usually is higher during expansions. Levels of economic factors reflect the current condition of the business cycle. Therefore, economic factors proxy expansions, and expansions drive extreme OpRisk losses. This latter is supported by the scenario analysis of Hambuckers et al. (2018).

The GPD distributed linear predictors depend on covariates, and are modelled by GAMLSSs. However, regular GAMLSS estimation procedures are unstable when many parameters need to be estimated. Statistical models frequently overfit data in the training set (Tibshirani (1996)), and we require variable selection methods to identify risk drivers. For that purpose, we apply regularization. We use L_1-norms, which penalize model coefficients by penalty functions, whereby tuning (penalty) parameters control the strength of shrinkage. L_1-norms apply uncontrolled shrinkage. This means that coefficients can immediately shrink to zero without any intermediate steps. We use adaptive weights to apply 'counter-pressure' to L_1-norms to ensure stable estimation procedures. We use gradient boosting optimizers to obtain adaptive weight estimates. As boosting optimizers, we introduce a derivative of the well-known *gamboostLSS* optimizer (Mayr et al. (2012)) to improve model parameter estimates. Moreover, penalty functions introduce biases in parameter estimates by construction. These biases could be controlled by adaptive weights.

In our simulation, we shed light on the performance of gradient and log-likelihood boosting estimation procedures. Our simulation study shows that the GPD scale parameter drives extreme OpRisk losses, and that maximizing log-likelihoods yields inferior parameter estimates. We therefore recommend investigating stopping criteria for these type of optimizers. Classical stopping criteria, such as Akaike Information Criteria (AIC; Akaike (1974)) and Bayesian Information Criteria (Bayesian Information Criteria (BIC; Schwarz (1978)) do not work since log-likelihood functions (LLFs) are involved. The level of LLFs largely depend on the level of the OpRisk losses and hence proper scaling procedures are required to proportionalize LLF-levels and loss-levels. Moreover, we alternatively update both GPD parameters and find that this slightly improves model performance.

We examine the performance of our proposed models in a real-world application in which we use UniCredit's OpRisk losses, which are categorized in seven risk categories, as well as twenty economic factors. UniCredit is the largest Italian bank and is classified as a global systemically important bank by the Financial Stability Board (latest update in 2021). This means a default of UniCredit could disrupt the global financial system and hence it is highly relevant to understand the drivers of their extreme OpRisk losses. We identify UniCredit's Tier-I capital ratios, UR IT, and a USA financial market volatility for the loss frequencies. For the loss severities, we detect UniCredit's deposit growth rate, UR IT, and the Milano Italia Borsa index. However, the joint effect of the two GPD parameters is not always clear. For example, when the GPD shape (scale) parameter increases (decreases), then we are left with a difficult question: "Does the increase in ξ estimates outweigh the decrease in σ estimates, or vice versa?" Lastly, all models in our study are identifiable and hence we never had to use L_2-norms to make models identifiable.

Our research has implications for the banking industry. We conduct a scenario study, and find that the aforementioned OpRisk risk drivers are able to explain differences in conditions of the business cycle. The associated scenario-specific and risk-category-specific parameter estimates are used to simulate losses, and hence we attain requested capital charge estimates for each scenario and risk category. Using our approach, banks can reduce their liquidity risk since the capital charge always is in a set of 720 charges, compared to the set of infinite charges under the metric covariate representation. When new quarterly economic factor information comes in, we check in which covariate groups these factor values fall, and the associated set of all covariate groups corresponds to one of the 720 covariate sets. Consequently, the new capital charge is equal to one of the 720 existing capital charges.

Section 2 contrasts our ideas with the literature. Sections 3-5 consider our data, methodology, and results, respectively. Section 6 concludes and discusses.

2. Related work.

Chavez-Demoulin et al. (2016) model (semi-)parametric LFEs and LSEs, and they include risk categories and time as covariates. We work with fully parametric models, and we use risk categories and economic factors as covariates, similar to Hambuckers et al. (2018) and Groll et al. (2019). Time is integrated in the GAMLSS model class. The last two papers use metric and categorical covariates, respectively. By contrast, we focus on categorical covariates, whereas metric covariates serve for robustness testing. Groll et al. (2019) simulate eight independent covariates in their simulation study, whereas their real-world application includes twenty highly comoving economic factors. We simulate up to fifteen highly comoving covariates in our simulation study to reproduce characteristics of real-world economic factors, although we focus on eight covariates due to computational reasons.

Similar to Hambuckers et al. (2018) and Groll et al. (2019), our research focuses on the underparametrized regime $p < n$, meaning that the number of model parameters p is lower than the number of observations n. We apply a modelling procedure that uses both a gradient boosting optimizer and regularization technique in a two-step approach. The link between the two is investigated by Hastie et al. (2022) with a focus on the proportional and overparametrized regimes $p \asymp n$ and $p > n$, respectively.

Hambuckers et al. (2018) use L_1-norms for metric covariates, whereas Groll et al. (2019) use 'fused L_1-norms' and group least absolute shrinkage and selection operator (LASSO) L_2-norms for categorical covariates. We focus on L_1-type norms. Fused LASSO (Tibshirani et al. (2005)) for metric covariates is the sum of two components: L_1-norms for coefficient dissimilarities and individual coefficients. Groll et al. (2019) extend the first component, 'fused L_1-norms', for categorical covariates. The second component is not considered because parameters of dummy encoded categorized covariates are undefined in the one-dimensional parameter space. Group LASSO L_2-norms shrink either all or none coefficients of groups within a dummy encoded categorical covariate to zero. If the coefficients of the risk categories would be shrunk to zero, then the associated capital charge estimates would be almost identical for all risk categories. However, the characteristics of the UniCredit losses, shown in Table 2, illustrate clear differences in the levels of the losses among risk categories. We therefore do not consider Group LASSO L_2-norms.

Groll et al. (2019) use a gradient boosting *gamboostLSS* optimizer to estimate unpenalized GAMLSSs. However, this optimizer maximizes the associated log-likelihood function (LLF) and thus does not concern about attaining low-biased parameter estimates for both GPD parameters. Hence, we simultaneously update model terms of the two GPD parameters by our own optimizer, namely *GAMLSSBoostAlt*.

Similar to Groll et al. (2019), we use well-specified models in our simulation study. The term well-specified means that the assumed probability distribution in fact is the true, unknown probability distribution. We do not consider misspecified models since our research is based on the assumption that LSEs follow a scaled GPD.

Similar to Hambuckers et al. (2018), we conduct a scenario analysis for conditions of the business cycle in which we simulate losses to subsequently estimate capital charges. Hambuckers et al. (2018) use an identical set of covariates for the loss frequency as well as the severity, whereas we use different covariate sets. Furthermore, we use categorical covariates for the loss severity in contrast to Hambuckers et al. (2018) who use metric covariates in their scenario analysis.

3. Data.

We describe UniCredit's OpRisk losses, motivate our choice for the threshold, and present the economic covariates that we use in our real-world application.

3.1. *Losses.*

UniCredit is an Italian international banking group and currently the world's 34th largest bank by total assets. They publicly published their OpRisk data set consisting of 40,871 quarterly losses above €2000, divided into seven Basel II recognized risk categories, shown in Table 1.

Table 1: Risk categories of the UniCredit loss data set, type and description.

Risk category	Description
BDSF	Business disruptions and system failures
CPBP	Clients, products and business practices
DPA	Damage to physical assets
EDPM	Execution, delivery and process management
EFRAUD	External fraud
EPWS	Employment practices and workplace safety
IFRAUD	Internal fraud

All losses are rescaled for privacy reasons and hence we can not draw conclusions with regard to the level of parameter estimates. Nevertheless, the relation between the losses and the economic factors remain unchanged. The losses can be downloaded from the data archive of the Journal of Applied Econometrics.[1] All OpRisks occurred between January 2005 and June 2014, and are categorized into 38 quarterly time periods. For illustration, OpRisk losses include losses due to fraud, unintentional failure to meet professional obligations and failed transaction processing, detailed in Appendix A.1. Activities in Italy and Germany account for approximately 50% and 20% of their revenue streams. Table 2 shows that CPBP and EDPM account for approximately 72% of all losses. The losses above our threshold (explained in the following section) capture approximately 87% of the total losses. CPBP and IFRAUD exhibit the highest mean, standard deviation, median and third quartile and thus would be dominant if one global threshold would be applied. Additionally, the standard errors show lower uncertainty when machines and/or processes, BDSF and DPA, are involved, in contrast to people.

3.2. *Threshold selection.*

We use risk-category-specific thresholds τ_e so that all seven risk categories are well-represented after applying the POT method. We set τ_e equal to the em-

[1] Source: http://qed.econ.queensu.ca/jae/2018-v33.6/hambuckers-et-al/

Table 2: Descriptive statistics of UniCredit's losses for 1) No threshold 2) threshold. The mean, standard deviation, median and third quartile are in thousands ($\times 10^3$), and the total losses are in millions ($\times 10^6$).

Risk category	Mean	Std	Median	Third quartile	Skewness	Number of losses	Total losses
No threshold							
BDSF	15.7	43.0	4.82	11.1	7.66	674	9.25
CPBP	77.5	1,587.2	10.37	27.1	48.7	16,138	1,218
DPA	10.9	70.5	3.46	5.73	23.4	896	8.00
EDPM	40.3	495.0	5.43	12.7	40.5	13,209	506.1
EFRAUD	21.9	490.3	4.78	10.6	73.5	6,391	127.4
EPWS	39.0	343.5	7.00	19.5	37.6	2,292	85.0
IFRAUD	151.8	1,143.3	13.6	53.0	21.3	1,271	190.5
All	54.5	1,077.9	6.76	18.6	64.9	40,871	2,145
Threshold							
BDSF	37.7	76.9	10.3	46.5	4.00	169	7.91
CPBP	255.8	3,166.6	29.4	105.5	24.4	4,035	1,133
DPA	28.2	138.7	3.25	15.5	11.8	224	7,16
EDPM	133.5	982.6	15.6	61.8	20.4	3,302	476.2
EFRAUD	64.0	979.0	15.3	46.5	36.8	1,598	122.3
EPWS	117.2	678.0	29.8	100.6	19.1	573	77.2
IFRAUD	515.0	2,237.0	90.4	350.6	10.8	318	180.0
All	178.0	2,147.8	21.2	84.3	32.5	10,219	1,998

Table 3: Economic factors, categorized by their data types. The term "EU" refers to all countries in the European Union.

Firm-specific	Macroeconomic	Financial
Risk category (RC)	EU / Italian growth rate (GDP EU / GDP IT)	10-year Italian government bond yield (LIR IT)
Deposit growth rate (DGR)	EU housing prices growth rate (HPI EU)	FTSE MIB index returns (MIB IT)
Leverage ratio (LR)	EU / Italian consumption loan rate (LOR EU / LOR IT)	3-month Italian interbank rate (SIR IT)
Revenue coming from fees (PRF)	EU monetary aggregate M1 growth rate (M1 EU)	S&P 500 returns (SP USA)
Tier I capital ratio (TCR)	EU / Italian unemployment rate (UR EU / UR IT)	TR EU Stock index returns (TRSI EU)
UniCredit stock returns (UCSR)		S&P based volatility index (VIX USA)
		FTSE 100 based volatility index (VFTSE UK)

pirical third quartile of the risk-category-specific unconditional loss series because Hambuckers et al. (2018) find a right balance between low variance of GPD coefficients and correct model specification for this particular threshold. In practice, we work with less than 25% of all losses since UniCredit only collected losses above €2000. Groll et al. (2019) did a similar study using a threshold that is based on a normality test of quantile residuals. We compare their and our choice for threshold in Appendix A.9.

3.3. *Covariates.* We use twenty economic factors, constructed by Hambuckers et al. (2018). These factors are firm-specific, macroeconomic and financial market products, summarized and more extensively detailed in Table 3 and Appendix A.2, respectively. We one-period lag economic factors to tackle two-way causality and to construct a convenient prediction framework. Credible unit root tests to investigate nonstationarity can not be performed as we have only 38 time periods, although Hambuckers et al. (2018) find that regularization adequately deals with nonstationarity.

Appendix A.3 shows the presence of multicollinearity among economic factors, which is the situation in which covariates are highly linearly related. Multicollinearity is to be expected since the levels of economic factors largely depend on the same underlying, that is the current

Table 4: Comparison of expected and realized number of observations (in %) after applying the POT method for our threshold choice.

Data type	Factor	Impact	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
Firm-specific	DGR	+	0.52	-5.6	-2.95	10.93	-4.15	0.79
	LR	+	-8.33	0.47	-1.77	8.40	4.09	NA
	PRF	+	0.19	-0.51	-3.06	0.76	11.02	NA
	TCR	*	-3.08	7.80	13.44	-2.63	-16.52	NA
	UCSR	+	1.90	-7.90	-9.92	4.49	-5.93	9.34
Macroeconomic	GDP EU	+	-7.90	-8.35	-1.31	11.16	3.73	-3.54
	GDP IT	+	-12.63	-1.77	6.57	10.73	-2.13	NA
	HPI EU	+	-1.43	-11.00	-3.62	0.56	8.06	12.78
	LOR EU	-	-7.18	-7.75	6.40	4.77	-10.10	NA
	LOR IT	-	-4.03	-10.73	4.21	4.76	0.61	NA
	M1 EU	+	-11.49	-3.10	-0.09	6.77	9.62	NA
	UR EU	-	-5.81	3.82	7.82	1.38	-5.70	-16.52
	UR IT	-	4.91	1.21	3.74	1.41	-13.93	-17.42
Financial market	LIR IT	-	6.23	4.60	-8.16	0.24	-13.18	NA
	MIB IT	+	-4.53	-10.93	1.67	1.41	9.34	NA
	SIR IT	-	-10.19	1.69	-3.51	7.34	2.07	-0.13
	SP USA	+	-4.38	-7.48	-1.99	8.24	4.83	-2.18
	TRSI EU	+	-6.21	-8.04	-2.51	10.31	1.70	NA
	VIX USA	*	0.54	13.06	1.43	-2.42	-11.1	-1.62
	VFTSE UK	*	12.78	-3.06	7.86	-4.79	-5.60	-5.01

Note: The table elements are calculated as: $\big(n_{C,df}^{\text{POT}} - \mathbb{E}(n_{C,df}^{\text{POT}})\big)/\mathbb{E}(n_{C,df}^{\text{POT}})$, where $n_{C,df}^{\text{POT}}$ and $\mathbb{E}(n_{\text{C,df}}^{\text{POT}})$ are the realized and expected number of observations for a categorized economic factor C and categorical covariate group df after applying the POT for our threshold choice. Positive values mean that the realized number of losses above our threshold is higher than expected. The expectations are calculated as $\mathbb{E}(n_{C,df}^{\text{POT}}) = n_{\text{C,df}}/4$, where $n_{\text{C,df}}$ is the number of observations for covariate C and group df before applying the POT method. Elements of the column “Impact” are determined using economic reasoning and hence are independent of the values in this table. Elements + (-) in column “Impact” mean that high values of the corresponding economic factor indicate an expansion (recession), and elements * mean that low values indicate an intermediate economic condition, whereas high values can indicate either an expansion or recession. Lastly, “NA” means not applicable since not all covariates have a sixth covariate group.

condition of the business cycle. The average absolute cross-correlation among covariates is .21. Correlations above and below .9 and -.8 are many, and we have extremes .97 and -.93 between financial indices VIX USA - VFTSE UK and UR EU - SIR IT. Financial indices are well-known to comove. UR EU and SIR IT are the EU unemployment rate and the Italian 3-month interbank rate, which is the price of short-term borrowing. The unemployment rate is low during expansions, in which regulators often increase interest rates to counteract the effects of the economic cycle, also known as counter-cyclical fiscal policy. Another perspective is that multicollinearity indicates redundant covariate information, which we tackle by regularization.

We dummy encode economic categorical covariates using interval limits, whereby DPA and lower-bound groups of economic factors serve as reference groups. Details about dummy encoding procedures, and the number of groups within categorical covariates are in Appendix A.4. To shed light on the impact of our choice for threshold on economic factors, Table 4 shows the difference between the realized and expected number of losses after applying the POT method. We use economic reasoning to determine the impact of economic factors on extreme OpRisks, shown in Column “Impact”, as we presume that economic factors proxy expansions, and that expansions increase both the severity and likelihood of extreme OpRisks losses. Therefore, we expect groups 1 and 2 to contain fewer observations than groups 5 and 6 after applying the

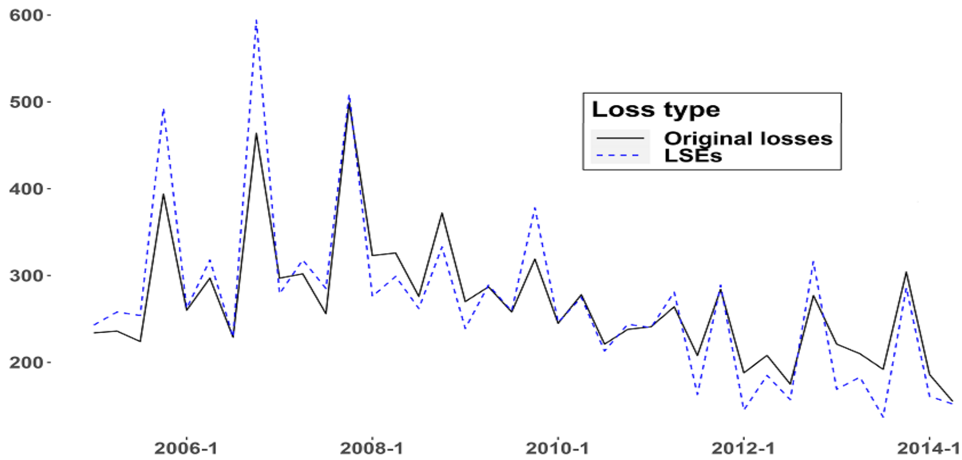


Figure 1: Number of scaled original UniCredit OpRisk losses, with a scaling factor of 1/4, and LSEs over time.

POT method when the impact of a factor is positive (+). All in all, using economic reasoning to determine the impact of economic factors is a valid approach since Table 4 demonstrates that our expectation of the impact of economic factors matches the associated economic factor patterns from group 1 to group 5/6 as we illustrate in the subsequent paragraph. We therefore are able to use economic reasoning to interpret model coefficients after model estimation.

The value -16.52% for group 5 of UR EU in Table 4 demonstrates that relatively few extreme OpRisk losses occurred during periods of high Italian unemployment rates. This is in line with our expectation that high UR IT proxies a recession, and a recession does not drive extreme OpRisks. From group 1 to group 6, UR EU differences on average decreases. Therefore, low UR EU values are linked to an increase in OpRisks. Furthermore, the pattern from group 1 and group 6 is not linear since the group differences are more extreme for the lower- (3.83–5.81) and upper-bound groups (-16.52–5.70) are more extreme than the differences for groups 2 to 4. Therefore, categorization of factors is highly informative since metric covariates implicitly assume linear covariate effects (effects between the covariate and the parameter). Similar findings hold for most other covariates. However, group 3 of TCR contains +13.44% more extreme OpRisk losses than expected. This means that relatively many extreme OpRisk losses occurred during time periods of moderate Tier-I capital ratios. Groll et al. (2019) explain that this could be due to the fact that banks reserve more capital when high future losses are anticipated. High negative LOR EU values for groups 1 and 5 indicate that relatively many extreme losses happened for moderate consumption loan rates in the European Union. Lastly, volatility in financial indices (VIX USA) is low for intermediate economic conditions, whereas VIX USA is high during both expansions and recessions.

Figure 1 shows that the frequency of losses on average slightly decrease over time. The peak frequency of losses are during the 2007-2010 financial crisis. Furthermore, a cyclical pattern, which can be linked to conditions of the business cycle, is present. The figure additionally illustrates that the number of losses above our threshold and the expected number of losses are quite similar. Therefore, our homogeneous threshold is valid and there is no evidence that indicates that the threshold should be time-varying.

4. Methodology.

We specify and estimate (un)penalized GAM(LSS)s, introduce the simulation setup, and assess model quality.

4.1. *Model specification.*

We have $T = 38$ time periods, whereby $t = 1, ..., 38$ is a single time period. We denote by $N_t(e)$ and $Z_{i,t}(e)$ the LFEs and loss severities for a risk category e in time period t above a threshold τ_e, where index i is the ith loss. Observed equivalents are $n_t(e)$ and $z_{i,t}(e)$. LSEs for a pair $\{e, t\}$ are $y_{i,t}(e) = z_{i,t}(e) - \tau_e$ if $z_{i,t}(e) - \tau_e \geq 0$, and collected by a $(n_t(e) \times 1)$ vector $\boldsymbol{y}_t(e) = [y_{1,t}(e), ..., y_{n_t(e),t}(e)]^\top$. We denote by $n_t = \sum_{e=1}^{7} n_t(e)$ and a $(n_t \times 1)$ vector $\boldsymbol{y}_t = [\boldsymbol{y}_t(1)^\top, ..., \boldsymbol{y}_t(7)^\top]^\top$ the LFEs and LSEs of all seven risk categories in time period t, where $\boldsymbol{y}_t$ consists of (n_t) elements y_{i_t} for $i_t = 1, ..., n_t$. We concatenate the LSEs over T time periods. We denote by $n = \sum_{t=1}^{T} n_t$ and a $(n \times 1)$ vector $\boldsymbol{y} = [\boldsymbol{y}_1^\top, ..., \boldsymbol{y}_T^\top]^\top$ the LFEs and LSEs over T time periods. The vector $\boldsymbol{y}$ consists of (n) elements y_i for $i = 1, ..., n$. We have a $(n \times p)$ model matrix $\boldsymbol{X}$ with matching n rows, where p is the total number of metric and dummy encoded categorical covariates.

We split $\boldsymbol{X} = [\boldsymbol{X}^{\text{met}}, \boldsymbol{X}^{\text{cat}}]$ into a $(n \times p^{\text{met}})$ and a $(n \times p^{\text{cat}})$ matrix with p^{met} and p^{cat} metric and dummy encoded categorical covariates, where $p^{\text{met}} + p^{\text{cat}} = p$. For metric covariates, $\boldsymbol{X}^{\text{met}} = [\boldsymbol{x}_1, ..., \boldsymbol{x}_{p^{\text{met}}}]$ consists of $(n \times 1)$ vectors $\boldsymbol{x}_j$ for $j = 1, ..., p^{\text{met}}$.

We have $J = 1, ..., C$ categorical covariates. Dummy encoding is the transformation of J into a set of $df_J + 1$ binary (dummy) variables with $df = 0, 1, ..., df_J$ groups within a dummy encoded covariate J, where $df = 0$ is the reference group, and $p^{\text{cat}} = \sum_{J=1}^{C} df_J$. The model matrix subset $\boldsymbol{X}^{\text{cat}} = [\boldsymbol{X}_1, ..., \boldsymbol{X}_C]$ consists of $(n \times df_J)$ matrices $\boldsymbol{X}_J = [\boldsymbol{x}_{J,1}, ..., \boldsymbol{x}_{J,df_J}]$ for $J = 1, ..., C$.

4.1.1. *GAMLSSs & GAMs.*

The cumulative distribution function (CDF) of the GPD is

$$\text{GPD}(y; \xi, \sigma) = \begin{cases} 1 - (1 + \frac{\xi y}{\sigma})^{-\frac{1}{\xi}} & \text{for} \quad \xi \neq 0, \\ 1 - e^{-\frac{y}{\sigma}} & \text{for} \quad \xi = 0, \end{cases} \tag{1}$$

where $y \geq 0$, $\xi \in \mathbb{R}$, and $\sigma > 0$. We assume $\xi > 0$, implicating heavy-tail dependence, which usually holds for OpRisk data for a sufficiently high threshold (for example Chavez-Demoulin et al. (2016)).

LSEs approximately follow a GPD, whereby the associated coefficients have their own linear predictors $\boldsymbol{\eta}_k$ for $k = 1, 2$. This relation is captured by the subsequent system of equations of (n) column vectors:

$$\boldsymbol{y} \sim GPD\big(\boldsymbol{\xi} = \exp(\boldsymbol{\eta}_1), \boldsymbol{\sigma} = \exp(\boldsymbol{\eta}_2)\big), \tag{2}$$

where exponential transformations of predictors ensure positivity of GPD coefficients. We model the predictors given the covariates in a GAMLSS as

$$\boldsymbol{\eta}_1 = \boldsymbol{\eta}_1(\boldsymbol{X}; \boldsymbol{\beta}_1) = \xi_0 + \sum_{j=1}^{p^{\text{met}}} \boldsymbol{x}_j \xi_j + \sum_{J=1}^{C} \boldsymbol{X}_J \boldsymbol{\xi}_J^\top, \tag{3}$$

$$\boldsymbol{\eta}_2 = \boldsymbol{\eta}_2(\boldsymbol{X}; \boldsymbol{\beta}_2) = \sigma_0 + \sum_{j=1}^{p^{\text{met}}} \boldsymbol{x}_j \sigma_j + \sum_{J=1}^{C} \boldsymbol{X}_J \boldsymbol{\sigma}_J^\top, \tag{4}$$

where we have model intercepts ξ_0 and σ_0, and scalars ξ_j and σ_j for $j = 1, ..., p^{\text{met}}$ which are collected by $\boldsymbol{b}^{\text{met}}$ consisting of scalars $b_{j,k}$ for $\{j, k\}$, where $b_{j,1} = \xi_j$ and $b_{j,2} = \sigma_j$. We have $(1 \times df_J)$ vectors $\boldsymbol{\xi}_J$ and $\boldsymbol{\sigma}_J$ for $J = 1, ..., C$ which are collected by $\boldsymbol{b}^{\text{cat}}$ consisting of $(1 \times df_J)$ vectors $\boldsymbol{b}_{J,k}$ for $\{J, k\}$, where $\boldsymbol{b}_{J,1} = \boldsymbol{\xi}_J$ and $\boldsymbol{b}_{J,2} = \boldsymbol{\sigma}_J$, and $\boldsymbol{b}_{J,k}$ consists of (df_J) scalars $b_{J,k,df}$ for $df = 1, ..., df_J$. The model coefficients of Equations (3) and (4) are collected in $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, which are subsequently collected in $\boldsymbol{B}$ for notational convenience. $\boldsymbol{x}_j \xi_j$ and $\boldsymbol{x}_j \sigma_j$ capture the product of scalar values for metric covariates, whereas $\boldsymbol{X}_J \boldsymbol{\xi}_J^\top$ and $\boldsymbol{X}_J \boldsymbol{\sigma}_J^\top$ collect all dummy encoded categorical covariates and associated coefficients.

The probability density function (PDF) of the GPD distributed GAMLSS linear predictors is given by $d_y(\boldsymbol{y}; \boldsymbol{\xi}, \boldsymbol{\sigma})$. The associated log-likelihood function is

$$\ell_{\text{LS}}(\boldsymbol{B}; \boldsymbol{y}, \boldsymbol{X}) = \sum_{i=1}^{n} \log\, d_y\big(y_i; \xi_i = \exp(\eta_1(\boldsymbol{x}_i; \boldsymbol{\beta}_1)), \sigma_i = \exp(\eta_2(\boldsymbol{x}_i; \boldsymbol{\beta}_2))\big) \tag{5}$$

for $i = 1, ..., n$, where $\boldsymbol{x}_i$ is the i-th row of $\boldsymbol{X}$.

We denote by $\boldsymbol{n}^{\text{LF}}$ the $(n \times 1)$ loss frequency exceedances vector containing (n) scalars, whereby the value of each element is equal to the number of losses $n_t(e)$ for the corresponding risk category e in associated time period t. We split the model matrix $\boldsymbol{X}^{\text{LF}} = [\boldsymbol{X}_{\text{e}}^{\text{LF}}, \boldsymbol{X}_{\text{r}}^{\text{LF}}]$ into a $(\boldsymbol{n} \times 20)$ and $(\boldsymbol{n} \times 6)$ matrix with twenty metric economic factors and six dummy encoded risk categories, respectively. Model matrix $\boldsymbol{X}_{\text{e}}^{\text{LF}} = [\boldsymbol{x}_1^{\text{LF}}, ..., \boldsymbol{x}_{20}^{\text{LF}}]$ consists of 20 $(n \times 1)$ vectors $\boldsymbol{x}_j^{\text{LF}}$ for $j = 1, ..., 20$. We assume that LFEs approximately follow a non-homogeneous Poisson process (NHPP; for example Chavez-Demoulin et al. (2016)). The NHPP rates $\boldsymbol{\kappa}$ have a linear predictor $\boldsymbol{\delta}$ as we assume linear relations between the covariates and the NHPP parameters. Mathematically, we have (n) column vectors in the subsequent system of equations:

$$\boldsymbol{n}^{\text{LF}} \sim \text{Poisson}(\boldsymbol{\kappa} = \exp(\boldsymbol{\delta})), \tag{6}$$

where an exponential link function ensures positivity of NHPP rates. We model the linear predictor given the covariates in a Generalized Additive Model (GAM; Hastie and Tibshirani (2017)) as

$$\boldsymbol{\delta} = \boldsymbol{\delta}(\boldsymbol{X}; \boldsymbol{k}) = \kappa_0 + \sum_{j=1}^{20} \boldsymbol{x}_j^{\mathrm{LF}} \kappa_j + \boldsymbol{X}_{\mathrm{r}}^{\mathrm{LF}} \boldsymbol{\kappa}_{\mathrm{r}}^{\top}, \tag{7}$$

where we have a model intercept κ_0, and scalars κ_j for all j, and a (1×6) vector $\boldsymbol{\kappa}_{\mathrm{r}}$, which collects the coefficients of the six dummy encoded risk categories, and $\boldsymbol{k}$ collects the model coefficients.

4.1.2. *Tuning parameter frameworks.* To specify penalized GAMLSS, we first need to introduce tuning parameter frameworks. Similar to Groll et al. (2019), we consider three tuning parameter frameworks: λ-, λ_k-, and λ_{jk}-type, explained in the following paragraphs.

We denote by $\lambda_{j,k}$ and $\lambda_{J,k}$ (or $\lambda_{j/J,k}$) the tuning parameters for $\{j = 1, ..., p^{\mathrm{met}}, J = 1, ..., C, k = 1, 2\}$, and all tuning parameters are collected in a $(2 \times (p^{\mathrm{met}} + C))$ matrix $\boldsymbol{\lambda}$.

A λ-type framework means that we use one global tuning parameter for both predictors and all covariates by imposing restrictions: $\lambda_{j,k} = \lambda$ and $\lambda_{J,k} = \lambda$ for all $\{j, J, k\}$. This framework is robust, reduces model complexity, and is computationally fast. However, we risk oversimplifying the model.

A λ_k-type framework uses two tuning parameters, one for each predictor. We thereby impose the restrictions $\lambda_{j,k} = \lambda_k$ and $\lambda_{J,k} = \lambda_k$ for all $\{j, J, k\}$. The intuition is that we can examine if either predictor is more sensible to overfitting by comparing the levels of the optimal in-sample tuning parameters after model estimation. The potential disadvantage of this framework is that two tuning parameters must be tuned. Hence, model complexity is increased.

A λ_{jk}-type framework does not impose any restrictions upon its tuning parameters. Hence, all covariates have their own tuning parameter. The advantage is that we can detect noisy and/or overfitting covariates by the level of the tuning parameters. The potential disadvantage of this framework is that the associated estimation procedure is more complex.

4.1.3. *Regularization.* Regularization techniques shrink coefficients to or towards zero, and covariates are consequently in- or excluded in the model after model estimation. Metric covariates are penalized individually, whereas we jointly penalize the groups within dummy encoded categorical covariates. Another perspective is that penalty functions introduce biases in the parameter estimates to decrease the associated variance, known as the bias-variance trade-off.

We extend the LLF of Equation (5) to allow for the inclusion of penalty functions:

$$\ell_{\text{LS}}^{\text{pen}}(\boldsymbol{B}, \boldsymbol{\lambda}; \boldsymbol{y}, \boldsymbol{X}) = \ell_{\text{LS}}(\boldsymbol{B}; \boldsymbol{y}, \boldsymbol{X}) - \sum_{k=1}^{2} \sum_{j=1}^{p^{\text{met}}} \lambda_{j,k} P^{(j,k)}(b_{j,k}) - \sum_{k=1}^{2} \sum_{J=1}^{C} \lambda_{J,k} P^{(J,k)}(\boldsymbol{b}_{J,k}), \quad (8)$$

where $P^{(j,k)}(b_{j,k})$ and $P^{(J,k)}(\boldsymbol{b}_{J,k})$ are scalar penalty functions for metric and dummy encoded categorical covariates.

L_1-norm regularization for metric covariates. Classical L_1-norm (LASSO) regularization for metric covariates (RLASSO$_m$; Tibshirani (1996)) shrinks coefficients to zero and hence ensures sparse parameter spaces. The scalar penalty function for a pair $\{j, k\}$ is given by

$$P_{\text{las}}^{(j,k)}(b_{j,k}) = |b_{j,k}|. \quad (9)$$

Adaptive weights are constructed as oracle estimators (Zou (2006)). These estimators are consistent in parameter estimation and variable selection, and have oracle properties: when the sample size is finite, in-sample non-zero coefficients are identified; when the sample size is infinite, then the estimators are unbiased and normally distributed with true variance. An adaptive weight for a metric covariate for a pair $\{j, k\}$ is given by

$$w_{j,k} = \frac{1}{|\hat{b}_{j,k}^{\text{ini}}|}, \quad (10)$$

where $\hat{b}_{j,k}^{\text{ini}}$ is an initialized coefficient. The penalty function ALASSO$_m$ is constructed by combining Equations (9) and (10) for a pair $\{j, k\}$ as

$$P_{\text{alas}}^{(j,k)}(b_{j,k}) = w_{j,k}|b_{j,k}| = \frac{|b_{j,k}|}{|\hat{b}_{j,k}^{\text{ini}}|}. \quad (11)$$

Fused L_1-norm regularization. Fused L_1-norms take spatial structures, which are interconnections between groups of dummy encoded categorical covariates, into account. We denote by $0 \leq l < m \leq df_J$ two groups within dummy encoded categorical covariates, and we fix $b_{J,k,0} = 0$, which is the reference group, for all $\{J, k\}$. Risk categories are nominal, and economic factors are categorized based on their levels and hence ordinal. Fused L_1-norms penalize dissimilar coefficients, thus parameter sparsity is encouraged since reference groups are fixed as zero.

For nominal and ordinal covariates, we penalize dissimilar coefficients for groups $\{l, m\}$ and $\{l-1, l\}$, respectively. The coefficient of group l is indirectly encouraged to shrink to zero by the linked structure of ordinal covariates. Fused L_1-norm penalty functions (RFUSE) for nominal and ordinal covariates are a sum of $((df_J^2 - df_J)/2)$ and (df_J) individual penalties. The functions are given by

$$P_{\text{fl-n}}^{(J,k)}(\boldsymbol{b}_{J,k}) = \sum_{0 \leq l < m \leq df_J} w_{lm}^{(J)} |b_{J,k,l} - b_{J,k,m}|, \quad (12)$$

$$P_{\text{fl-o}}^{(J,k)}(\boldsymbol{b}_{J,k}) = \sum_{l=1}^{df_J} w_l^{(J)} |b_{J,k,l} - b_{J,k,l-1}|, \tag{13}$$

where $w_{lm}^{(J)}$ and $w_l^{(J)}$ are 'standardizing weights' to account for differences in the number of groups and number of observations within these groups. Bondell and Reich (2009) suggested the following weights:

$$w_{lm}^{(J)} = 2(df_J + 1)^{-1} \sqrt{\frac{n_l^{(J)} + n_m^{(J)}}{n_q}}, \tag{14}$$

$$w_l^{(J)} = \sqrt{\frac{n_l^{(J)} + n_{l-1}^{(J)}}{n_q}}, \tag{15}$$

where $n_l^{(J)}$ and $n_m^{(J)}$ are the number of observations in groups l and m within J. The term $(df_J + 1)^{-1}$ ensures that nominal and ordinal covariates can be combined in a model with a single tuning parameter. Adaptive weight variants are given by

$$w_{alm}^{(J)} = w_{lm}^{(J)} |b_{J,k,l}^{\text{ini}} - b_{J,k,m}^{\text{ini}}|^{-1}, \tag{16}$$

$$w_{al}^{(J)} = w_l^{(J)} |b_{J,k,l}^{\text{ini}} - b_{J,k,l-1}^{\text{ini}}|^{-1}, \tag{17}$$

where the superscript 'ini' refers to initialized coefficients. We construct adaptive fused L_1-norm penalty functions (AFUSE) by replacing regular with adaptive weights. The economic implications of fused L_1-norms, as well as a downside of using adaptive weights are elaborated in Appendix A.5.

L_2-norm regularization. L_2-norms do not encourage parameter sparsity as much as L_1-norms. We therefore only use L_2-norms to obtain adaptive weights in case of identification issues, that is if we cannot obtain a Maximum Likelihood (ML) estimator. We construct classical L_2-norms for metric and dummy encoded categorical covariates for pairs $\{j, k\}$ and $\{J, k\}$ as

$$P_{\text{rm}}^{(j,k)}(b_{j,k}) = b_{j,k}^2, \tag{18}$$

$$P_{\text{rc}}^{(J,k)}(\boldsymbol{b}_{J,k}) = \|\boldsymbol{b}_{J,k}\|_2^2 = b_{J,k,1}^2 + \ldots + b_{J,k,df_J}^2. \tag{19}$$

4.2. *Model estimation.*

We present methods for model estimation. Our models are trained over either $T-1$ or T time periods. We validate models, that are trained over $T-1$ time periods, by the remaining time period t. Such a framework is feasible since we work with models without a time-series structure. To validate models, we first filter the LFEs/LSEs for all T time periods, then we exclude a time period t, fit the model, and lastly predict the remaining period t using parameter estimates resulting from a model that trains over $T-1$ time periods, combined with economic factors in the remaining time period t.

4.2.1. *Unpenalized GAMLSS & GAMLSSBoostAlt.*

We estimate unpenalized GAMLSSs by gradient descent type optimizers to attain good parameter estimates. These optimizers use a prespecified step size ν to update model coefficients. In principle, gradient descent updating schemes are interpolators, estimators that achieve zero training error, for a sufficiently high number of iterations n^{iter}. However, interpolators overfit the data, which could result in inferior parameter estimates for underparametrized regimes. Hence, we early stop gradient descent type optimizers. Groll et al. (2019) use the gradient descent optimizer *gamboostLSS*. First, we show the limitations of *gamboostLSS*, and we thereafter present an altered version of *gamboostLSS*, *GAMLSSBoostAlt*, which simultaneously updates the predictors of both GPD parameters. The performance of both optimizers is evaluated in our simulation study.

gamboostLSS updates coefficients in a two-step approach: *(1)* coefficients are zero-initialized; *(2)* partial derivatives of both GPD predictors with respect to the coefficients of the covariates are calculated. Then, only one covariate is updated, namely the one that improves the associated LLF the most. Afterwards, step *(2)* is repeated iteratively until either the LLF converges or the maximum number of iterations is reached. In practice, model coefficients of one GPD parameter may impact the associated LLF substantially more. Consequently, the coefficients belonging to one GPD parameter are updated many times, whereas the coefficients of the other GPD parameter are updated only a few times. Hence, *gamboostLSS* may yield highly biased coefficients for the latter GPD parameter. Without many coefficient updates, there is a large difference between model parameter estimates and true non-zero parameters. Moreover, highly biased parameter estimates likely yield worse linear predictor estimates for $\boldsymbol{\eta}_1$ and $\boldsymbol{\eta}_2$.

GAMLSSBoostAlt splits step *(2)* into two steps, namely *(2a)* and *(2b)*. In step *(2a)*/*(2b)*, the partial derivatives of $\boldsymbol{\eta}_1/\boldsymbol{\eta}_2$ with respect to its coefficients are calculated. Afterwards, only coefficient set, $\boldsymbol{\eta}_1$ or $\boldsymbol{\eta}_2$, is updated. Hence, *GAMLSSBoostAlt* updates the two GPD parameters every other time. Therefore, *GAMLSSBoostAlt* likely yields lower biased coefficients for the GPD parameter that received few updates under *gamboostLSS*. We consequently expect a decrease in associated Root Mean Squared Errors (RMSEs) of parameter estimates for this GPD predictor. We examine whether this is indeed the case in our simulation study. Furthermore, we may expect a decrease in RMSEs of the out-of-sample GPD predictor estimates $\hat{\boldsymbol{\eta}}_1$ and $\hat{\boldsymbol{\eta}}_2$, compared to the true linear predictor parameters. A disadvantage of *GAMLSSBoostAlt* is that the overperforming predictor under *gamboostLSS* gets fewer updates under *GAMLSSBoostAlt*. To take this into account, we iterate a sufficient number of times so that both predictors are in fact updated many times.

We fix the number of iterations to 25,000 for both *gamboostLSS* and *GAMLSSBoostAlt* and hence the two GPD parameters are updated unknown (between 0 and 25,000) and exactly 12,500 times, respectively. As step size ν for gradient descent optimizers, we consider $\nu = \{.01, .05, .01\}$. We increase the number of iterations for $\nu = .01$ since a lower step size requires more updates.

Both optimizers are implemented in the R package *BAMLSS* (Umlauf et al. (2019)) by *opt-boost*. We implement *GAMLSSBoostAlt* ourselves in a branch of the *BAMLSS* package. We thereby extend the *select.type* input option, which determines how model terms should be selected by log-likelihood contributions, of the *BAMLSS* function by a new option: *select.type = 3*.[2] This option uses *GAMLSSBoostAlt*.

4.2.2. *Tuning parameters.* We obtain optimal tuning parameters by minimizing BIC, which find the best fitted model among a set of candidate models by penalizing the number of non-zero coefficients. Using BIC, we are able to decrease overfitting issues. Hambuckers et al. (2018) demonstrate that BIC yields good variable selection performance, although biases in parameter estimates are on average increased.

We set the number of tuning parameters in a 1- and 2-dimensional grid equal to 200 and 30^2 for λ- and λ_k-type models with lower- and upper bounds equal to exp(-3) and exp(7), respectively. Iterating over a $(2 \times (p^{\text{met}} + C))$-dimensional grid in a λ_{jk} framework is computationally infeasible. Hence, we update the associated coefficients stepwise (Umlauf et al. (2018)). For λ, λ_k, and $\lambda_{j,k}$ frameworks, we use the *opt-lasso* (first two) and *opt-bfit* (third) optimizers in *BAMLSS*.

4.2.3. *Penalized GAMLSSs & GAMs.* We quadratically approximate penalty functions to allow for a combination of penalties in a model, and to ensure continuity and differentiability of penalty functions (Oelker and Tutz (2017)). This approximation yields the LLF: $\ell_{\text{LS}}^{\text{pen-a}}(\boldsymbol{B}, \boldsymbol{\lambda}; \boldsymbol{y}, \boldsymbol{X})$, which replaces the LLF of Equation (8). To illustrate this approximation, a metric L_1-norm penalty function for $\{j, k\}$ is approximated as

$$|b_{j,k}| \approx \sqrt{b_{j,k}^2 + 10^{-8}}, \tag{20}$$

which is continuous and differentiable in $\mathbb{R}$. To estimate GAMLSS parameters, we use Iterated Weighted Least Squares (IWLS) updating schemes. We obtain penalized and unpenalized ML estimators $\widehat{\boldsymbol{B}}_\lambda$ and $\widehat{\boldsymbol{B}}$ by maximizing the LLFs of $\ell_{\text{LS}}^{\text{pen-a}}(\boldsymbol{B}, \boldsymbol{\lambda}; \boldsymbol{y}, \boldsymbol{X})$ and Equation (5) with respect to $\boldsymbol{B}$ as

$$\widehat{\boldsymbol{B}}_\lambda = \arg\max_{\boldsymbol{B}} \left\{ \ell_{\text{LS}}^{\text{pen - a}}(\boldsymbol{B}, \boldsymbol{\lambda}; \boldsymbol{y}, \boldsymbol{X}) \right\}, \tag{21}$$

$$\widehat{\boldsymbol{B}} = \arg\max_{\boldsymbol{B}} \left\{ \ell_{\text{LS}}(\boldsymbol{B}, \boldsymbol{y}, \boldsymbol{X}) \right\}. \tag{22}$$

No analytical solution exists and hence we use numerical methods (Umlauf et al. (2019)), and we have the limiting case $\lim_{\lambda \to 0}\{\widehat{\boldsymbol{B}}_\lambda\} = \widehat{\boldsymbol{B}}$.

The LLF of NHHP distributed unpenalized GAMs is $\ell_{\text{LF}}(\boldsymbol{k}; \boldsymbol{y}, \boldsymbol{X})$.We obtain an ML esti-

[2] http://www.bamlss.org/reference/boost.html

mator for $\widehat{\boldsymbol{k}}$ by maximizing this LLF with respect to $\boldsymbol{k}$ as

$$\widehat{\boldsymbol{k}} = \arg\max_{\boldsymbol{k}} \ell_{\mathrm{LF}}(\boldsymbol{k}; \boldsymbol{y}, \boldsymbol{X}). \tag{23}$$

No analytical solution exists. We obtain a ML estimator by Penalized Iteratively Re-Weighted Least Squares (P-IRLS; Wood (2011)) updating schemes by Generalized Cross Validation (GCV), implemented in the *MGCV* R package (Wood and Wood (2015)).

4.2.4. *Modelling details & benchmark.* Our optimizers can not, and usually do not, guarantee optimal solutions, which means that LLFs do not converge during model estimation. We therefore initialize coefficients in a λ_{jk}-type framework with λ_k coefficients in a similar context. There is a trade-off for initialization paths: bad starting coefficients likely result in an inferior solution path, whereas 'too specific' initialization could result in instant solutions as the optimizer gets stuck in, or close to, the previous found solution.

We benchmark *gamboostLSS* to evaluate *GAMLSSBoostAlt* in our simulation study. In our real-world application, we benchmark the best performing model of Groll et al. (2019), according to their own conclusions. This model uses dummy encoded categorical covariates, fused L_1-norms in a λ_{jk} framework with λ_k starting coefficients, and adaptive weights. The adaptive weight coefficients are obtained by the estimation of an unpenalized GAMLSS using *gamboostLSS*. Since unpenalized GAMLSS estimation procedures are a derivative of the original GAM backfitting estimation procedures, updating procedures for a given iteration heavily depends on the coefficients of the previous iteration. Hence, the number of potential coefficient paths is infinitely many. Therefore, we experiment with multiple optimizers to examine which optimizers are the best fit for our application.

4.2.5. *Training and testing.* We standardize all covariates to ensure that all covariates are weighted equally before applying regularization. We simulate both loss severities and covariates in our study study. Thereafter, we randomly split the rows into a 80% and 20% train (tr) and test (te) set. For the real-world application, we first train models over all 38 time periods for LFE- and LSE-models. Furthermore, we train loss frequency models over 37 time periods. Then, we use these in-samples model parameter estimates, combined with economic factor information in the remaining time period, to predict model parameters in the remaining time period. We thereby introduce time-variation in the NHPP distributed GAM parameters. A similar procedure could also be applied to the loss severity models but we chose not to do so due to time-related as well as computational reasons.

4.3. *Assessing model quality.*

We elaborate our simulation setup, and present model evaluation techniques for the simulation study as well as the real-world application.

4.3.1. *Simulation study, setup.* We conduct a simulation study for well-specified models to evaluate the performance of our proposed LSE-models. In our simulation study, we use similar notation as introduced in Section 4.1. What is different, is that we differentiate between informative (inf) and uninformative (unf) covariates in the simulation. We denote by $C^{\text{sim}} = C^{\text{inf}} + C^{\text{unf}}$ and $p^{\text{sim}} = p^{\text{inf}} + p^{\text{unf}}$ the number of regular and dummy encoded simulated categorical covariates, whereby $J^{\text{inf}} = 1, ..., C^{\text{inf}}$, $J^{\text{unf}} = 1, ..., C^{\text{unf}}$, and $J^{\text{sim}} = 1, ..., C^{\text{inf}}, C^{\text{inf}} + 1, ..., C^{\text{sim}}$, and $p^{\text{inf}} = \sum_{J=1}^{C^{\text{inf}}} df_J^{\text{inf}}$ and $p^{\text{unf}} = \sum_{J=1}^{C^{\text{unf}}} df_J^{\text{unf}}$, where df_J^{inf} and df_J^{unf} are the number of groups within J^{inf} and J^{unf}.

We solely simulate categorical covariates, and not metric covariates since we can not compare two non-identical covariate sets in a simulation, explained in Appendix A.6. We denote by n^{sim} the number of simulated losses. Let $\boldsymbol{X}^{\text{sim}}$ be the $(n^{\text{sim}} \times p^{\text{sim}})$ model matrix, that we split as $\boldsymbol{X}^{\text{sim}} = [\boldsymbol{X}^{\text{inf}}, \boldsymbol{X}^{\text{unf}}]$ into a $(n^{\text{sim}} \times p^{\text{inf}})$ and a $(n^{\text{sim}} \times p^{\text{unf}})$ matrix consisting of $(n^{\text{sim}} \times df_J^{\text{inf}})$ and $(n^{\text{sim}} \times df_J^{\text{unf}})$ matrices $\boldsymbol{X}_J^{\text{inf}}$ and $\boldsymbol{X}_J^{\text{unf}}$.

We use two nominal and two ordinal informative covariates, thus $C^{\text{inf}} = 4$. We consider two uninformative covariate frameworks $C^{\text{unf}} = \{4, 11\}$, whereby the number of nominal and ordinal uninformative covariates is equal to 2 and $C^{\text{unf}} - 2$, respectively. Therefore, the total number of simulated covariates is equal to $C^{\text{sim}} = \{8, 15\}$. We use this covariate setup to reflect real-world data sets since real-world data sets usually include many ordinal covariates and a few nominal covariates (for example Hambuckers et al. (2018), Groll et al. (2019)). Many ordinal covariates are included since economic factors can usually be sorted on their levels.

To simulate covariates, we first denote by $\boldsymbol{P}$ a $(C^{\text{sim}} \times C^{\text{sim}})$ correlation matrix with non-diagonal and diagonal elements ρ and 1, whereby we examine three correlation frameworks with an average cross-correlation of $\rho = \{0, .3, .6\}$. We thereby use the multivariate normal distribution (MVND) to simulate metric covariates. Metric covariates are subsequently transformed into categorical covariates with 4 or 8 groups, whereby we alternate between both group sizes. Afterwards, these covariates are dummy encoded into $\boldsymbol{X}_J^{\text{inf}}$ and $\boldsymbol{X}_J^{\text{unf}}$ for $J = J^{\text{sim}} = 1, ..., C^{\text{sim}}$. Initially, we attempted to construct simulated covariates with correlations above $\rho = .6$. Unfortunately, this is not possible, as explained in Appendix A.7, because we use the MVND in our simulation procedure.

To proxy extreme losses, we draw samples from the GPD as $\boldsymbol{y}^{\text{sim}} \sim \text{GPD}\big(\boldsymbol{\xi}^{\text{inf}} = \exp(\boldsymbol{\eta}_1^{\text{inf}}), \boldsymbol{\sigma}^{\text{inf}} = \exp(\boldsymbol{\eta}_2^{\text{inf}})\big)$ of size $n^{\text{sim}} = \{5,000; 10,000; 20,000\}$, where $i^{\text{sim}} = 1, ..., n^{\text{sim}}$, where the true non-zero dummy parameters $\boldsymbol{\xi}^{\text{inf}}$ and $\boldsymbol{\sigma}^{\text{inf}}$ are fixed to values that we present in Table 5. We combine informative dummy encoded categorical covariates and associated true non-zero parameters for the data generating process (DGP), given by

$$\boldsymbol{\eta}_1^{\text{inf}} = \xi_0^{\text{sim}} + \sum_{J=1}^{4} \boldsymbol{X}_J^{\text{inf}} (\boldsymbol{\xi}_J^{\text{inf}})^\top, \quad (24) \qquad \boldsymbol{\eta}_2^{\text{inf}} = \sigma_0^{\text{sim}} + \sum_{J=1}^{4} \boldsymbol{X}_J^{\text{inf}} (\boldsymbol{\sigma}_J^{\text{inf}})^\top, \quad (25)$$

where ξ_0^{sim} and σ_0^{sim} are true model intercepts, and $(1 \times df_J^{\text{inf}})$ vectors $\boldsymbol{\xi}_J^{\text{inf}}$ and $\boldsymbol{\sigma}_J^{\text{inf}}$ contain the true parameters. This procedure yields losses that mirror real-world extreme OpRisk losses since we assume that LSEs follow a GPD. Therefore, our simulation study is able to reflect all kind of loss data, and no assumptions on the underlying distribution of original loss series are necessary. For real-world loss data, it is the responsibility of the modeller to find a sufficiently high threshold such that their LSEs indeed follow a scaled GPD. UniCredit's loss data include just over 10,000 LSEs for our threshold choice (Section 3.2), and 10,000 is within the range of n^{sim}.

Table 5: True non-zero parameters in our simulation.

GPD parameter	Parameter	True parameters
	ξ_0^{sim}	-0.3
	$\boldsymbol{\xi}_1^{\text{inf}}$	(0, 0.3, 0.3, 0.3, 0.3,-0.5,-0.5,-0.5)
ξ	$\boldsymbol{\xi}_2^{\text{inf}}$	(0, -0.4, -0.4)
	$\boldsymbol{\xi}_3^{\text{inf}}$	(0, -0.4, -0.4, -0.8, -0.8, -1.1, -1.1)
	$\boldsymbol{\xi}_4^{\text{inf}}$	(0, -0.5, -0.5)
	σ_0^{sim}	-0.2
	$\boldsymbol{\sigma}_1^{\text{inf}}$	(0, -0.6, 0.3, 0, -0.6, 0.3, 0)
σ	$\boldsymbol{\sigma}_2^{\text{inf}}$	(0.4, 0, 0.4)
	$\boldsymbol{\sigma}_3^{\text{inf}}$	(0, 0, -0.4, -0.4, -0.4, -0.9, -0.9)
	$\boldsymbol{\sigma}_4^{\text{inf}}$	(0, -0.3, -0.3)

4.3.2. *Simulation study, evaluation.* We exclude notational superscripts 'sim' for writing convenience. The true model and subsequent GPD parameters are known in our simulation study. We compare true and estimated model parameters by RMSEs for an individual dummy encoded categorical covariate J and a GPD parameter k as

$$RMSE_{b_{J,k}} = \sqrt{\frac{1}{df_J} \sum_{df=1}^{df_J} (\hat{b}_{J,k,df} - b_{J,k,df})^2}. \tag{26}$$

To obtain a single performance measure for all C covariates, we aggregate the RMSEs as follows:

$$RMSE_{b_k}^{\text{agg}} = \sqrt{\sum_{J=1}^{C} \left(\frac{df_J}{\sum_{J=1}^{C} df_J} \right) (RMSE_{b_{J,k}})^2}, \tag{27}$$

where we use a rescaling factor $\left(\frac{df_J}{\sum_{J=1}^{C}} \right)$ to take into account that the number of groups within categorical covariates differ among covariates.

In Section, 4.2 we demonstrate how to estimate our models, and we thereby obtain in-sample model parameter estimates that we use to predict out-of-sample GPD parameters in the simulation. Let $n^{\text{te}} = n \times 0.2$ be the number of observations in the test set. We denote by $\boldsymbol{X}^{\text{te}}$ the $(n^{te} \times p)$ model matrix consisting of $(n^{\text{te}} \times df_J)$ matrices $\boldsymbol{X}_J^{\text{te}}$ for $J = 1, ..., 4, 5, ..., C$, where $J = 4$ and $J = 5$ are linked to the last informative covariate and the first uninformative covariate. We predict the two GPD parameters by their corresponding predictors as

$$\hat{\boldsymbol{\eta}}_1 = \hat{\xi}_0 + \sum_{J=1}^{4} \boldsymbol{X}_J^{\text{te}} \left(\hat{\boldsymbol{\xi}}_J^{\text{inf}}\right)^\top + \sum_{J=5}^{C} \boldsymbol{X}_J^{\text{te}} \left(\hat{\boldsymbol{\xi}}_J^{\text{unf}}\right)^\top, \tag{28}$$

$$\hat{\boldsymbol{\eta}}_2 = \hat{\sigma}_0 + \sum_{J=1}^{4} \boldsymbol{X}_J^{\text{te}} \left(\hat{\boldsymbol{\sigma}}_J^{\text{inf}}\right)^\top + \sum_{J=5}^{C} \boldsymbol{X}_J^{\text{te}} \left(\hat{\boldsymbol{\sigma}}_J^{\text{unf}}\right)^\top, \tag{29}$$

where we have $(n^{te} \times 1)$ predictor estimates on the left-hand side. On the right-hand side, we have model intercept estimates $\hat{\xi}_0$ and $\hat{\sigma}_0$, and $(n^{te} \times df_J)$ matrices $\boldsymbol{X}_J^{\text{te}}$, and $(1 \times df_J)$ parameter estimates $\hat{\boldsymbol{\xi}}_J$ and $\hat{\boldsymbol{\sigma}}_J$ for $J = 1, ..., C$, where $\hat{\boldsymbol{\xi}}_J^{\text{unf}}$ and $\hat{\boldsymbol{\sigma}}_J^{\text{unf}}$ are model coefficients of uninformative covariates. We compare predicted and true GPD parameters by their predictor estimates for a distributional parameter k as

$$RMSE_{\eta_k} = \sqrt{\frac{1}{n^{\text{te}}} \sum_{i=1}^{n^{\text{te}}} (\hat{\eta}_{i,k} - \eta_{i,k})^2}. \tag{30}$$

We aggregate the RMSEs of Equations (27) and (30) over the two GPD parameters as follows:

$$RMSE_b^{\text{agg}} = \sqrt{\frac{1}{2} \sum_{k=1}^{2} (RMSE_{b_k}^{\text{agg}})^2}, \tag{31}$$

$$RMSE_\eta^{\text{agg}} = \sqrt{\frac{1}{2} \sum_{k=1}^{2} (RMSE_{\eta_k})^2}. \tag{32}$$

In Equation (31), a scaling factor 1/2 is used because the RMSEs of the two GPD parameters are calculated by vectors of identical length (n^{te}), and a scaling factor 1/2 in Equation (32) is used because the model matrices of the two GPD parameters are identical. Therefore, the number of groups within categorical covariates are identical as well since $df_{J,1} = df_{J,2}$ for all J, where $df_{J,k}$ is the number of groups within a dummy encoded categorical covariate J for distributional parameter k.

Moreover, we evaluate the variable selection performance to identify: true fused categories, uninformative and informative covariates, by false negative and positive rates, FNRs and FPRs, respectively. FNRs (FPRs) are the likelihood that covariates or fused groups within covariates are falsely excluded (included) in a model. A sensitivity measure, set to 10^{-4}, determines ex- or inclusion in models. Lastly, we report the in-sample and out-of-sample log-likelihood scores (INS-LLH; OOS-LLH), BIC, and EDF, which is the approximated number of non-zero coefficients.

4.3.3. *Real-world application.* Evaluating real-world data is difficult because the true model and GPD parameters are unknown in the physical world. To evaluate real-world data, we compute in-sample BICs and EDFs as well as OOH-LLH scores using the validation set.

5. Empirical analysis.

We present empirical findings for both the simulation study and real-world application. An overview of the R functions that we use in our analyses is presented in Appendix A.13.

5.1. *Simulation study.*

We evaluate the performance of models that use *gamboostLSS*, which we refer to as B1-type models (or B1), and *GAMLSSBoostAlt*, which we refer to as B2-type models (or B2). The main analysis of the two boosting optimizers is for $n^{\text{sim}} = 5000$, $C^{\text{sim}} = 8$ and $\rho = 0$. We thereby validate our findings over 10 samples (due to high computational time) over a grid of n^{iter} since the optimal n^{iter} is unknown.

The structure is as follows: in Section 5.1.1 we summarize our findings. In Sections 5.1.2, 5.1.3, and 5.1.4, we evaluate both optimizers in both an penalized and unpenalized framework, and we compare updating schemes of boosting optimizers, respectively.

5.1.1. *Summary of findings.*

Our core goals are to improve model parameter estimates and to better perform variable selection. In Table 6, we find that B2 models slightly outperform B1 models according to the $\text{RMSE}_b^{\text{agg}}$ evaluation measure. Thereby, the best performing models are AFUSE(λ_k)-B2 and AFUSE(λ_{jk})-B2. These models use our *GAMLSSBoostAlt* optimizer, include adaptive weights, apply L_1-norms, and use λ_k/λ_{jk}-type tuning parameter frameworks. However, finding good stopping criteria for gradient boosting optimizers remains challenging. Figure 2 demonstrates that model parameters as well as GPD parameter estimates deteriorate after a bliss number of iterations n_*^{iter} and hence stopping criteria must be determined. We consider AIC and BIC, and find that AIC does not satisfy optimality conditions since the first derivative is never zero, whereas BIC shows potential but requires losses to be adequately scaled to proportionalize the level of the log-likehood function and the number of model parameters (Figure 3). BIC requires adequately scaling of loss severities since the magnitude of the losses determines the level of associated log-likelihoods, and the BIC penalization component (number of model parameters times the logarithm of the number of loss severities) is fixed. With regard to the variable selection performance of B1 and B2, B2-type models identify fused groups on average slightly better (Table 6), but these differences are marginal. We subsequently provide an explanation for the improved model parameter estimates under B2 as follows: B1-type models yield unstable updating schemes, as shown in Figure 6, whereas B2 updating schemes are stable by construction. The unstable updating of B1 is due to the GPD scale parameter (σ), which is the driving force of the involved log-likelihood function. Therefore, σ covariates are updated many times for the first thousands of iterations (further detailed in Section 5.1.4).

5.1.2. *Comparison of gradient boosting optimizers in a penalized framework.* Next, we evaluate our regularization methods. We use coefficients that result from a model with parameters $\{n^{\text{iter}} = 5,000, \nu = .05\}$ as adaptive weights. Table 6 demonstrates that regular weight models are inferior to adaptive weight models. This can be explained by the fact that regular weight models use IWLS updating schemes, and IWLS insensibly iteratively updates all model coefficients. Since we dummy encode categorical covariates, the number of parameters that need to be estimated is high and hence IWLS updating schemes do not work well in our application. Another perspective is that adaptive weights successfully act as stabilizing force by applying counter-pressure to L_1-norms.

With regard to the parameter estimates, Table 6 shows that both the model parameter estimates slightly improve under B2, compared to B1. Although, regularization techniques worsen GPD parameter estimates, compared to unpenalized models. This could be explained by the fact that boosting optimizers update the model parameter estimates. In this process, the GPD parameter estimates are not involved and hence these coefficients do not improve when applying L_1-norms. Moreover, σ model parameter estimates substantially improve when applying L_1-norms, whereas ξ model parameter estimates improve less substantially. We use Equations (31) and (32) to aggregate RMSEs, and find that the GPD shape parameter dominates aggregated RMSEs. This is due to the relatively high RMSEs for the GPD shape parameter, compared to the GPD scale parameter.

We explored three tuning parameter frameworks. The λ_k/λ_{jk}-frameworks on average yields better parameter estimates than a λ framework (Table 6). This means that the two GPD parameters differ in their sensitivity to overfitting. Furthermore, a λ_{jk}-framework on average slightly increases the model EDFs. Hence, such a framework should only be considered when the use of a λ-/λ_k-framework in a similar context results in a sufficiently sparse solution.

With regard to variable selection, Table 6 demonstrates that fused L_1-norms frequently detect fused groups within categorical covariates, although detection becomes more difficult for $\rho = .6$. The false negative rates show that informative covariate parameter estimates are sometimes shrunk to zero, which is the case when the optimal high tuning parameter(s) is/are sufficiently high. Since L_1-norms add a penalty to the log-likelihood function, the associated log-likelihood value by definition worsen due to the increase in biasses in model parameter estimates. Nevertheless, BICs show substantial improvements, when applying L_1-norms, due to the decreased equivalent degrees of freedoms. EDFs of unpenalized models frequently are close to 82, which is the total number of model parameters for $C^{\text{sim}} = 8$. Our regularization methods reduce the EDF to on average 15. This indicates that L_1-norms successfully attain sparse model

Table 6: Evaluation measures for of our models for $C^{\text{sim}} = 4$ and $\nu = .05$

ρ	Model	$\text{RMSE}_{\eta}^{\text{agg}}$	RMSE_{η_1}	RMSE_{η_2}	$\text{RMSE}_{b}^{\text{agg}}$	$\text{RMSE}_{b_1}^{\text{agg}}$	$RMSE_{b_2}^{agg}$	FPR^{fuse}	$\text{FPR}^{\text{noise}}$	FNR	LLH	BIC	EDF
	B1	**1**	**1**	1	1	1	1	0.98	0.88	0.02	**-2929.74**	6529.64	80.8
		(0)	(0)	(0)	(0)	(0)	(0)	(0.05)	(0.12)	(0.05)	(98.48)	(190.48)	(1.55)
	B2	1.01	**1**	1.01	1.03	1.03	1.02	0.99	0.86	**0.01**	-2930.62	6526.43	80.2
		(0.02)	(0.02)	(0.02)	(0.04)	(0.04)	(0.05)	(0.04)	(0.09)	(0.04)	(98.82)	(193.71)	(2.1)
	RFUSE(λ)	1.5	1.56	1.02	1.2	1.23	0.99	0.46	0.26	0.25	-2980.21	6092.58	15.93
		(0.14)	(0.15)	(0.16)	(0.12)	(0.13)	(0.16)	(0.1)	(0.1)	(0)	(101.46)	(199.7)	(0.89)
	AFUSE(λ)-B1	1.1	1.16	**0.62**	0.91	0.95	**0.57**	0.35	**0.16**	0.14	-2959.41	6038.37	14.41
		(0.09)	(0.11)	(0.11)	(0.12)	(0.13)	(0.12)	(0.14)	(0.15)	(0.09)	(98.98)	(197.9)	(1.03)
0	AFUSE(λ_k)-B1	1.02	1.07	**0.62**	0.86	0.89	**0.57**	0.34	**0.16**	0.12	-2957.86	**6035.23**	14.41
		(0.1)	(0.11)	(0.14)	(0.11)	(0.13)	(0.14)	(0.13)	(0.15)	(0.08)	(99.96)	(198.84)	(0.93)
	AFUSE(λ_{jk})-B1	1.02	1.07	**0.62**	0.86	0.89	**0.57**	0.34	**0.16**	0.12	-2957.86	6044.34	15.51
		(0.1)	(0.11)	(0.14)	(0.11)	(0.13)	(0.14)	(0.13)	(0.15)	(0.08)	(99.96)	(199.67)	(0.69)
	AFUSE(λ)-B2	1.1	1.16	**0.62**	0.91	0.95	0.58	0.36	0.19	0.15	-2959.26	6038.82	14.5
		(0.08)	(0.1)	(0.12)	(0.12)	(0.13)	(0.13)	(0.12)	(0.15)	(0.1)	(98.8)	(197.93)	(0.85)
	AFUSE(λ_k)-B2	**1**	1.05	0.63	**0.84**	**0.87**	0.58	**0.3**	**0.16**	0.14	-2957.95	6035.53	14.42
		(0.09)	(0.1)	(0.13)	(0.12)	(0.14)	(0.14)	(0.12)	(0.15)	(0.09)	(97.98)	(198.82)	(0.83)
	AFUSE(λ_{jk})-B2	**1**	1.05	0.63	**0.84**	**0.87**	0.58	**0.3**	**0.16**	0.14	-2957.95	6044.96	15.56
		(0.09)	(0.1)	(0.13)	(0.12)	(0.14)	(0.14)	(0.12)	(0.15)	(0.09)	(97.98)	(198)	(0.63)
	B1	**1**	**1**	1	1	1	1	0.96	0.85	**0**	**-2908.1**	6473.91	79.3
		(0)	(0)	(0)	(0)	(0)	(0)	(0.06)	(0.08)	(0)	(97.33)	(197.05)	(1.7)
	B2	1.01	1.01	1.02	1.01	1.01	1.01	0.98	0.85	0.01	-2908.72	6475.15	79.3
		(0.02)	(0.02)	(0.02)	(0.04)	(0.04)	(0.06)	(0.05)	(0.08)	(0.04)	(97.16)	(202.94)	(1.7)
	RFUSE(λ)	1.7	1.8	1.06	1.27	1.32	0.95	0.46	0.26	0.25	-2980.21	6092.58	15.93
		(0.28)	(0.31)	(0.29)	(0.3)	(0.34)	(0.17)	(0.1)	(0.1)	(0)	(101.46)	(199.7)	(0.89)
	AFUSE(λ)-B1	1.24	1.32	0.65	0.95	1.01	0.56	0.35	0.16	0.14	-2959.41	6038.37	14.41
		(0.23)	(0.24)	(0.26)	(0.2)	(0.22)	(0.17)	(0.14)	(0.15)	(0.09)	(98.98)	(197.9)	(1.03)
.3	AFUSE(λ_k)-B1	1.13	1.19	**0.64**	0.9	0.95	**0.55**	0.34	**0.16**	0.12	-2957.86	**6035.23**	14.41
		(0.16)	(0.19)	(0.21)	(0.19)	(0.21)	(0.15)	(0.13)	(0.15)	(0.08)	(99.96)	(198.84)	(0.93)
	AFUSE(λ_{jk})-B1	1.13	1.19	**0.64**	0.9	0.95	**0.55**	0.34	**0.16**	0.12	-2957.86	6044.34	15.51
		(0.16)	(0.19)	(0.21)	(0.19)	(0.21)	(0.15)	(0.13)	(0.15)	(0.08)	(99.96)	(199.67)	(0.69)
	AFUSE(λ)-B2	1.24	1.33	0.65	0.95	1.01	0.57	0.36	0.19	0.15	-2959.26	6038.82	14.5
		(0.23)	(0.25)	(0.26)	(0.21)	(0.23)	(0.17)	(0.12)	(0.15)	(0.1)	(98.8)	(197.93)	(0.85)
	AFUSE(λ_k)-B2	1.12	1.18	**0.64**	**0.88**	**0.93**	0.56	**0.3**	**0.16**	0.14	-2957.95	6035.53	14.42
		(0.18)	(0.19)	(0.22)	(0.19)	(0.21)	(0.15)	(0.12)	(0.15)	(0.09)	(97.98)	(198.82)	(0.83)
	AFUSE(λ_{jk})-B2	1.12	1.18	**0.64**	**0.88**	**0.93**	0.56	**0.3**	**0.16**	0.14	-2957.95	6044.96	15.56
		(0.18)	(0.19)	(0.22)	(0.19)	(0.21)	(0.15)	(0.12)	(0.15)	(0.09)	(97.98)	(198)	(0.63)
	B1	1	1	1	1	1	1	0.95	0.93	0.16	**-2916.83**	6503.82	80.8
		(0)	(0)	(0)	(0)	(0)	(0)	(0.06)	(0.06)	(0.09)	(102.85)	(200.19)	(1.55)
	B2	**0.99**	**0.98**	1.01	1	1	1.03	0.95	0.9	**0.16**	-2917.11	6499.4	80.2
		(0.03)	(0.03)	(0.01)	(0.03)	(0.03)	(0.04)	(0.06)	(0.08)	(0.06)	(102.78)	(204.16)	(1.55)
	RFUSE(λ)	1.6	1.68	1.05	1.18	1.23	0.91	0.43	0.35	0.25	-2980.21	6092.58	15.93
		(0.29)	(0.34)	(0.23)	(0.26)	(0.33)	(0.14)	(0.1)	(0.1)	(0)	(101.46)	(199.7)	(0.89)
	AFUSE(λ)-B1	1.18	1.25	**0.64**	0.89	0.94	**0.54**	0.34	0.24	0.19	-2959.41	6038.37	14.41
		(0.24)	(0.26)	(0.23)	(0.16)	(0.2)	(0.16)	(0.16)	(0.11)	(0.09)	(98.98)	(197.9)	(1.03)
.6	AFUSE(λ_k)-B1	1.07	1.13	0.65	0.83	0.88	**0.54**	**0.24**	**0.22**	0.17	-2957.86	**6035.23**	14.41
		(0.19)	(0.22)	(0.25)	(0.15)	(0.19)	(0.17)	(0.13)	(0.13)	(0.09)	(99.96)	(198.84)	(0.93)
	AFUSE(λ_{jk})-B1	1.07	1.13	0.65	0.83	0.88	**0.54**	**0.24**	**0.22**	0.17	-2957.86	6044.34	15.51
		(0.19)	(0.22)	(0.25)	(0.15)	(0.19)	(0.17)	(0.13)	(0.13)	(0.09)	(99.96)	(199.67)	(0.69)
	AFUSE(λ)-B2	1.18	1.25	**0.64**	0.88	0.94	0.55	0.36	0.25	0.19	-2959.26	6038.82	14.5
		(0.23)	(0.26)	(0.24)	(0.17)	(0.2)	(0.17)	(0.12)	(0.11)	(0.09)	(98.8)	(197.93)	(0.85)
	AFUSE(λ_k)-B2	1.06	1.12	0.65	**0.82**	**0.86**	**0.54**	0.25	0.24	**0.16**	-2957.95	6035.53	14.42
		(0.2)	(0.22)	(0.25)	(0.16)	(0.19)	(0.18)	(0.12)	(0.14)	(0.1)	(97.98)	(198.82)	(0.83)
	AFUSE(λ_{jk})-B2	1.06	1.12	0.65	**0.82**	**0.86**	**0.54**	0.25	0.24	**0.16**	-2957.95	6044.96	15.56
		(0.2)	(0.22)	(0.25)	(0.16)	(0.19)	(0.18)	(0.12)	(0.14)	(0.1)	(97.98)	(198)	(0.63)

Note: The evaluation measures are validated over 10 samples, whereby the associated standard deviations are between parentheses. RMSEs of model and GPD parameter estimates are relative to model B1 as follows: $\frac{\text{RMSE(B)}}{\text{RMSE(B1)}}$, where B is a model in the column "Model".

coefficient spaces for the optimal in-sample tuning penalty parameter(s) when using BIC.

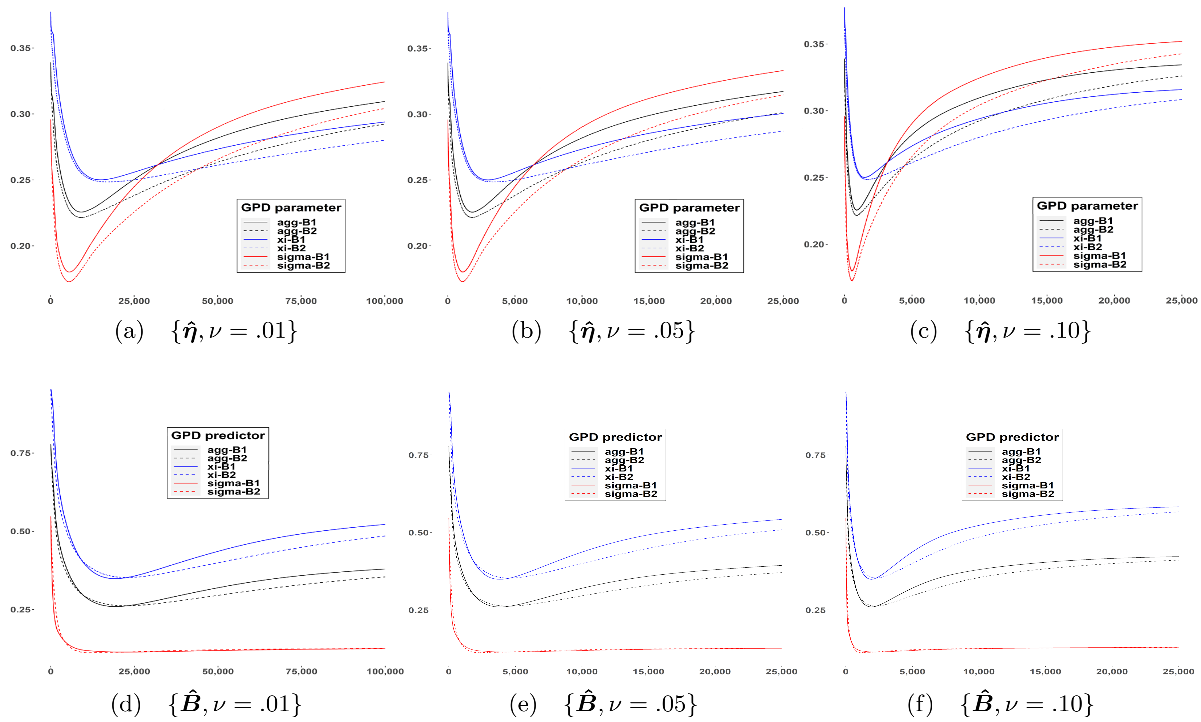
With regard to the number of observation included in the simulation study: the performance of models for $n^{\text{sim}} = \{10,000; 20,000\}$ is similar to $n^{\text{sim}} = 5,000$ and therefore not shown. This can be explained as follows: we do not run into identification issues since $n^{\text{sim}} = 5,000 >> \max(p^{\text{sim}})$, and $n^{\text{sim}} = 5,000$ is sufficiently large so that asymptotic properties can hold. However, we do emphasize that overfitting issues are more likely to arise for $n^{\text{sim}} > 5,000$. Therefore, optimal tuning parameters for $n^{\text{sim}} = 5,000$ should be considered as a conservative estimate for the optimal tuning parameters for $n^{\text{sim}} > 5,000$. In our test samples, this issue was restricted by considering BICs since BICs penalize log-likelihoods by a product of the estimated EDFs, and a scalar that is equal to $\log(n^{\text{sim}})$, which increases in n^{sim}. For $n^{\text{sim}} > 5,000$, there seems to be a relation between n^{sim} and the optimal stopping iteration of gradient descent optimizers (negative relation), and the optimal tuning parameters (positive relation). This can be explained by the fact that overfitting issues are more likely to arise when the model is trained over a higher number of observations n^{sim}. Therefore, by stopping the boosting optimizers earlier or by choosing higher tuning parameters, the risk of overfitting and consequently model misspecification can be mitigated.

5.1.3. *Comparison of gradient boosting optimizers in an unpenalized framework.* We subsequently investigate both B1 and B2 in an unpenalized framework for comparison as well as to attain a deeper understanding of these model optimizers. We provide figures to demonstrate differences between B1- and B2-type models. Hereby, we mainly evaluate one sample. Although there are marginal differences among the 10 samples that we consider, data characteristics of one sample also hold for other samples.

Figure 2 shows that B2 slightly improves model and GPD parameter estimates. This can be explained by the fact that few GPD shape parameters are updated for low to intermediate n^{iter} for B1-type models (later illustrated in Section 5.1.4). The boosting optimizer step size ν does not substantially impact the level of optimal RMSEs, albeit that the slopes of the associated RMSEs are much steeper for high ν. We find that the optimal n_*^{iter}, based on the RMSEs of the model and GPD parameter estimates, is inversely proportional to ν. After reaching n_*^{iter}, parameter estimates deteriorate, as shown in Figure 2. Therefore, B1 and B2 do not behave as interpolators since zero training error is not achieved. The reason is that boosting optimizers maximize log-likelihood functions and hence they do not directly optimize model and GPD parameters. Particularly, $\hat{\boldsymbol{\beta}}_{\mathbf{1}}$ and $\hat{\boldsymbol{\eta}}_1$ estimates are substantially improved under *GAMLSSBoostAlt* because the associated covariates of the GPD shape parameter receive many more updates. The $\hat{\boldsymbol{\eta}}$ parameter estimates are more volatile over n^{iter} than $\widehat{\boldsymbol{B}}$ estimates since the GPD parameter is represented by a single instance, whereas the GPD model parameters are many, and only one covariate can be updated per iteration.

With regard to the information criteria, Figure 3 demonstrate that AIC does not penalize coefficients sufficiently severely since AICs do not reach a global minimum. On the other hand, BICs attains a global minimum for low n^{iter}. However, BICs achieve optimality when n^{iter} is much lower than the number of iterations for which the model coefficients are optimally fit (n_*^{iter}; Figure 2). Upward shifts in AICs and BICs (Figure 3) occur when the associated optimizer updates a new covariate group for the first time.

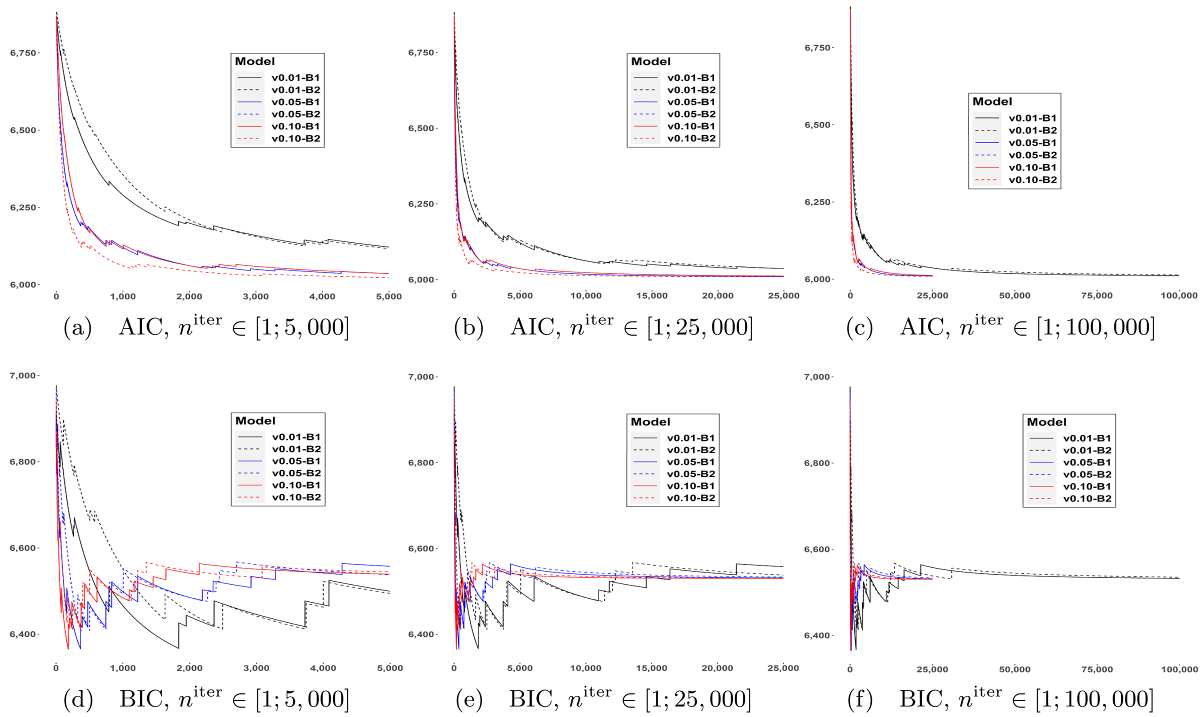
Figures 4a, 4c, and 4d show that in-sample log-likelihoods converge, whereas out-of-sample



(a) $\{\hat{\boldsymbol{\eta}}, \nu = .01\}$ (b) $\{\hat{\boldsymbol{\eta}}, \nu = .05\}$ (c) $\{\hat{\boldsymbol{\eta}}, \nu = .10\}$

(d) $\{\hat{\boldsymbol{B}}, \nu = .01\}$ (e) $\{\hat{\boldsymbol{B}}, \nu = .05\}$ (f) $\{\hat{\boldsymbol{B}}, \nu = .10\}$

Figure 2: RMSE (y-axis) over the number of iterations (x-axis), whereby $\text{RMSE}_{b_k}^{\text{agg}}/\text{RMSE}_{\eta_k}^{\text{agg}}$ for $k = 1, 2$ are visualized by lines "xi-B1", "xi-B2", "sigma-B1", and "sigma-B2". Additionally, the lines "agg-B1" and "agg-B2" are linked to $\text{RMSE}_{b}^{\text{agg}}/\text{RMSE}_{\eta}^{\text{agg}}$. Solid and dotted lines refer to model types B1 and B2, respectively.

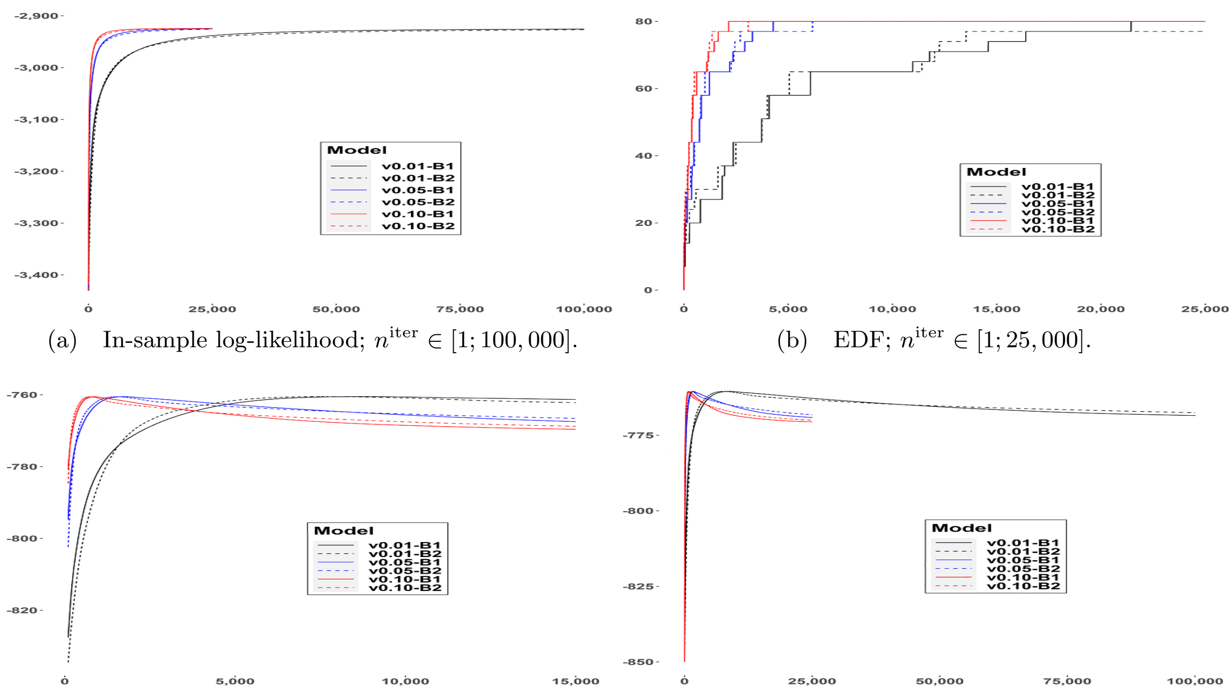


(a) AIC, $n^{\text{iter}} \in [1; 5,000]$ (b) AIC, $n^{\text{iter}} \in [1; 25,000]$ (c) AIC, $n^{\text{iter}} \in [1; 100,000]$

(d) BIC, $n^{\text{iter}} \in [1; 5,000]$ (e) BIC, $n^{\text{iter}} \in [1; 25,000]$ (f) BIC, $n^{\text{iter}} \in [1; 100,000]$

Figure 3: AIC/BIC (y-axis) over the number of iterations (x-axis). The terms "v0.01/v0.05/v10" refer to the step size v. The lines "v0.05-B1" and "v0.05-B2" are almost identical in figures (a), (b), and (c).

log-likelihoods reach a global maximum and worsen afterwards. Notice that log-likelihoods of *gamboostLSS* are not always better than *GAMLSSBoostAlt*. *GamboostLSS* directly maximizes log-likelihood functions, whereas *GAMLSSBoostAlt* imposes the restriction of alternatively updates. Such a restriction introduces a bias in model parameters and subsequently in associated log-likelihoods functions. However, B2-type models do not necessarily yield worse log-likelihoods, both in- and out-of-sample. This could be explained as follows: log-likelihood functions depend on the shape as well as the scale parameter. Since *gamboostLSS* updates one GPD parameter consecutively for many iterations, the resulting coefficient solution paths could deteriorate. Therefore, log-likelihoods in subsequent iterations do not always improve under B1 since the current position of the model coefficients could be worse compared to the previous iteration. For high n^{iter}, log-likelihoods of both optimizers converge towards similar values. Figure 4b illustrates that covariate groups, that have not yet been updated, on average are updated sooner when ν is high.



(a) In-sample log-likelihood; $n^{\text{iter}} \in [1; 100,000]$.

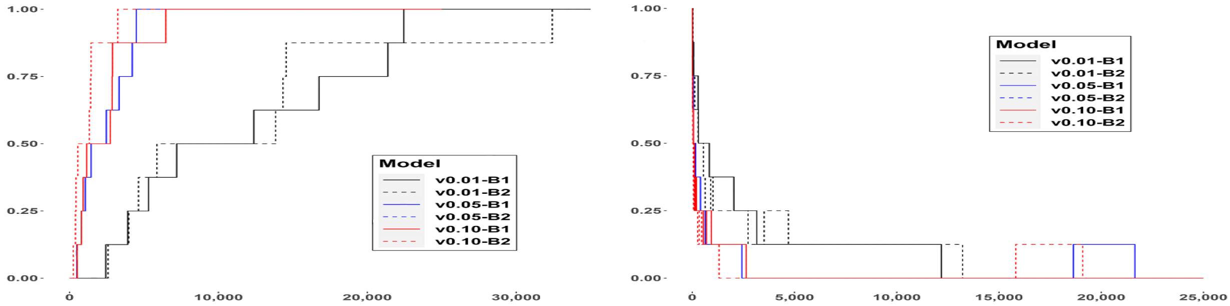
(b) EDF; $n^{\text{iter}} \in [1; 25,000]$.

(c) OOS-LLH; $n^{\text{iter}} \in [100; 15,000]$.

(d) OOS-LLH; $n^{\text{iter}} \in [1; 100,000]$.

Figure 4: Log-likelihoods and equivalent degrees of freedom (y-axis) over the number of iterations (x-axis). The terms "v0.01/v0.05/v10" refer to the step size v.

With regard to the variable selection, Figures 5b and 5a demonstrate that FNR and FPR of informative and uninformative covariates start at 1 and 0 since coefficients are zero-initialized, and then steadily converge to 0 and 1 for n^{iter} sufficiently high. The speed of converge positively depends on step size ν. Moreover, there is a trade-off between the false positive rates of noisy covariates and the false negative rates of the informative covariates. Therefore, optimizers *gamboostLSS* and *GAMLSSBoostAlt* are not able to optimize both rates and hence we use regularization techniques (Section 5.1.2) to tackle this issue. Additionally, not all covariates



(a) FPR; $n^{\text{iter}} \in [1; 35,000]$.

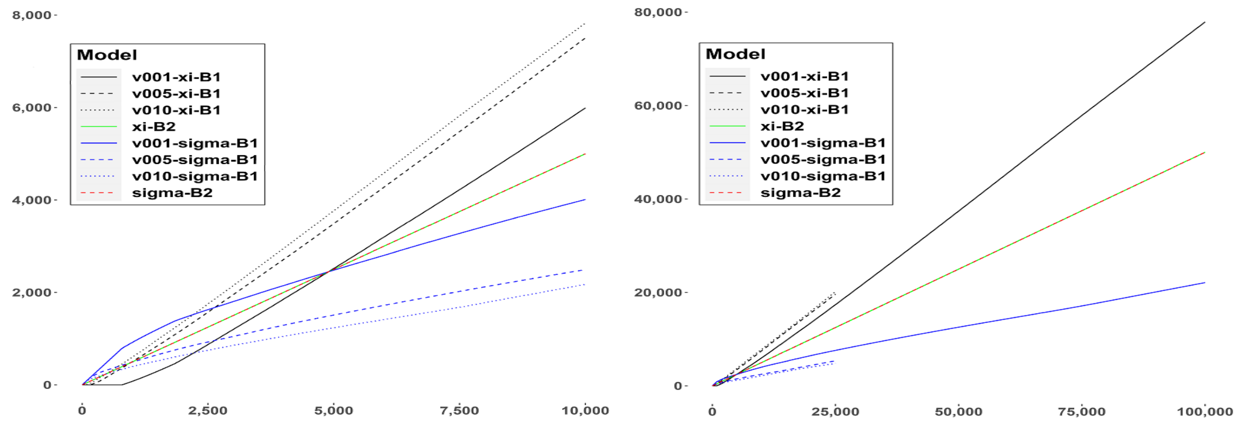
(b) FNR; $n^{\text{iter}} \in [1; 25,000]$.

Figure 5: FPR of uninformative and FNR of informative covariates (y-axis) over the number of iterations (x-axis). The terms "v0.01/v0.05/v10" refer to the step size v.

are updated once when $n^{\text{iter}} = 100,000$. In our simulation study, such covariates frequently are uninformative. However, this phenomenon is undesired for real-world covariate data sets, since zero-coefficients in unpenalized frameworks effectively ensure that subsequent regularization procedures likely will not update these coefficients to non-zero when using adaptive weights, explained in Appendix A.5.

5.1.4. *Comparison of gradient boosting updating schemes.* Let r^{iter} be the aggregated number of GPD shape parameter updates over the total boosting optimizer iterations n^{iter}. In Figure 6, we observe that GPD shape and scale parameters are on average updated a few and many times for the first thousand iterations, respectively. In fact, for $\nu = .01$ under *gamboostLSS* (line "v001-xi-B1"), the GPD shape is never updated in the first 500 iterations. Furthermore, Figure 6 demonstrates that the ratio r^{iter} increases when n^{iter} increases under *gamboostLSS*. In fact, ξ coefficients are updated many times when $n^{\text{iter}} > 5,000$. In our test cases, the associated LLFs never converged. Hence, the number of GPD shape and scale parameter updates heavily depend on n^{iter} under *gamboostLSS*. Due to this dependency, we find that the updating scheme of *gamboostLSS* is unpredictable, whereas the updating scheme of *GAMLSSBoostAlt* is fixed and known before model estimation and thus predictable. All in all, the function of GPD shape (scale) parameter with respect to the log-likelihood function is extremely flat (steep) for the first iterations, whereas this function is steep (flat) for a sufficiently high n^{iter}. Moreover, Figure 6 demonstrates that the intersection of the number of GPD shape and scale parameter updates ($r^{\text{iter}} = 0.5$), is inversely proportional to the step size ν. Additionally, the update ratio r^{iter} appears to converge towards 1 for sufficiently high n^{iter}.

For other specifications, Tables 20 and 21 in Appendix A.8 show that $C^{\text{sim}} = 15$ on average yields lower ratios r^{iter}. The reason is that, due to the higher number of covariates, more iterations are required to attain a bliss level of GPD scale parameter updates before GPD shape parameters impact the associated LLF substantially more. Moreover, these tables illustrate that



(a) $n^{\text{iter}} \in [1; 10,000]$.

(b) $n^{\text{iter}} \in [1; 100,000]$.

Figure 6: Number of updates (y-axis) over the number of iterations (x-axis). The terms "v0.01/v0.05/v10" refer to the step size v, "xi/sigma" refer to the corresponding GPD parameter, and "B1/B2" refer to the two optimizers B1 and B2. The lines 'xi-B2' and 'sigma-B2' are almost identical because B2-type models alternately update the two GPD parameters.

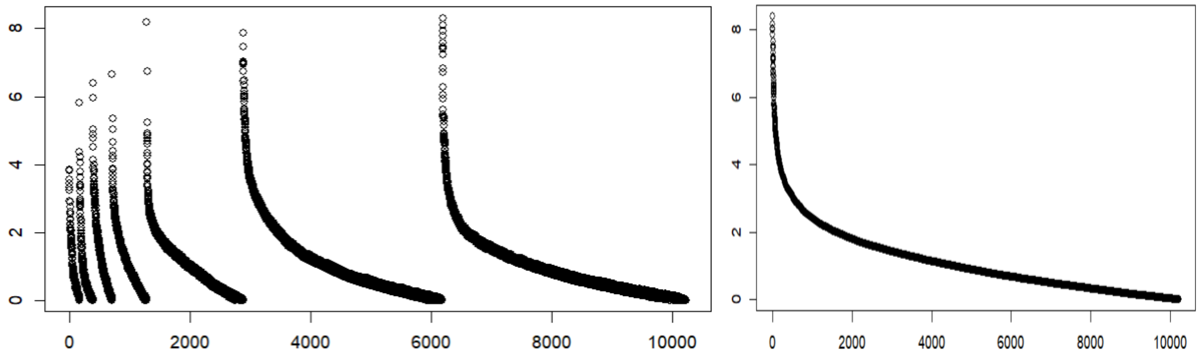
high ρ requires high n^{iter} to filter out the noise.

5.2. *Real-world application.* We investigate the performance of our proposed models using UniCredit's OpRisk loss data. We model LFEs/LSEs conditional on the following covariate sets, which include the seven risk categories, and additionally the following covariates: none (model M1), 20 economic factors (model M2, M2-met/M2-cat), the covariates in model M2, whereby we apply a variable selection method (model M3, M3-met/M3-cat).

The structure of this section is as follows: we present our loss severity and frequency models in Sections 5.2.1 and 5.2.2. In Section 5.2.3, we conduct a scenario analysis and subsequently estimate capital charges.

5.2.1. *Loss severity exceedances.* Figure 7 presents the log-LSEs of the losses above our threshold. We observe that the shape and scale of the losses differ among the risk categories. Particularly, the distribution of the risk-category-specific LSEs differs between (1) BDSF, DPA, EPWS, and IFRAUD; (2) CPBP, EDPM, and EFRAUD. Hence, associated GPD parameters should reflect these differences.

To identify risk drivers of the loss severity exceedances, we use a model in which we apply L_1-norms to filter the signal from the noise and to tackle overfitting, and apply adaptive weights from a similar unpenalized model, and use optimizer B1. The results of this model (model M3-cat) are shown in Table 7. We thereby find that the factor Tier-I capital ratios is identified for the GPD shape parameter, whereby the three upper covariate groups have been fused. Furthermore, UR IT and VIX USA exhibit high coefficients. With regard to the risk categories, CPBP-EPWS and BDSF-EDPM have been fused. This means that the LSE distributions are similar for these



(a) Risk-category-specific log-LSEs (b) Aggregated log-LSEs

Figure 7: Log-LSEs (y-axis) over the loss indices (x-axis), sorted on the risk-category-specific LSE levels. The risk categories in (a) from left to right are BDSF, DPA, IFRAUD, EPWS, EFRAUD, EDPM, and CPBP.

Table 7: Estimated LSE parameters over all 38 time periods using model M3-cat in a penalized λ framework, whereby the optimal tuning parameter using BIC is $\lambda^{\text{opt}} = 1.572$. We use use unpenalized M2-cat estimates in a similar context (Table 9) as adaptive weights.

	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
Intercept*	0.01	NA	NA	NA	NA	NA
RC*	0.08	-1.16	-0.32	-0.33	-0.68	-0.68
DGR	0	0	0	0	0	NA
LR	0	0	0	0	NA	NA
PRF	0	0	0	0	NA	NA
TCR	-0.01	0.14	0.14	0.14	NA	NA
UCSR	0	0	0.05	0.02	0.02	NA
GDP EU	0	0	0	0	0	NA
GDP IT	0	0	-0.03	-0.03	NA	NA
HPI EU	0	0	0	0	0	NA
LOR EU	0	0	0	0	NA	NA
LOR IT	0	-0.04	-0.04	-0.04	NA	NA
M1 EU	0	0	0	0	NA	NA
UR EU	-0.05	-0.08	-0.08	-0.08	-0.08	NA
UR IT	0	-0.07	-0.09	-0.09	-0.19	NA
LIR IT	0	0	0	0	NA	NA
MIB IT	0	0	0	0	NA	NA
SIR IT	0	0	0	0	0	NA
SP USA	0	0	0	0	0	NA
TRSI EU	0	0	0	0	NA	NA
VIX USA	0.06	-0.01	-0.01	-0.01	0.05	NA
VFTSE UK	0	0	0	0	0	NA

Note: There is only 1 intercept for each GPD parameter, and parameter estimates of the risk categories are in the order: IFRAUD, EFRAUD, EPWS, CPBP, EDPM, BDSF. In the rows that correspond to the economic factors, "NA" values mean that the covariate group does not exist for that economic factor. The interval limits and number of groups for each economic factor are in Appendix A.4, Table 17.

pairs. Parameter estimates of σ are not shown since the coefficients of the economic are all shrunk to zero, whereas the parameters belonging to the risk categories are identical to those of Table 10. All factors that are linked to the scale parameter are unidentified due to the levels of ξ and σ being vastly different. It is challenging to shrink both GPD parameters adequately since differences in levels require differences in the levels of the tuning parameters.

In this and the following paragraphs, we present other LSE-models to attain a deeper un-

Table 8: Estimated LSE parameters over all 38 time periods using model M2-cat in an unpenalized GAMLSS framework using the IWLS optimizer.

	ξ						σ					
	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
Intercept*	0	NA	NA	NA	NA	NA	0	NA	NA	NA	NA	NA
RC*	-0.95	-0.99	-0.76	-0.82	-0.28	-0.28	3.39	1.73	2.4	2.41	1.65	1.3
DGR	-0.39	-0.69	-0.54	-0.22	-0.32	NA	1.69	2.04	1.89	1.71	1.27	NA
LR	0.57	0.69	0.28	-0.03	NA	NA	0.25	0.25	0.29	1	NA	NA
PRF	0.2	0.1	0.21	0.12	NA	NA	-0.08	0.13	-0.09	-0.05	NA	NA
TCR	-0.33	-0.23	0.05	0.23	NA	NA	0.37	-0.04	0.15	-0.22	NA	NA
UCSR	-0.14	-0.09	0.07	0.03	0.07	NA	0.2	-0.04	0.01	-0.4	-0.31	NA
GDP EU	0.16	-0.09	-0.09	0.2	0.38	NA	-0.35	0.29	-0.07	-0.17	-0.42	NA
GDP IT	0.04	0.05	-0.16	0.02	NA	NA	0.04	0.02	-0.13	-0.11	NA	NA
HPI EU	-0.09	0.14	0.16	-0.2	0.11	NA	-0.08	0.09	0.02	-0.18	-0.07	NA
LOR EU	0.03	-0.15	0.03	-0.07	NA	NA	0.27	0	-0.09	0.3	NA	NA
LOR IT	-0.16	-0.15	0.02	0.1	NA	NA	0.38	0.16	-0.06	0.01	NA	NA
M1 EU	0.03	-0.15	0.22	-0.01	NA	NA	-0.27	0.27	-0.15	0.29	NA	NA
UR EU	-0.14	0	0.18	-0.21	0.14	NA	0.21	-0.08	-0.04	-0.24	-0.44	NA
UR IT	0.3	0	-0.13	0.31	-0.38	NA	-0.51	0.22	-0.1	0.16	0.18	NA
LIR IT	0.05	-0.21	-0.02	-0.01	NA	NA	0.04	0.13	-0.03	-0.19	NA	NA
MIB IT	-0.09	0.12	-0.24	0.2	NA	NA	-0.12	-0.04	0.27	-0.52	NA	NA
SIR IT	-0.05	0.16	-0.05	-0.07	-0.11	NA	0.03	-0.3	0.21	0.07	0.02	NA
SP USA	-0.08	-0.03	0.21	-0.13	0	NA	0.09	-0.21	-0.27	0.48	-0.01	NA
TRSI EU	-0.14	0.13	-0.01	-0.04	NA	NA	-0.07	0.02	0.08	-0.05	NA	NA
VIX USA	0.15	-0.18	0.21	-0.38	-0.01	NA	0.08	-0.22	0	0.15	-0.05	NA
VFTSE UK	0.14	0.04	-0.09	-0.11	-0.01	NA	-0.18	-0.23	0.36	0.08	-0.01	-

Note: There is only 1 intercept for each GPD parameter (presented in the column "Group 1"), and risk-category-specific parameter estimates are from left to right: IFRAUD, EFRAUD, EPWS, CPBP, EDPM, BDSF (presented in the row "RC*"). In the rows that correspond to the economic factors, "NA" values mean that the covariate group does not exist for that economic factor. The interval limits and number of groups for each economic factor are in Appendix A.4, Table 17.

Table 9: Estimated LSE parameters over all 38 time periods using model M2-cat in an unpenalized framework and the *gamboostLSS* gradient boosting optimizer for $n^{\text{iter}} = 100,000$ and $\nu = .01$.

	ξ						σ					
	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
Intercept*	-0.34	-	NA	NA	NA	NA	7.14	-	NA	NA	NA	NA
RC*	-0.48	-0.55	-0.31	-0.36	0.16	0.15	2.94	1.29	1.95	1.96	1.21	0.86
DGR	0	0	0	0	0	NA	0	-0.01	-0.06	0.01	-0.1	NA
LR	0	0	-0.01	0	NA	NA	-0.03	-0.12	0.04	-0.07	NA	NA
PRF	0	0.01	0	-0.02	NA	NA	-0.01	0	-0.01	0.02	NA	NA
TCR	-0.14	0.06	0.06	-0.03	NA	NA	-0.12	0.06	0	-0.04	NA	NA
UCSR	-0.05	-0.21	0.05	-0.04	-0.02	NA	-0.02	-0.03	0.01	-0.09	-0.09	NA
GDP EU	0	0	0	-0.01	-0.01	NA	-0.03	0.02	-0.02	-0.08	-0.16	NA
GDP IT	0.04	0.04	-0.09	-0.02	NA	NA	-0.01	0.01	-0.02	-0.02	NA	NA
HPI EU	-0.01	0	0.01	-0.02	0.01	NA	0	0	0	0	0	NA
LOR EU	0	0	0	0	NA	NA	0.02	-0.02	-0.01	0.02	NA	NA
LOR IT	0.04	-0.05	-0.01	0.01	NA	NA	0.01	-0.04	-0.01	0.02	NA	NA
M1 EU	0.09	-0.04	-0.01	0	NA	NA	-0.11	0.02	-0.03	-0.05	NA	NA
UR EU	-0.07	-0.11	0.1	0.05	-0.03	NA	0	-0.04	0.01	-0.01	-0.01	NA
UR IT	0	0.05	-0.05	0.06	-0.09	NA	-0.01	-0.02	0.01	0	-0.02	NA
LIR IT	0.03	-0.03	0.03	-0.03	NA	NA	-0.01	-0.03	0.02	-0.06	NA	NA
MIB IT	0	0	0	0	NA	NA	0	0	0	0	NA	NA
SIR IT	-0.03	0.02	-0.01	0	0	NA	-0.01	-0.09	-0.01	0.03	-0.02	NA
SP USA	0.05	0.08	-0.07	-0.08	-0.02	NA	0.03	-0.02	0.02	-0.02	0	NA
TRSI EU	0	0	0	0	NA	NA	-0.12	-0.03	0.05	-0.01	NA	NA
VIX USA	0.09	-0.01	0.03	-0.06	0.04	NA	0.08	-0.08	0.01	-0.05	0	NA
VFTSE UK	0	0	0	0	0	NA	-0.17	-0.04	-0.07	-0.02	0	NA

Note: Identical to Table 8.

derstanding of the relation between the loss severity exceedances and the involved economic factors. Table 8 presents the results from the using IWLS updating schemes, whereby BIC is the stopping criteria. Unlike the gradient boosting optimizers, the log-likelihood function of the IWLS optimizer converged, namely after 382 iterations. IWLS does not put any weight into the model intercept, whereas economic factors exhibit high weights. The parameter estimates in Table 8 illustrate the trade-off between the two GPD parameters. For example, the GPD shape parameter estimates of the leverage ratios decrease from group 1 to 6, whereas GPD scale parameter estimates increase from group 1 to 6. Therefore, the effect of LR is unclear. Furthermore, Table 8 indicates a relation between the signs of the two GPD parameters ξ and σ for the economic factors. Economic factors that have positive ξ coefficient frequently have negative associated σ coefficient for the same group. Moreover, we identify high absolute levels of the deposit growth rate factor. High deposit growth rates are associated with high shape and scale parameter estimates. Similarly, the Tier-I capital ratio parameter estimates for both GPD parameters are highly negative in column "Group 1". This means that low financial market volatility indicates low OpRisk losses. The factors UR EU and UR IT, which are highly correlated, illustrate the consequence of multicollinearity. Associated coefficients are substantially non-zero, and the signs of these factors are complementary. Therefore, UR EU and UR IT effectively dominate other factors. This could be explained by the fact that two highly comoving factors exhibit substantially diversification effects from a modelling perspective.

Next, we model LSEs using model M2-cat and *gamboostLSS* for $n^{\text{iter}} = 100,000$ and $\nu = 0.01$, as shown in Table 9.[3] We thereby discover that $0 < \xi < 1$ holds for all risk categories, after applying an exponential transformation to the risk-category-specific parameter estimates to revert the logarithmic link function to the covariates. For illustration, reverted ξ estimates for DPA and IFRAUD are $\exp(-0.34)$ and $\exp(-0.34 - 0.48)$. In line with Hambuckers et al. (2018), IFRAUD exhibits the highest GPD scape parameter estimate. With regard to the economic factors, it appears that leverage ratios, tier-I capital ratios, unemployment rates in the European Union, and USA S&P volatility index explain model variation since these factors have at least one group that has an absolute coefficient above 0.10. A plausible explanation is that optimal stopping conditions for gradient boosting optimizers are yet unclear and thus we let the gradient boosting optimizer run for $n^{\text{iter}} = 100,000$. Therefore, the differences between B1 and B2 are minimal since our findings in Section 4.1.1 illustrate that B1 and B2 perform similarly for sufficiently high n^{iter}.

Table 10 compares the updates and log-likelihood contributions of the two gradient boosting optimizers using model M2-cat over all 38 time periods. We calculate these contributions as follows: in the current iteration, we evaluate what would happen if we would update a single involved covariate (we have 40 involved covariates under B1, and 20 under B2). Due to the changes in model parameter estimates for a particular covariate, we obtain a changed linear predictor estimate. We then calculate the associated log-likelihood for this particular covariate using the changed linear predictor estimates. After calculating the set of all log-likelihoods, we

[3]B2-type models perform similar to B1-type models, thus these results are not presented.

Table 10: Percentage of updates per covariate and log-likelihood improvement for model M2-cat using optimizers B1 and B2 for $n^{\text{iter}} = 100,000$ and $\nu = 0.01$.

	GamboostLSS				GAMLSSBoostAlt			
	ξ		σ		ξ		σ	
Factor	%	LLH$^+$	%	LLH$^+$	%	LLH$^+$	%	LLH$^+$
Intercept	100	6.17	100	34.58	100	4.68	100	29.65
RC	15.45	49.06	45.5	379.94	20.03	314.94	37.0	359.17
DGR	0	0	1.84	1.9	0.04	0.01	1.23	1.1
LR	0.26	0.06	2.1	3.15	0.23	0.2	1.6	2.7
PRF	0.75	0.16	0.47	0.24	0.29	0.08	0.37	0.69
TCR	1.23	6.51	0.8	5.25	1.77	16.72	0.88	3.49
UCSR	3.29	3.42	1.44	3.41	3.49	11.87	1.07	2.59
GDP EU	0.21	0.07	1.12	2.97	0.65	0.96	0.96	2.12
GDP IT	1.59	1.71	0.54	0.27	2.36	7.26	0.46	0.16
HPI EU	0.55	0.14	0.1	0.02	0.65	0.8	0.01	0.02
LIR IT	1.72	0.48	1.11	0.84	3.17	4.67	0.84	0.86
LOR EU	0	0	0.22	1.61	0.09	1.02	0.27	1.04
M1 EU	2.21	1.21	1.58	3.7	2.94	4.46	1.24	3.61
UR EU	1.81	6.61	0.2	1.86	1.92	13.29	0.09	0.84
UR IT	2.21	1.27	0.68	0.37	3.06	5.27	0.32	0.31
LIR IT	1.72	0.48	1.11	0.84	3.17	4.67	0.84	0.86
MIB IT	0.23	0.05	0.04	0.02	0.84	1.03	0	0
SIR IT	0.23	1.17	1.22	2.52	0.16	2.09	0.86	2.51
SP USA	2.23	2.04	0.14	2.07	3.42	7.41	0.16	2.4
TRSI EU	0	0	0.89	4.76	0	0	0.8	4.61
VIX USA	2.41	1.29	0.5	6.84	2.46	7.29	0.57	6.25
VFTSE UK	0	0	1.87	10.96	0.84	0.92	1.22	10.08
All	37.47	82.14	62.53	469.52	50	411.17	50	436.07

Note: Columns "%" show the percentage of updates per factor for $n^{\text{iter}} = 100,000$, whereby the model intercepts are updated in each iteration. Columns LLH$^+$ illustrate the total log-likelihood improvement for all updates of the corresponding covariate. Table elements that are exactly equal to 0 mean that the corresponding covariate received no updates whatsoever.

update the covariate that positively impacts the log-likelihood the most (which is the covariate that maximizes the log-likelihood). The associated contribution, summed over all iterations, is presented in columns "LLH$^+$".

In Table 10, we find that model terms that belong to the GPD scale parameter receive the most updates under B1 (62.53%). This is mainly due to the model terms of the GPD scale parameter of the risk categories (45.5%) contributed the most to the log-likelihood. Economic factors TCR, UCSR and UR EU contribute the most to the log-likelihood of the GPD shape parameter. The high contribution of TCR is in line with Groll et al. (2019) and could be due to banks anticipating large losses and hence they improve their core capital positions in advance of the anticipated future extreme losses. Moreover, TCR, VFTSE UK, and VIX USA contribute the most to the GPD scale parameters. The impact of the financial market volatility indices could be explained by the fact that high associated values signal an expansion. However, high financial market volatility indices also indicate a recession and hence caution is warranted in the interpretation of these factors. For the interested reader, in Appendix A.10; Table 23, we present LSE estimates over all 38 time periods for models M1 and M2-met using unpenalized IWLS updating schemes, whereby we use IWLS because the number of model parameters is sufficiently low for IWLS to yield good performance.

5.2.2. *Loss frequency exceedances.* Next, we evaluate our loss frequency exceedances models. Similar to LSE models, we use models M1 - M3 but exclusively using metric covariates. Model M3 is constructed as follows: we re-estimate model M2 after removing insignificant economic factors (p-values > 0.05). Thereafter, we keep removing covariates based on p-values until three economic factors remain to attain a parsimonious model, namely model M4.

Table 11 presents LFE-models M1 - M4. We thereby find that the remaining economic factors in model M4 are DGR, MIB IT, and UR IT. Using the impact of these factors based on economic reasoning (explained in Section 3.3), we interpret these factors as follows: high deposit growth rates, low Italian unemployment rate, and high Milano Italia Borsa stock returns increase the likelihood of the occurrence of extreme OpRisk losses. Moreover, we discover that the LFEs significantly differ among risk categories due to associated close-to-zero p-values.

Table 11: NHPP distributed $\hat{\boldsymbol{\kappa}}$ coefficients over all 38 time periods. whereby the corresponding p-values are in parentheses. The row "CV-OOS-LLH" presents the out-of-sample cross-validated log-likelihood scores, aggregated over all risk categories and time periods.

Factor	Model M1	Model M2	Model M3	Model M4
Intercept	3.024 (0)	2.982 (0)	2.983 (0)	2.998 (0)
IFRAUD	0.123 (0)	0.123 (0)	0.123 (0)	0.123 (0)
EFRAUD	0.692 (0)	0.692 (0)	0.692 (0)	0.692 (0)
EPWS	0.331 (0)	0.331 (0)	0.331 (0)	0.331 (0)
CPBP	1.018 (0)	1.018 (0)	1.018 (0)	1.018 (0)
BDSF	-0.068 (0.05)	-0.072 (0.039)	-0.071 (0.04)	-0.069 (0.052)
EDPM	0.948 (0)	0.948 (0)	0.948 (0)	0.948 (0)
DGR	-	0.068 (0)	0.081 (0)	0.072 (0)
LR	-	-0.035 (0.073)	-	-
PRF	-	-0.013 (0.625)	-	-
TCR	-	0.006 (0.89)	-	-
UCSR	-	-0.161 (0)	-0.138 (0)	-
GDP EU	-	-0.073 (0.139)	-	-
GDP IT	-	0.031 (0.415)	-	-
HPI EU	-	0.025 (0.736)	-	-
LOR EU	-	0.313 (0.003)	0.269 (0)	-
LOR IT	-	-0.128 (0.116)	-	-
M1 EU	-	0.047 (0.24)	-	-
UR EU	-	0.686 (0)	0.699 (0)	-
UR IT	-	-0.594 (0)	-0.575 (0)	-0.271 (0)
LIR IT	-	-0.032 (0.303)	-	-
MIB IT	-	0.218 (0)	0.194 (0)	0.084 (0)
SIR IT	-	0.235 (0.002)	0.161 (0)	-
SP USA	-	-0.107 (0.006)	-0.102 (0)	-
TRSI EU	-	-0.009 (0.85)	-	-
VIX USA	-	-0.112 (0.04)	-0.097 (0)	-
VFTSE UK	-	0.004 (0.949)	-	-
AIC	3012.83	2208.67	**2201.2**	2343.05
CV-OOS-LLH	1554.9	1943.4	**1232.4**	1241.0

With regard to the economic factors, model M2 in Table 11 illustrates that many are significant. Nevertheless, the inclusion of economic factors barely impacts the level of the risk-category-specific parameter estimates. Model M3 is preferred by comparing AICs and out-of-sample cross-validated log-likelihood scores (CV-OOS-LLH). Moreover, model M2 exhibits an inferior CV-OOS-LLH score. Model M2 includes all twenty economic factors, and these factors are highly collinear. Therefore, taking large positive and negative positions in such factors could potentially yield vast diversification effects, resulting in unstable out-of-sample LFE predictions and consequently an inferior CV-OOS-LLH score. Another perspective is that model M2 overfits the data (due to the inclusion of many economic factors) and hence underperforms out-of-sample. For the interested reader, we conduct further analyses to attain a better understanding of the relation between loss frequency exceedances and economic factors in Appendix A.11.

5.2.3. *Scenario analysis for economic conditions of the business cycle.* Independent scenario analyses are conducted for both the loss frequency and severity for three economic conditions: recession (SC1), intermediate (SC2), and expansion (SC3). We compare these conditions to a constant scenario (SC0), which uses long-term average risk-category-specific parameter estimates. Our approach for the scenario analysis is to fix the involved economic factors to plausible values such that they represent scenarios SC1 - SC3. In this procedure, we evaluate the impact of the scenarios by "impact coefficients". These coefficients are calculated as the sum of the product of predetermined covariate values and associated parameter estimates. Thereafter, we add the impact parameter, which is scenario-specific, to the risk-category-specific parameter estimates of the constant scenario to obtain scenario-specific parameter estimates for the risk categories.

For the loss severity, we use the key risk driving economic factors TCR, UR IT, and VIX USA (elaborated in Section 5.2.1). We estimate a model over all 38 time periods using the risk categories and these economic factors as covariates. Re-estimation is necessary since other models include economic factors that we do not use in the scenario analysis and hence these factors distort the parameter estimates of the covariates that we do use. Furthermore, the penalized M2-cat model, which identifies these factors, primarily serves to identify risk drivers since the connected parameter estimates are unstable. All in all, we estimate a new model using B1, whereby $n^{\text{iter}} = 100,000$ and $\nu = .01$. The corresponding parameter estimates that belong to the risk categories represent the constant scenario SC0, and are presented in Table 12. Although we already interpreted parameter estimates (Section 5.2.1), we further analyse the impact of TCR due to its anomalous behaviour. Moderate Tier-I capital ratios yield a drastic increase in ξ (group 3) and σ (group 2) parameters, whereas extremely low Tier-I capital ratios slightly substantially decrease model coefficients. Valencia (2016) argues that banks self-insure against future shocks through holding more bank capital, whereby a higher uncertainty around future losses is associated with a higher level of self-insurance. Another possible explanation is that extremely low and high Tier-I capital ratios are not necessarily related to an economic condition but reflect the internal risk policy and/or UniCredit's state of health at that time.

Table 12: Scenario analysis, LSEs; unpenalized GAMLSS parameter estimates.

	ξ						σ					
	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
Intercept*	-0.49	-	-	-	-	-	8.17	-	-	-	-	-
RC*	-0.54	-0.62	-0.36	-0.40	0.12	0.11	2.98	1.32	1.99	1.98	1.23	0.91
UR IT	0.07	0.04	-0.15	-0.14	-0.15	NA	0.01	-0.05	-0.20	-0.16	-0.21	NA
VIX USA	0.18	0.20	0.25	0.17	0.21	NA	0.09	-0.13	0.02	-0.02	0.05	NA
TCR	-0.20	0.07	0.19	0.07	NA	NA	-0.01	0.23	0.10	0.14	NA	NA

Note: There is only 1 intercept for each GPD parameter, and parameter estimates of the risk categories are in the order: IFRAUD, EFRAUD, EPWS, CPBP, EDPM, BDSF. In the rows that correspond to the economic factors, "NA" values mean that a particular covariate group does not exist for the corresponding economic factor.

Since we categorize covariates for LSE-models, we link specific covariate groups (one group per scenario) to economic conditions using economic reasoning. Table 13 presents these predetermined covariate groups, whereby each covariate group is linked to a corresponding model parameter estimate (Table 12) and interval limit (Appendix A.4; Table 17). For illustration, UR IT has a negative impact on the loss severity, thus we select groups 4, 1, and 0 for scenarios SC1 - SC3. The factor that proxies U.S.' financial market volatility (VIX USA) requires additional explanation: we presume high values during recessions (group 4), low values during an intermediate scenario (group 1), and relatively high values during expansions (group 3), in line with Hambuckers et al. (2018). For categorized LSE-models, impact coefficients are obtained as the sum of the product of the covariate groups (conditional on the scenario) and associated coefficients, whereby only one group in a categorical covariate is equal to 1, and all other groups are equal to 0. For illustration, the impact coefficient in a recession is equal to the sum of 0 (TCR), -0.14 (UR IT) and 0.17 (VIX USA) = 0.03 (Table 13).

Table 13: Scenario analysis, LSEs; predetermined covariate groups that represent scenarios SC1 - SC3. For clarification, we added the corresponding coefficients and interval limits.

		Recession				Intermediate				Expansion			
	Impact	Group	$\hat{\xi}$	$\hat{\sigma}$	Interval	Group	$\hat{\xi}$	$\hat{\sigma}$	Interval	Group	$\hat{\xi}$	$\hat{\sigma}$	Interval
TCR	+	0	0	0	$[-\infty, 0.06]$	1	-0.20	-0.01	[0.06, 0.07]	2	0.07	0.23	[0.07, 0.08]
UR IT	-	4	-0.14	-0.16	[10,12]	1	0.07	0.01	[6,7]	0	0	0	$[-\infty, +6]$
VIX USA	*	4	0.17	-0.02	[15,18]	1	0.18	0.09	[11.9, 13]	3	0.19	0.02	[13, 15]
Impact coefficient	NA	NA	0.03	-0.18	NA	NA	0.05	0.09	NA	NA	0.26	0.25	NA

Note: Group 0 refers to the reference group. The coefficients of this table match the coefficients in Table 12, and interval limits are as presented in Appendix A.4; Table 17.

For the loss frequency, we use the average number of losses above our threshold to represent SC0. For illustration, BDSF has an average of 169/38 losses (number of losses are in Table 2). For scenarios SC1 - SC3, we use the included economic factors in LFE-model M4, which are DGR, MIB IT, and UR IT. We use both the impact (+/-/*) and levels of economic factors to determine plausible values for the scenarios, shown in Table 14. Since all covariates are transformed to standard normal distributions before model estimation, we use normalized equivalents of the factors in the scenario analysis. Normalized equivalents are obtained by an approximation: we add the prespecified economic factor value to the associated vector in the model matrix that consists of 38 scalars (equal to the number of time periods) for a given economic factor, and thereafter standard normalize this (1×39) vector. We subsequently extract the value for the factor that we added. Impact coefficients for LFE scenarios SC1 - SC3 are calculated as the sum of the product of normalized covariate values (Table 14) and associated model coefficients (Table 11; model M4). For illustration, the impact coefficient in a recession is equal to the sum of -1.08×0.072 (DGR) - 1.81×0.084 (MIB IT) + 2.67×-0.271 (UR IT) = -0.95 (Table 14).

Table 14: Scenario analysis, LFEs; original factor values and normalized equivalents using model M4 that represent scenarios SC0 - SC3.

Factor	Impact	Constant	Recession		Intermediate		Expansion	
		Norm. value	Value	Norm. value	Value	Norm. value	Value	Norm. value
DGR	+	0	-0.15	-1.08	0.03	-0.04	0.20	0.95
MIB IT	+	0	-0.20	-1.81	0.00	-0.03	0.20	1.76
UR IT	-	0	0.13	2.67	0.09	0.54	0.05	-1.59
Impact coefficient	NA	0	-0.95		-0.14		0.66	

Final scenario- and risk-category-specific parameter estimates for LFEs and LSEs are in Table 15, and are calculated as the sum of the corresponding risk-category-specific parameter estimate in a constant scenario and the impact coefficient, conditional on the scenario. For illustration, we calculate the Poisson and GPD parameter estimates for SC0 and SC1 for risk category DPA. The Poisson parameter estimates for SC0 and SC1 are 224/38 = 5.89 and exp(log(5.89)-0.95) = 2.27. The GPD shape parameter estimates are exp(-0.49) = 0.61 and exp(log(0.61)+0.03) = 0.63, and the GPD scale parameter estimates are exp(8.17) = 3,528 and exp(log(3,528)-0.18) = 2,948 for SC0 and SC1, respectively. Thereby, -0.95 ($\hat{\kappa}$), 0.03 ($\hat{\xi}$), and -0.18 ($\hat{\sigma}$) are the impact coefficients in a recession, and -0.49 and 8.17 are the model coefficients as presented by Table 12.

Table 15 demonstrates that loss severity exceedances are on average higher during favourable economic conditions due increased $\hat{\xi}$ and $\hat{\sigma}$, in line with our hypothesis that expansions drive extreme OpRisk losses. Therefore, the involved factors TCR, UR IT, and VIX USA are able to capture differences among economic conditions. The constant scenario frequently represents an average scenario, which means that the associated coefficients lie within the range of other scenarios.

Table 15: LFE and LSE estimates for scenarios: constant (SC0), crisis (SC1), intermediate scenario (SC2), and expansion (SC3). The column "Exp." shows the expected coefficient, calculated as $.25 \times (\hat{\kappa}_{\mathrm{SC1}} + \hat{\kappa}_{\mathrm{SC3}} + 2 \times \hat{\kappa}_{\mathrm{SC2}})$.

	LFE; $\hat{\kappa}$					LSE; $\hat{\xi}$				LSE; $\hat{\sigma}$			
	SC0	SC1	SC2	SC3	Exp.	SC0	SC1	SC2	SC3	SC0	SC1	SC2	SC3
BDSF	4.45	1.71	3.85	8.62	4.51	0.69	0.71	0.72	0.94	8,778	7,334	9,657	11,284
CPBP	106.18	40.93	92.00	205.74	107.67	0.41	0.43	0.43	0.56	25,559	21,355	28,121	32,857
DPA	5.89	2.27	5.11	11.42	5.98	0.61	0.63	0.64	0.84	3,528	2,948	3,882	4,536
EDPM	86.89	33.50	75.29	168.37	88.11	0.69	0.71	0.72	0.95	12,038	10,058	13,245	15,476
EFRAUD	42.05	16.21	36.44	81.48	42.64	0.33	0.34	0.35	0.46	13,198	11,027	14,521	16,967
EPWS	15.08	5.81	13.06	29.22	15.29	0.43	0.45	0.45	0.59	25,925	21,662	28,524	33,328
IFRAUD	8.37	3.23	7.25	16.21	8.48	0.36	0.37	0.37	0.49	69,433	58,014	76,393	89,259
All	268.91	103.66	233.00	521.06	272.68	-	-	-	-	-	-	-	-

With regard to the loss frequency exceedances, Table 15 illustrates that the number of estimated loss frequency exceedances is higher during favourable economic conditions, again in line with our hypothesis. We therefore conclude that the LFE parameter estimates for scenarios SC1 - SC3 are plausible and hence the associated economic factors DGR, MIB IT, and UR IT help in explaining differences in LFEs among economic scenarios. The estimate, aggregated over the risk categories, is around a factor 5 (521.06 versus 103.66) higher during expansions than crises. Let us assume that a crisis, intermediate condition, and expansion occur with probabilities .25, .5, and .25, respectively. Then, the expected number of extreme OpRisk losses is equal to $.25 \times (103.66 + 521.06 + 2 \times 233.00) \approx 272.68$, which is close to the true number of average losses, namely 268.91. A similar analysis for LFEs using LFE-model M3 (our preferred LFE-model) yields counter-intuitive insights, elaborated in Appendix A.12.

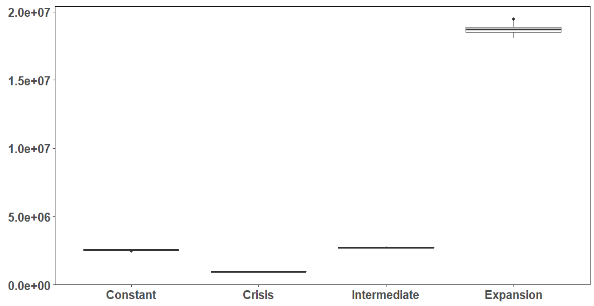
We subsequently estimate requested capital charge estimates, which are scenario- and risk-category-specific. To simulate total losses above our threshold, we use the associated parameter estimates as presented in Table 15, whereby 10^6 instances for 100 samples for each scenario/risk-category pair are drawn from the Poisson (LFE) and GPD (LSE) distributions. The 100 samples serve to quantify the uncertainty around the simulated instances. After the simulation of these instances, we add the risk-category-specific threshold to the loss series since we originally subtracted the threshold by applying the POT method. We fix the losses under our threshold to zero. Hence, this loss series consists of a 40,871-10,219 = 30,652 vector of zeros because our threshold (Section 3.2) excludes 75% of all losses.

The requested capital charge estimate is equal to the 99.9% quantile of the concatenated sorted loss series (also known as the 99.9% Value-at-Risk), which is a vector that includes the losses under as well as above our threshold, for a given scenario and risk category. Banks and regulators commonly use the 99.9% VaR measure to predict the level of extreme losses. We present the risk-category-specific capital charge estimates in Figure 8. We thereby observe that a crisis (SC1) yields the lowest charge, followed by the constant scenario (SC0) and intermediate scenario (SC2), whereas the expansion scenario (SC3) is responsible for most total losses. These findings are in contrast with Hambuckers et al. (2018) since their estimated capital charges are highest in their constant scenario. This is not in line with the general principle that a constant scenario, which is constructed by the long-term average parameter estimates, is by definition not an extreme case. Our capital charge analysis shows that the constant scenario yields moderate capital charge estimates. The clear pattern from the total loss increases for SC1 - SC2 - SC3 supports that our hypothesis, that economic factors proxy expansions, and that expansions drive extreme OpRisk losses. Ratios between the charge estimates among the scenarios are quite alike, in line with Hambuckers et al. (2018). This could be due to the fact that shifts in economic conditions impact all risk categories to a similar extent.

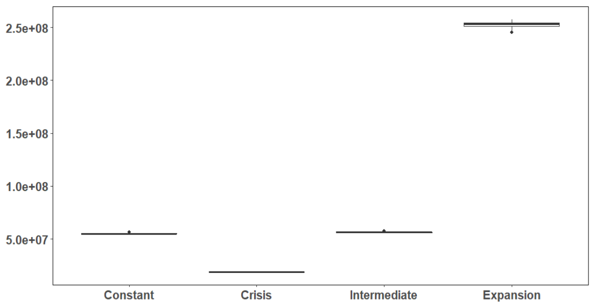
An upside of our capital charge framework is that the charges are partially based on categorized covariates. Since the LSE covariates are categorical, we now have a fixed number of possible capital charges, namely $5 \times 6 \times 6 \times 4 = 720$. This number is partially obtained by the fact that we use three economic factors in our final LSE model, namely TCR, UR IT and VIX USA, which contain 5, 6, and 6 covariate groups, respectively. Although the capital charges are stochastic, a final capital charge could be approximated for a sufficient number of instances. The second component concerns the loss frequency. Since LSEs drive total losses, we argue that LFEs could be fixed to one of the associated model estimates (Table 15) conditional on the current condition of the business cycle. This is a plausible framework since it is typically known what the current condition of the business cycle is and hence the associated $\hat{\kappa}$ parameter estimates can represent this condition.

With our approach, banks can reduce their liquidity risk because the set of 720 charges covers all possible charges. When the covariates in a new time period are known, we could place these covariate values into one of our dummy encoded categorical covariate groups. This set of associated covariate groups corresponds to one of the 720 known charges. Although Figure 8 only represents four out of 720 charges, all other charges can be estimated using a similar approach.

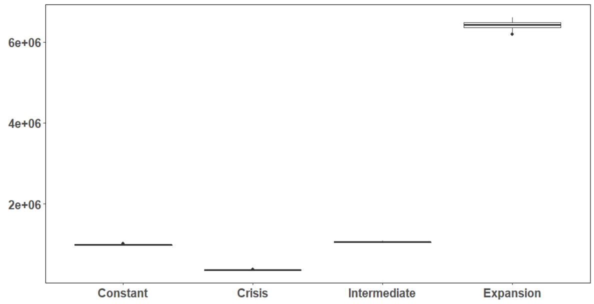
We emphasize that our capital charge estimates may be underestimates since we set the losses under our threshold to zero as an approximation. Nevertheless, it is unlikely that the losses under threshold are equal or above the 99.9% quantile and hence these losses likely do not affect the calculation of the requested capital charge. Additionally, we stress that we can not directly draw conclusions with regard to the levels of the capital charge estimates since UniCredit rescaled their losses with an unknown factor for privacy reasons. Therefore, the face values of the capital charge estimates can not be interpreted. Nevertheless, these capital charge estimates shed light on the differences of these charges among risk categories as well as the



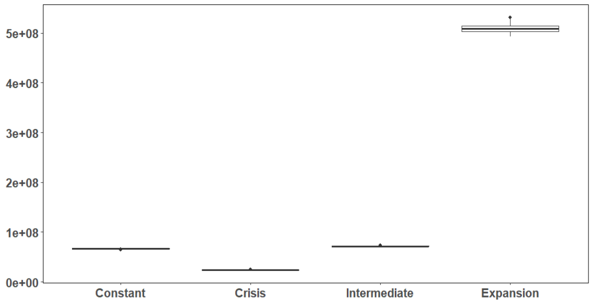
(a) BDSF



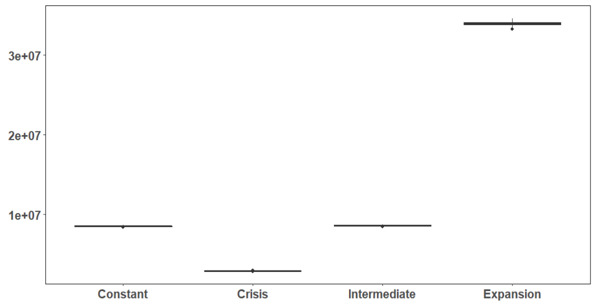
(b) CPBP



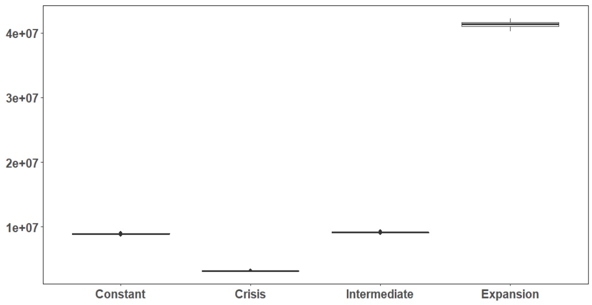
(c) DPA



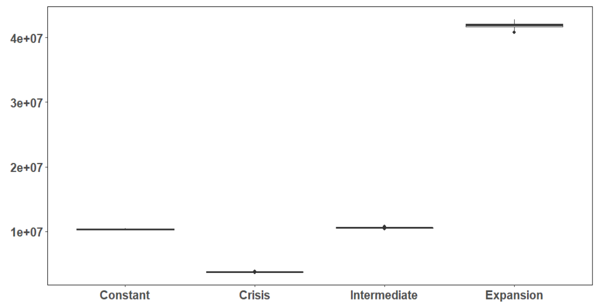
(d) EDPM



(e) EFRAUD



(f) EPWS



(g) IFRAUD

Figure 8: Scenario-specific and risk-category-specific capital charges estimates using Monte Carlo experiments.

sensitivity of the risk categories to several economic conditions of the business cycle.

6. Conclusion.

We independently model the loss frequency and severity using the hypothesis that economic factors proxy expansions, and expansions drive extreme operational risk losses. This hypothesis is empirically supported in Section 3.3. We thereby investigate several model optimizers to attain a better understanding of the relation between extreme OpRisk losses and economic categorized factors. Our simulation study illustrates that the scale parameter σ of the GPD is the driver of log-likelihood functions, when using gradient boosting type optimizers. Alternatively updating ξ and σ parameters yields slightly improves model parameter estimates. Furthermore, within our extreme value framework, it is necessary to early-stop boosting optimizers since associated in-sample and out-of-sample parameter estimates deteriorate after a bliss level of boosting optimizer iterations. However, finding adequate stopping conditions is challenging and hence must be further researched.

In our real-world application, we model extreme OpRisk losses of the global systemically important bank UniCredit. We thereby use 20 economic factors capturing a wide spectrum of financial information. Due to high levels of multicollinearity, it remains a challenge to truly identify economic key risk drivers of OpRisks. Our analysis demonstrates that TCR, UR IT, and VIX USA explain variation in loss severity exceedances, whereas the loss frequency exceedances could be partially explained by DGR, MIB, and UR IT. Our scenario analysis illustrates that these factors are able to explain differences in loss frequencies and severities of extreme losses among economic scenarios. Subsequently, the associated parameter estimates are used to estimate capital charges.

The banking industry can benefit from our methodology. The required capital charge is reduced from a infinite number of possibilities to 720 charges by the use of categorized economic factors. These charges are known after model estimation. Conditional on new economic information, the new capital charge can always be mapped to an existing capital charge. Therefore, banks could reduce liquidity risk, which subsequently reduces default risk. Additionally, the economic factors that drive OpRisk losses could be predicted and tracked to anticipate changes in future extreme OpRisk losses. Furthermore, we add to the literature by attaining a deeper understanding of current gradient and log-likelihood optimizers. Besides the banking industry, our *GAMLSSBoostAlt* optimizer could be used for other probability distributions than the GPD. Note that the GAMLSS model class can be used to model any probability distribution with up to four distribution parameters. Therefore, it seems plausible that other probability distributions, especially those that exhibit more than two distributional parameters, could benefit from alternatively updating its distributional parameters using *GAMLSSBoostAlt*.

With regard to suggestions for future research, Herawati et al. (2018) argue that a downside of L_1-norm regularization is the impossibility of properly dealing with multicollinearity among covariates. We therefore suggest investigating ways to deal with collinearity among covariates. Moreover, Ali et al. (2019) find that early-stopping gradient descent type optimizers are closely linked to L_2-norm regularization for proportional and overparametrized regimes. Similarly, Tibshirani (2015) find a link, within a classification framework, between L_1-norm regularization and gradient descent optimizers that behave as interpolators. We therefore recommend further researching gradient boosting optimizers using adequate stopping conditions. These optimizers may be modified such that multicollinearity could be tackled. Additionally, parameter uncertainty could be tackled by applying our methodology in a Bayesian framework.

Furthermore, we suggest researching a framework in which OpRisk losses are not immediately known in the current time period but one or a few subsequent time period(s). It seems likely that losses that occur in a given time period, are not identified immediately. For example, such losses could be related to internal/external fraud, software errors, or unintentional contract breaches. In particular, such an analysis is especially relevant since the majority of the total losses occur during expansions. Hence, part of these losses are presumably noticed during the next economic condition, which could be a recession. Liquidity positions of banks generally worsen during recessions. Therefore, huge "lagged OpRisk losses" could potentially put an immense amount of pressure on the liquidity of banks.

Lastly, our research is based on the hypothesis that economic factors proxy expansions, and expansions proxy extreme OpRisk losses. Since economic factors highly comove, another approach is to exclusively use a small subset of economic factors that represent all factors. The conditions of the business cycle could be modelled immediately, conditional on the associated subset of covariate groups that proxy that particular condition. The economic upside of such an approach is that capital requirements do not change when the current condition of the business cycle is equal to the previous condition thereof. Additionally, Markov processes could be used to model the likelihood to jump or stay in conditions of the business cycle based on recent economic information. Therefore, changes in capital requirement could be predicted. This approach potentially solves the issue of multicollinearity since the subset of factors could be constructed such that collinearity among these factors is low.

Bibliography

Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6):716–723.

Ali, A., Kolter, J. Z., and Tibshirani, R. J. (2019). A continuous-time view of early stopping for least squares regression. In *The 22nd international conference on artificial intelligence and statistics*, pages 1370–1378. PMLR.

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

BCBS (2004). Basel ii: International convergence of capital measurement and capital standards: a revised framework. *Bank for International Settlements, Basel (June 2004)*.

Bondell, H. D. and Reich, B. J. (2009). Simultaneous factor selection and collapsing levels in anova. *Biometrics*, 65(1):169–177.

Chavez-Demoulin, V., Embrechts, P., and Hofert, M. (2016). An extreme value approach for modeling operational risk losses depending on covariates. *Journal of Risk and Insurance*, 83(3):735–776.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (2013). *Modelling extremal events for Insurance and Finance*, volume 33. Springer Science & Business Media.

Groll, A., Hambuckers, J., Kneib, T., and Umlauf, N. (2019). Lasso-type penalization in the framework of generalized additive models for location, scale and shape. *Computational Statistics & Data Analysis*, 140:59–73.

Hambuckers, J., Groll, A., and Kneib, T. (2018). Understanding the economic determinants of the severity of operational losses: A regularized generalized pareto regression approach. *Journal of Applied Econometrics*, 33(6):898–935.

Hastie, T., Montanari, A., Rosset, S., and Tibshirani, R. J. (2022). Surprises in high-dimensional ridgeless least squares interpolation. *The Annals of Statistics*, 50(2):949–986.

Hastie, T. J. and Tibshirani, R. J. (2017). *Generalized additive models*. Routledge.

Herawati, N., Nisa, K., Setiawan, E., Nusyirwan, N., and Tiryono, T. (2018). Regularized multiple regression methods to deal with severe multicollinearity. *International Journal of Statistics and Applications*, 8(4):167–172.

Mayr, A., Fenske, N., Hofner, B., Kneib, T., and Schmid, M. (2012). Generalized additive models for location, scale and shape for high dimensional data—a flexible approach based on boosting. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 61(3):403–427.

Oelker, M.-R. and Tutz, G. (2017). A uniform framework for the combination of penalties in generalized structured models. *Advances in Data Analysis and Classification*, 11(1):97–120.

Rigby, R. A. and Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54(3):507–554.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, pages 461–464.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288.

Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., and Knight, K. (2005). Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(1):91–108.

Tibshirani, R. J. (2015). A general framework for fast stagewise algorithms. *J. Mach. Learn. Res.*, 16(1):2543–2588.

Umlauf, N., Klein, N., Simon, T., and Zeileis, A. (2019). bamlss: A lego toolbox for flexible bayesian regression (and beyond). *arXiv preprint arXiv:1909.11784*.

Umlauf, N., Klein, N., and Zeileis, A. (2018). Bamlss: Bayesian additive models for location, scale, and shape (and beyond). *Journal of Computational and Graphical Statistics*, 27(3):612–627.

Valencia, F. (2016). Bank capital and uncertainty. *Journal of Banking & Finance*, 69:S1–S9.

Wood, S. and Wood, M. S. (2015). Package 'mgcv'. *R package version*, 1(29):729.

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(1):3–36.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101(476):1418–1429.

A. Appendix

A.1. Description of risk categories

BDSF - losses due to business disruption and system failures, including soft- and hardware issues.

CPBP - losses from unintentional failure to meet professional obligations, including contract, fiduciary and privacy breaches, unlicensed activities, model errors.

DPA - losses due to damaged physical assets, where damage could arise from both natural and human causes.

EDPM - losses due to failed transaction processing or process management, resulting from interaction with counterparties and vendors. These losses include incorrect information, miscommunication, data issues, and others.

EFRAUD - losses due to committed fraud by externals, including cyberfraud, falsely forged documents, theft.

EPWS - losses due to breaches of external employment, health or safety laws or internal discrimination/diversity events.

IFRAUD - losses due to committed fraud by internal employees, including unauthorized transactions, bribes, tax evasion, insider trading.

A.2. Description of economic factors

Firm-specific UniCredit factors

Deposit growth rate (DGR) - proportional variation of total deposits from quarter to quarter.

Leverage ratio (LR) - value of total assets divided by value of total liabilities.

Percentage of revenue coming from fees (PRF) - proportion of non-interest income.

Tier-I capital Ratio (TCR) - ratio between core capital and total assets of a bank. A higher TCR ensures a better liquidity position for the minimum capital requirement for extreme losses under Basel IV.

UniCredit stock returns (UCSR) - quarterly log-returns of UniCredit publicly tradeable stock price.

Macroeconomic factors

GDP growth rate (GDP EU / GDP IT) - compares the year-over-year (or quarterly) change of the European Union / Italian economic output to measure how fast an economy is growing, adjusted for seasonality.

Housing prices growth rate (HPI EU) - tracks movement of single-family house prices in the EU.

Consumption loan rate (LOR EU / LOR IT) - floating rate for European Union and Italian consumption loans.

Monetary Aggregate M1 growth rate (M1 EU) - the growth rate of the sum of currency in circulation and overnight deposits.

Unemployment rate (UR EU / UR IT) - percentage of the total labour force that is unemployed but actively seeking employment and willing to work in the European Union (UR Eu) and in Italy (UR IT).

Financial market factors

10-year Italian government bond yield (LIR IT) - long-term interest rate for Italy.

Milano Italia Borse returns (MIB IT) - quarterly log-returns of MIB index.

3-month Italian interbank rate (SIR IT) - short-term interest rate for Italy.

SP USA - quarterly log-returns of S&P 500 stock index.

Thomson Reuters European stock index log-returns (TRSI EU) - quarterly log-returns of TRSI index.

VIX USA - volatility index, calculated from a variety of S&P 500 options prices.

VFTSE UK - volatility index, based on options of the FTSE 100 index, based on stocks in the UK.

A.3. Correlation matrix of covariates

Table 16 shows the correlation among the seven risk categories and economic factors.

A.4. Categorical covariates, intervals limits and group sizes

We construct interval limits based on intervals. We set $-\infty$ and $+\infty$ as lower and upper bound since we do not know the out-of-sample covariate values. Observations are not equally divided between groups, although all groups capture a substantial number of observations to ensure consistent parameter estimates of all groups. The exact interval limits are shown in Table 17, and Tables 18 and 19 show the number of observations within the covariates before and after applying the POT method for our threshold choice.

Table 16: Pearson correlation coefficients among the risk categories and economic factors.

	IFRAUD	EFRAUD	EPWS	CPBP	BDSF	EDPM	DPA	UR EU	UR IT	GDP EU	GDP IT	HPI	M1	LOR EU	LOR IT	LIR	SIR	UCSR	TRSI	SP	VIX	VFTSE	MIB	PRF	DGR	TCR	LR
IFRAUD	1	-0.08	-0.04	-0.14	-0.02	-0.12	-0.03	-0.01	-0.01	0.01	0.01	0.04	0.01	0	0.02	-0.01	0.02	0	0	-0.01	-0.02	-0.02	0	0.03	0.02	-0.01	0.01
EFRAUD		1	-0.1	-0.35	-0.06	-0.3	-0.06	-0.01	-0.01	0.04	0.04	0.09	0.05	-0.01	0.02	-0.05	0.01	0.02	0.03	0.01	-0.06	-0.06	0.03	0.08	0.01	-0.03	0.03
EPWS			1	-0.2	-0.03	-0.17	-0.04	-0.02	-0.03	0.02	0.03	0.03	0.01	0.01	0.02	-0.01	0.02	-0.01	-0.01	0	-0.01	-0.01	-0.01	0.02	-0.01	-0.01	0.01
CPBP				1	-0.1	-0.56	-0.12	0	0.01	0	0	-0.05	-0.02	0.02	-0.01	0.02	0	0.01	0.01	0.02	0.02	0.02	0	-0.05	0.01	0.01	-0.02
BDSF					1	-0.09	-0.02	-0.03	-0.01	0	0	0.03	-0.01	0.03	0.03	-0.02	0.03	-0.01	0.01	0	-0.01	-0.01	0	0.03	0	-0.02	0.02
EDPM						1	-0.1	0.02	0.02	-0.04	-0.04	-0.04	-0.02	-0.01	-0.02	0.02	-0.03	-0.02	-0.03	-0.02	0.04	0.04	-0.02	-0.03	-0.02	0.02	-0.01
DPA							1	0.02	0.01	-0.02	-0.02	-0.03	-0.01	-0.02	-0.03	0.03	-0.03	-0.01	0	0.01	0.02	0.02	0	-0.03	-0.02	0.01	-0.02
UR EU								1	0.9	-0.1	-0.03	-0.47	0.08	-0.86	-0.84	0.02	-0.93	0.25	0.23	0.32	-0.12	-0.22	0.3	-0.46	-0.16	0.93	-0.67
UR IT									1	-0.16	-0.14	-0.45	-0.04	-0.73	-0.76	-0.02	-0.78	0.17	0.12	0.21	-0.15	-0.22	0.2	-0.35	-0.08	0.85	-0.6
GDP EU										1	0.94	0.65	0.39	-0.13	0.05	-0.25	0.24	0.38	0.45	0.43	-0.66	-0.63	0.4	0.45	-0.04	-0.09	0.21
GDP IT											1	0.61	0.48	-0.16	0.05	-0.31	0.13	0.41	0.48	0.44	-0.62	-0.63	0.42	0.39	-0.01	-0.05	0.09
HPI												1	0.37	0.18	0.49	-0.42	0.56	0.13	0.18	0.09	-0.61	-0.6	0.14	0.83	0.18	-0.52	0.54
M1													1	0.04	0.32	-0.74	-0.19	0.6	0.51	0.41	-0.42	-0.52	0.52	0.25	0.09	-0.07	0.07
LOR EU														1	0.9	-0.19	0.77	-0.12	-0.21	-0.31	0.25	0.33	-0.2	0.25	0.16	-0.83	0.56
LOR IT															1	-0.4	0.73	0.01	-0.07	-0.21	0.02	0.06	-0.07	0.46	0.18	-0.87	0.62
LIR																1	0.02	-0.58	-0.39	-0.27	0.43	0.51	-0.5	-0.36	-0.08	0.13	-0.09
SIR																	1	-0.24	-0.19	-0.27	-0.05	0.07	-0.28	0.54	0.19	-0.84	0.71
UCSR																		1	0.85	0.72	-0.59	-0.63	0.92	-0.1	0.01	0.06	-0.04
TRSI																			1	0.91	-0.68	-0.7	0.93	-0.08	-0.06	0.08	-0.02
SP																				1	-0.64	-0.65	0.86	-0.19	-0.02	0.21	-0.08
VIX																					1	0.97	-0.62	-0.41	-0.08	-0.02	-0.18
VFTSE																						1	-0.66	-0.39	-0.06	-0.1	-0.12
MIB																							1	-0.11	-0.05	0.13	-0.07
PRF																								1	0.22	-0.44	0.44
DGR																									1	-0.19	-0.02
TCR																										1	-0.72
LR																											1

Table 17: Interval limits of dummy encoded categorical covariates.

Data type	Factor	Impact	Groups	Minimum	Maximum	Intervals
Firm-specific	RC	NA	7	NA	NA	NA
	DGR	+	6	-0.32	0.44	-0.2, 0, 0.02, 0.1, 0.25
	LR	+	5	14.0	30.8	15, 18, 21, 24
	PRF	+	5	0.24	0.42	0.27, 0.31, 0.35, 0.39
	TCR	+	5	0.05	0.11	0.06, 0.07, 0.08, 0.1
	UCSR	+	6	-0.26	0.21	-0.23, -0.14, -0.05, 0.05, 0.14
Macroeconomic	GDP EU	+	6	-3.00	1.10	-0.5, 0, 0.5, 0.75, 1
	GDP IT	+	5	-2.90	1.00	-0.5, 0, 0.5, 0.75
	HPI EU	+	6	-4.45	7.52	-2, 0, 3, 5, 7.3
	LOR EU	-	5	5.09	8.75	5.3, 6.5, 7.5, 8.5
	LOR IT	-	5	5.42	11.2	5.9, 8.5, 10, 10.8
	M1 EU	+	5	0.60	12.6	3, 6, 9, 12
	UR EU	-	6	7.00	12.0	7.5, 8.5, 9.5, 10.5, 11.5
	UR IT	-	6	6.00	14.0	6, 7, 8, 10, 12
Financial market	LIR IT	-	5	3.09	6.61	4, 4.3, 4.6, 5
	MIB IT	+	5	-0.27	0.23	-0.17, -0.6, 0.06, 0.17
	SIR IT	-	6	0.20	4.98	0.5, 1, 2, 3, 4.5
	SP USA	+	6	-0.26	0.14	-0.1, -0.02, 0.02, 0.05, 0.08
	TRSI EU	+	5	-0.27	0.22	-0.04, 0, 0.04, 0.10
	VIX USA	-	6	11.4	44.1	11.9, 13, 15, 18, 25
	VFTSE UK	-	6	10.1	39.7	11, 13, 16, 18, 2

Note: This table includes the number of groups, minimum and maximum covariate values, interval limits, whereby we exclude the lower- and upper bound, which are fixed to $-\infty$ and $+\infty$ for all covariates. The term "RC" is an abbreviation for the risk categories. Elements of the column "Impact" are + (-) when high values of the corresponding economic factor indicate an expansion (recession), and "NA" means not applicable.

Table 18: Number of observations within dummy encoded economic categorized factors before applying the POT method.

Data type	Factor	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
Firm-specific	DGR	4,035	11,920	8,726	8,045	3,887	4,258
	LR	8,506	16,446	6,713	6,229	2,977	NA
	PRF	1,034	9,689	15,850	10,746	3,552	NA
	TCR	12,720	6,144	5,630	11,306	5,071	NA
	UCSR	1,134	3,716	7,246	20,910	5,555	2,310
Macroeconomic	GDP EU	3,716	9,223	13,799	8,902	3,127	2,104
	GDP IT	8,132	13,560	11,032	3,878	4,269	NA
	HPI EU	8,074	7,912	8,824	5,521	6,186	4,354
	LOR EU	5,569	8,715	10,622	12,292	3,673	NA
	LOR IT	6,604	7,680	5,441	14,507	6,639	NA
	M1 EU	6,141	9,275	13,708	9,491	2,256	NA
	UR EU	5,697	8,760	8,761	9,187	3,395	5,071
	UR IT	3,020	9,053	12,417	9,776	3,632	2,973
Financial market	LIR IT	7,637	12,558	8,380	8,045	4,251	NA
	MIB IT	4,666	5,005	18,534	10,356	2,310	NA
	SIR IT	6,880	6,119	6,836	7,661	7,948	5,427
	SP USA	5,549	5,529	8,388	7,719	5,356	8,330
	TRSI EU	8,841	4,071	11,092	8,248	8,619	NA
	VIX USA	2,375	6,211	5,366	11,610	5,964	9,345
	VFTSE UK	4,354	3,809	8,219	4,749	11,671	8,069

Note: "NA" means not applicable for factors that do not have 6 groups. The rows sum to 40,871, which is the total number of observations before applying the POT for our threshold (Section 3.2).

Table 19: Number of observations within dummy encoded economic categorized factors after applying the POT method.

Data type	Factor	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
Firm-specific	DGR	1,014	2,822	2,119	2,258	933	1,073
	LR	1,963	4,131	1,649	1,700	776	NA
	PRF	2,59	2,410	3,845	2,707	998	NA
	TCR	3,085	1,666	1,626	2,754	1,088	NA
	UCSR	289	861	1,648	5,473	1,311	637
Macroeconomic	GDP EU	861	2,128	3,405	2,505	812	508
	GDP IT	1,805	3,331	2,952	1,086	1,045	NA
	HPI EU	1,990	1,782	2,129	1,388	1,682	1,248
	LOR EU	1,299	2,022	2,837	3,227	834	NA
	LOR IT	1,587	1,734	1,420	3,808	1,670	NA
	M1 EU	1,377	2,249	3,424	2,545	624	NA
	UR EU	1,346	2,277	2,376	2,329	803	1,088
	UR IT	794	2,291	3,225	2,479	797	633
Financial market	LIR IT	2,036	3,291	1,937	2,016	939	NA
	MIB IT	1,116	1,128	4,712	2,626	637	NA
	SIR IT	1,561	1,556	1,651	2,067	2,029	1,355
	SP USA	1,329	1,286	2,056	2,103	1,407	2,038
	TRSI EU	2,081	942	2,705	2,299	2,192	NA
	VIX USA	597	1,786	1,361	2,834	1,342	2,299
	VFTSE UK	1,248	924	2,230	1,133	2,763	1,921

Note: "NA" means not applicable for factors that do not have 6 groups. The rows sum to 10,219, which is the total number of observations after applying the POT for our threshold (Section 3.2).

A.5. Economic interpretation of fused L_1-norms, and a downside of using adaptive weights

For illustration, we use the economic metric factor GDP EU, which reflects the GDP growth rate in the EU. We transform this factor, based on their levels, into an ordered categorical covariate J, which we dummy encoded into 6 groups $df = 0, 1, ..., 5$. In is likely that not all (or none) covariate groups explain much variation in the LSEs. In that case, the associated coefficients are shrunk to/towards zero by the (in)direct link with the reference category. If one or more groups do explain much variation in the model, then fused L_1-norms can successfully filter these groups as follows: most coefficients are shrunk to zero, hence fused L_1-norms encourage 'relevant' groups to shrink to zero as well. However, relevant groups do not 'give in', and 'stand their ground'. In this particular case, since expansions drive extreme OpRisks, we expect the upper-bound group of GDP EU to be significantly higher than zero.

We subsequently illustrate why penalized estimation procedures likely yield zero coefficients when adaptive weights are zero. Assume that we have a AFUSE$_m$ penalty function for a pair $\{j, k\}$, combined with a corresponding close to zero adaptive weight $\hat{b}_{j,k}^{\text{ini}} = 10^{-8}$, as given by Equation (10). Therefore, we have $P_{\text{alas}}^{j,k}(b_{j,k}) = \frac{|b_{j,k}|}{10^{-8}}$. If $b_{j,k}$ is not close to zero after an iteration during model estimation, then $P_{\text{alas}}^{j,k}(b_{j,k})$ becomes huge. We subtract the product of tuning parameters and penalty functions from the original loss severity log-likelihood function. Hence, the log-likelihood, which we maximize, consequently shrinks drastically. Therefore, in practice, regularization techniques do not update model coefficients to substantial non-zero values if the corresponding adaptive weight is close to zero. All in all, boosting optimizers implicitly perform variable selection when there still are zero-coefficients after model estimation, although we want our regularization methods to select driving covariates and not the boosting optimizers. This interpretation can be extended to other adaptive weight specifications.

A.6. *Finding comparable true linear predictors for different covariate sets in GAMLSSs*

The 'true losses' in a simulation study are the result of random draws from the GPD, and these draws depend on true linear predictors, which are constructed by true distributional parameters. To investigate whether data transformations (metric to categorical, or vice versa) increase modelling performance, we must model the same vector of losses. Unfortunately, the losses are based on random draws. Now, let us assume that we can find a reasonable way to obtain deterministic draws from the probability distribution. For example, instead of a random draw from the probability distribution, one might consider taking a deterministic draw from predetermined quantiles of the probability distribution.

We denote by M1 and M2 two models. In a GAMLSS framework, we have $(n \times 1)$ vectors $\boldsymbol{y}^{\mathrm{M1}} \sim GPD\Big(\boldsymbol{\xi}^{\mathrm{M1}} = \exp(\boldsymbol{\eta}_1^{\mathrm{M1}}), \boldsymbol{\sigma}^{\mathrm{M1}} = \exp(\boldsymbol{\eta}_2^{\mathrm{M1}})\Big)$ and a $\boldsymbol{y}^{\mathrm{M2}} \sim GPD\Big(\boldsymbol{\xi}^{\mathrm{M2}} = \exp(\boldsymbol{\eta}_1^{\mathrm{M2}}), \boldsymbol{\sigma}^{\mathrm{M2}} = \exp(\boldsymbol{\eta}_2^{\mathrm{M2}})\Big)$. In our simulation, we predetermine the true distributional parameters. Then, we use the product of these parameters and simulated covariates to obtain true linear predictors. Consequently, even if the random draw from a probability distribution can be fixed in some way, we still have to solve $\boldsymbol{y}^{\mathrm{M1}} = \boldsymbol{y}^{\mathrm{M2}}$. Equivalently, we solve $\boldsymbol{\eta}_1^{\mathrm{M1}} = \boldsymbol{\eta}_1^{\mathrm{M2}}$ and $\boldsymbol{\eta}_2^{\mathrm{M1}} = \boldsymbol{\eta}_2^{\mathrm{M2}}$. We define $j^{\mathrm{M1}} = 1, .., p^{\mathrm{M1}}$ and $j^{\mathrm{M2}} = 1, .., p^{\mathrm{M2}}$ as the number of metric covariates, and $J^{\mathrm{M1}} = 1, ..., C^{\mathrm{M1}}$ and $J^{\mathrm{M2}} = 1, ..., C^{\mathrm{M2}}$ as the number of categorical covariates for models M1 and M2. The categorical covariates J^{M1} or J^{M2} contain groups of size df_J^{M1} and df_J^{M2}. We have the subsequent set of equations:

$$\begin{aligned}
\boldsymbol{\eta}_1^{\mathrm{M1}} &= \xi_0^{\mathrm{M1}} + \sum_{j^{\mathrm{M1}}=1}^{p^{\mathrm{M1}}} \boldsymbol{x}_j^{\mathrm{M1}} \xi_j^{\mathrm{M1}} + \sum_{J^{\mathrm{M1}}=1}^{C^{\mathrm{M1}}} \boldsymbol{X}_j^{\mathrm{M1}} \boldsymbol{\xi}_J^{M1^\top} \qquad (33)\\
&= \xi_0^{\mathrm{M2}} + \sum_{j^{\mathrm{M2}}=1}^{p^{\mathrm{M2}}} \boldsymbol{x}_j^{\mathrm{M2}} \xi_j^{\mathrm{M2}} + \sum_{J^{\mathrm{M2}}=1}^{C^{\mathrm{M2}}} \boldsymbol{X}_j^{\mathrm{M2}} \boldsymbol{\xi}_J^{M2^\top} = \boldsymbol{\eta}_1^{\mathrm{M2}},\\
\boldsymbol{\eta}_2^{\mathrm{M1}} &= \sigma_0^{\mathrm{M1}} + \sum_{j^{\mathrm{M1}}=1}^{p^{\mathrm{M1}}} \boldsymbol{x}_j^{\mathrm{M1}} \sigma_j^{\mathrm{M1}} + \sum_{J^{\mathrm{M1}}=1}^{C^{\mathrm{M1}}} \boldsymbol{X}_j^{\mathrm{M2}} \boldsymbol{\sigma}_J^{M1^\top} \qquad (34)\\
&= \sigma_0^{\mathrm{M2}} + \sum_{j^{\mathrm{M2}}=1}^{p^{\mathrm{M2}}} \boldsymbol{x}_j^{\mathrm{M2}} \sigma_j^{\mathrm{M2}} + \sum_{J^{\mathrm{M2}}=1}^{C^{\mathrm{M2}}} \boldsymbol{X}_j^{\mathrm{M2}} \boldsymbol{\sigma}_{J,2}^{M2^\top} = \boldsymbol{\eta}_2^{\mathrm{M2}},
\end{aligned}$$

where $\boldsymbol{x}_j^{(\cdot)}$ is a $(n \times 1)$ vector for all j, and $\boldsymbol{X}_j^{(\cdot)}$ is a $(n \times df_J^{(\cdot)})$ matrix for all J. We have scalars $\xi_0^{(\cdot)}$ and $\sigma_0^{(\cdot)}$, and $(1 \times df_J^{(\cdot)})$ coefficient vectors $\boldsymbol{\xi}_J^{(\cdot)}$ and $\boldsymbol{\sigma}_J^{(\cdot)}$.

Why is this relevant? If we can solve the set of equations above, then we can evaluate whether data transformations in a GAMLSS framework improve modelling performance in a simulation study. Then, we can compare different covariate sets or identical covariate sets consisting of different numbers of covariates and/or number of groups within these covariates.

A.7. Simulation study, limitations of the MVND when simulating comoving covariates

We use the MVND to simulate comoving covariates. In this procedure, we start by performing a Cholesky decomposition on the correlation matrix $\boldsymbol{P}$. In order to perform such a decomposition, a necessary condition is that $\boldsymbol{P}$ needs to be positive definite (PD). However, for high correlations ρ and/or high numbers of covariates C^{sim}, $\boldsymbol{P}$ frequently is not PD. In our simulation study, we work with many covariates $C^{\text{sim}} = \{8, 15\}$. In our application, $\boldsymbol{P}$ is usually only PD for $\rho = 0$ and not for $\rho = \{.3, .6\}$. To attain the nearest PD matrix $\boldsymbol{P}^{\text{PD}}$, we use the *nearPD* function of the *Matrix* R-package.[4] However, much aimed correlation is lost when using this function. Therefore, we are not able to simulate extremely high average levels of cross-correlations among simulated categorical covariates in our simulation study.

[4] https://cran.r-project.org/web/packages/Matrix/index.html

A.8. Simulation study, number of GPD parameter updates

Tables 20 and 21 compare the updates schemes of B1 and B2 over a grid of n^{iter}.

Table 20: Mean, standard deviation and ratio r^{iter} of the number of GPD parameter updates over ten samples, where the standard deviation is in parentheses, and $n^{\text{iter}} = [1; 10,000]$.

			n^{iter}	100			500			1,000			5,000			10,000		
Optimizer	ρ	C^{sim}	ν	ξ	σ	r^{iter}	ξ	σ	r^{iter}	ξ	σ	r^{iter}	ξ	σ	r^{iter}	ξ	σ	r^{iter}
gamboostLSS	0	8	.01	0	100	0	0	500	0	86	914	0.09	2,010	2,990	0.4	5,025	4,975	0.5
				(0)	(0)	-	(0)	(0)	-	(43)	(43)	-	(318)	(318)	-	(508)	(508)	-
	0	8	.05	0	100	0	147	353	0.29	404	596	0.4	3,202	1,798	0.64	7,143	2,857	0.71
				(0)	(0)	-	(27)	(27)	-	(63)	(63)	-	(166)	(166)	-	(273)	(273)	-
	0	8	.1	9	91	0.09	203	297	0.41	508	492	0.51	3,584	1,416	0.72	7,621	2,379	0.76
				(4)	(4)	-	(31)	(31)	-	(50)	(50)	-	(135)	(135)	-	(168)	(168)	-
	0	15	.01	0	100	0	0	500	0	53	947	0.05	1,562	3,438	0.31	4,363	5,637	0.44
				(0)	(0)	-	(0)	(0)	-	(40)	(40)	-	(232)	(232)	-	(376)	(376)	-
	0	15	.05	0	100	0	109	391	0.22	314	686	0.31	2,971	2,029	0.59	6,789	3,211	0.68
				(0)	(0)	-	(21)	(21)	-	(46)	(46)	-	(158)	(158)	-	(291)	(291)	-
	0	15	.1	6	94	0.06	158	342	0.32	441	559	0.44	3,408	1,592	0.68	7,451	2,549	0.75
				(4)	(4)	-	(23)	(23)	-	(37)	(37)	-	(144)	(144)	-	(224)	(224)	-
	.3	8	.01	0	100	0	6	494	0.01	112	888	0.11	1,997	3,003	0.4	4,961	5,039	0.5
				(0)	(0)	-	(18)	(18)	-	(73)	(73)	-	(241)	(241)	-	(367)	(367)	-
	.3	8	.05	1	99	0.01	157	343	0.31	402	598	0.4	3,177	1,823	0.64	7,045	2,955	0.7
				(4)	(4)	-	(28)	(28)	-	(48)	(48)	-	(181)	(181)	-	(276)	(276)	-
	.3	8	.1	12	88	0.12	201	299	0.4	502	498	0.5	3,535	1,465	0.71	7,499	2,501	0.75
				(7)	(7)	-	(24)	(24)	-	(37)	(37)	-	(137)	(137)	-	(246)	(246)	-
	.3	15	.01	0	100	0	0	500	0	96	904	0.1	1,894	3,106	0.38	4,698	5,302	0.47
				(0)	(0)	-	(0)	(0)	-	(44)	(44)	-	(194)	(194)	-	(295)	(295)	-
	.3	15	.05	0	100	0	144	356	0.29	380	620	0.38	3,031	1,969	0.61	6,822	3,178	0.68
				(0)	(0)	-	(18)	(18)	-	(38)	(38)	-	(96)	(96)	-	(145)	(145)	-
	.3	15	.1	10	90	0.1	191	309	0.38	475	525	0.47	3,425	1,575	0.69	7,391	2,609	0.74
				(4)	(4)	-	(19)	(19)	-	(29)	(29)	-	(71)	(71)	-	(138)	(138)	-
	.6	8	.01	0	100	0	0	500	0	78	922	0.08	1,994	3,006	0.4	4,849	5,151	0.48
				(0)	(0)	-	(0)	(0)	-	(49)	(49)	-	(260)	(260)	-	(366)	(366)	-
	.6	8	.05	0	100	0	152	348	0.3	401	599	0.4	3,076	1,924	0.62	6,895	3,105	0.69
				(0)	(0)	-	(29)	(29)	-	(52)	(52)	-	(118)	(118)	-	(266)	(266)	-
	.6	8	.1	8	92	0.08	202	298	0.4	490	510	0.49	3,462	1,538	0.69	7,511	2,489	0.75
				(5)	(5)	-	(26)	(26)	-	(36)	(36)	-	(131)	(131)	-	(233)	(233)	-
	.6	15	.01	0	100	0	2	498	0	157	843	0.16	2,076	2,924	0.42	4,925	5,075	0.49
				(0)	(0)	-	(5)	(5)	-	(48)	(48)	-	(178)	(178)	-	(286)	(286)	-
	.6	15	.05	0	100	0	170	330	0.34	417	583	0.42	3,053	1,947	0.61	6,772	3,228	0.68
				(1)	(1)	-	(19)	(19)	-	(36)	(36)	-	(68)	(68)	-	(148)	(148)	-
	.6	15	.1	16	84	0.16	209	291	0.42	497	503	0.5	3,400	1,600	0.68	7,223	2,777	0.72
				(5)	(5)	-	(18)	(18)	-	(28)	(28)	-	(74)	(74)	-	(180)	(180)	-
GAMLSSBoostAlt	[-1,1]	> 1	$(0,\infty)$	50	50	.5	250	250	.5	500	500	.5	2,500	2,500	.5	5,000	5,000	.5
				(0)	(0)	-	(0)	(0)	-	(0)	(0)	-	(0)	(0)	-	(0)	(0)	-

Note: C^{sim}, ρ, and ν are the number of simulated categorical covariates, the average absolute correlation among covariates and the step size of the gradient descent optimizers, respectively. The standard deviations of both GPD parameters are equal because n^{iter} is fixed as the sum of the number of ξ and σ updates. Therefore, the number of σ updates is equal to the number of total iterations minus the number of ξ updates, and the standard deviation is not effected by a constant and a change of sign.

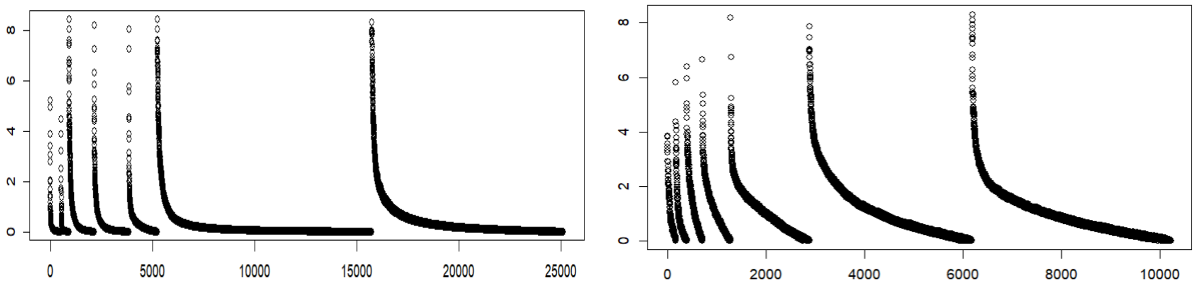
Table 21: Mean, standard deviation and ratio r^{iter} of the number of GPD parameter updates over ten samples, where the standard deviation is in parentheses, and $n^{\text{iter}} = [25,000; 100,000]$.

			n^{iter}	25,000			50,000			75,000			100,000		
Optimizer	ρ	C^{sim}	ν	ξ	σ	r^{iter}	ξ	σ	r^{iter}	ξ	σ	r^{iter}	ξ	σ	r^{iter}
gamboostLSS	0	8	.01	15,950	9,050	0.64	35,616	14,384	0.71	55,702	19,298	0.74	75,832	24,168	0.76
				(835)	(835)	-	(1,372)	(1,372)	-	(1,704)	(1,704)	-	(1,732)	(1,732)	-
	0	8	.05	19,212	5,788	0.77	-	-	-	-	-	-	-	-	-
				(345)	(345)	-	-	-	-	-	-	-	-	-	-
	0	8	.1	19,662	5,338	0.79	-	-	-	-	-	-	-	-	-
				(361)	(361)	-	-	-	-	-	-	-	-	-	-
	0	15	.01	14,792	10,208	0.59	33,838	16,162	0.68	53,726	21,274	0.72	74,078	25,922	0.74
				(795)	(795)	-	(1,462)	(1,462)	-	(1,957)	(1,957)	-	(2,281)	(2,281)	-
	0	15	.05	18,983	6,017	0.76	-	-	-	-	-	-	-	-	-
				(495)	(495)	-	-	-	-	-	-	-	-	-	-
	0	15	.1	19,952	5,048	0.8	-	-	-	-	-	-	-	-	-
				(429)	(429)	-	-	-	-	-	-	-	-	-	-
	.3	8	.01	15,827	9,173	0.63	35,123	14,877	0.7	-	-	-	-	-	-
				(906)	(906)	-	(1,396)	(1,396)	-	-	-	-	-	-	-
	.3	8	.05	18,931	6,069	0.76	-	-	-	-	-	-	-	-	-
				(646)	(646)	-	-	-	-	-	-	-	-	-	-
	.3	8	.1	19,427	5,573	0.78	-	-	-	-	-	-	-	-	-
				(700)	(700)	-	-	-	-	-	-	-	-	-	-
	.3	15	.01	15,095	9,905	0.6	34,001	15,999	0.68	53,549	21,451	0.71	73,475	26,525	0.73
				(483)	(483)	-	(738)	(738)	-	(1,025)	(1,025)	-	(1,400)	(1,400)	-
	.3	15	.05	18,782	6,218	0.75	-	-	-	-	-	-	-	-	-
				(373)	(373)	-	-	-	-	-	-	-	-	-	-
	.3	15	.1	19,675	5,325	0.79	-	-	-	-	-	-	-	-	-
				(423)	(423)	-	-	-	-	-	-	-	-	-	-
	.6	8	.01	15,318	9,682	0.61	34,358	15,642	0.69	-	-	-	-	-	-
				(593)	(593)	-	(1,345)	(1,345)	-	-	-	-	-	-	-
	.6	8	.05	19,077	5,923	0.76	-	-	-	-	-	-	-	-	-
				(473)	(473)	-	-	-	-	-	-	-	-	-	-
	.6	8	.1	19,687	5,313	0.79	-	-	-	-	-	-	-	-	-
				(281)	(281)	-	-	-	-	-	-	-	-	-	-
	.6	15	.01	15,205	9,795	0.61	33,758	16,242	0.68	52,570	22,430	0.7	71,767	28,233	0.72
				(345)	(345)	-	(744)	(744)	-	(1,262)	(1,262)	-	(1,818)	(1,818)	-
	.6	15	.05	18,318	6,682	0.73	-	-	-	-	-	-	-	-	-
				(483)	(483)	-	-	-	-	-	-	-	-	-	-
	.6	15	.1	19,299	5,701	0.77	-	-	-	-	-	-	-	-	-
				(483)	(483)	-	-	-	-	-	-	-	-	-	-
GAMLSSBoostAlt	[-1,1]	> 1	$(0,\infty)$	12,500	12,500	.5	25,000	25,000	.5	37,500	37,500	.5	50,000	50,000	.5
				(0)	(0)	-	(0)	(0)	-	(0)	(0)	-	(0)	(0)	-

Note: Identical to Table 20.

A.9. Reflection upon the threshold analysis of Groll et al. (2019)

Groll et al. (2019) filter 25,134 LSEs based a normality test of resulting quantile residuals.[5] Figure 9 shows their and our log-LSEs after our respective threshold analyses. We observe that their log-LSEs are smoother and hence the lower-tail of their LSEs likely better fits a scaled GPD compared to our own log-LSEs. However, by including many more losses, the fit of the upper-tail likely worsen. This argument is in line with their own conclusion, which is that 2-98% quantiles of the LSEs are adequately modelled but not the upper 98-100% quantile. Moreover, it is unknown how their analysis performs out-of-sample since they do not present out-of-sample evaluation measures. On the other side, an advantage of their framework is that they include more losses for all risk categories except EFRAUD, as shown in Table 22. Therefore, the risk categories that do not contain many losses are better represented.



(a) Log-LSEs; Groll et al. (2019) (b) Log-LSEs; our choice for threshold (Section 3.2)

Figure 9: Comparison of log-LSEs. From left to right: BDSF, DPA, IFRAUD, EPWS, EFRAUD, EDPM, and CPBP.

Table 22: Comparison of number of losses for different thresholds.

	Threshold		
	None	Groll et al. (2019)	Ours
IFRAUD	1,271	1,233	318
EFRAUD	6,391	1,392	1,598
EPWS	2,292	1,706	5,73
CPBP	16,138	9,378	4,035
BDSF	674	542	169
EDPM	13,209	10,499	3,302
DPA	896	384	224
All	40,871	25,134	10,219

[5]Source: https://doi.org/10.1016/j.csda.2019.06.005

A.10. Real-world application, loss severity exceedances

Table 23: Estimated LSE parameters using models M0, M1-met and M2-met. Model M0 is a model in which the LSEs are conditional on the model intercepts.

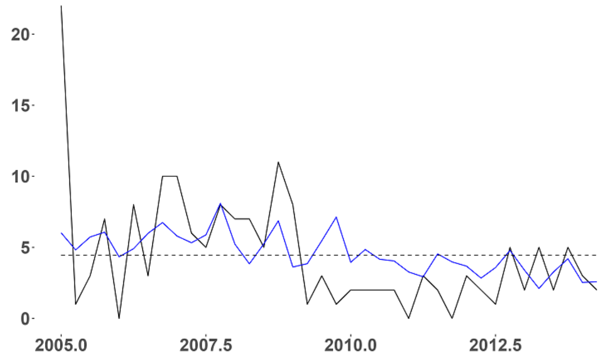
	M0		M1-met		M2-met	
	ξ	σ	ξ	σ	ξ	σ
Intercept	-0.343	3.822	-0.516	3.862	-0.545	3.87
IFRAUD	-	-	-0.162	0.585	-0.159	0.578
EFRAUD	-	-	-0.359	0.61	-0.353	0.601
EPWS	-	-	-0.162	0.543	-0.167	0.539
CPBP	-	-	-0.377	1.148	-0.378	1.141
BDSF	-	-	-0.036	0.164	-0.033	0.155
EDPM	-	-	-0.11	0.741	-0.112	0.738
DGR	-	-	-	-	0.334	0.488
LR	-	-	-	-	-0.204	-0.223
PRF	-	-	-	-	0.012	0.061
TCR	-	-	-	-	0.014	-0.037
UCSR	-	-	-	-	0.09	-0.04
GDP EU			-	-	-0.007	-0.11
GDP IT	-	-	-	-	0.337	0.021
HPI EU	-	-	-	-	-0.326	0.184
LOR EU	-	-	-	-	0.017	-0.027
LOR IT			-	-	0.028	0.199
M1 EU	-	-	-	-	-0.012	-0.055
UR EU	-	-	-	-	0.074	-0.02
UR IT	-	-	-	-	-0.066	0.048
LIR IT	-	-	-	-	0.137	-0.003
MIB IT	-	-	-	-	-0.19	-0.002
SIR IT	-	-	-	-	-0.077	0.059
SP USA	-	-	-	-	-0.015	0.075
TRSI EU	-	-	-	-	0.027	-0.019
VIX USA	-	-	-	-	-0.11	0.035
VFTSE UK	-	-	-	-	-0.095	-0.055

A.11. Real-world application, loss frequency exceedances

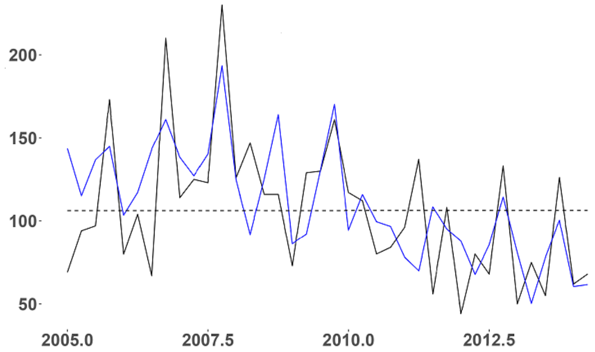
We investigate time-varying LFE-models. Thereby, we use model M3 to train in-sample on all but one time period, where the remaining time period serves as validation set. To predict the number of losses in the validation set, we use the in-sample coefficients and the economic factors in the validation set. Figure 10 shows the time-varying NHPP distributed GAM parameters, whereby the horizontal lines show the average number of losses over all 38 time periods. We thereby observe a downward trend for all risk categories. The most extreme upward peaks are in 2006, just before the 2007-2010 financial crisis. Furthermore, CPBP and EDPM exhibit the highest number of losses, and the associated parameter estimates are quite close to the true parameters over time periods in which the volatility of the LFEs is high for CPBP and EDPM. However, the peaks belonging to the other risk categories are frequently difficult to estimate, especially for risk categories with a low number of losses. For example, the peak of the sixth subfigure belongs to fourth quarter in 2006 of EPWS, whereby we estimate 22.87 losses, whereas 71 losses in fact occurred. The reason is that we use a logarithmic link function to the LFE predictor and hence the parameter estimates are transformed back by an exponential transformation to these estimates. For illustration, a one-level shift in parameter estimates from 3 to 4 has a much larger impact than a change from 2 to 3 since $(e^4 - e^3) >> (e^3 - e^2)$.

Tables 24 - 27 present the out-of-sample log-likelihood scores of the validation set, and demonstrate huge differences in log-likelihoods scores, both in the cross-section of risk categories and over time. The inclusion of a few economic factors (model M3) substantially improves OOS-LLH scores for EFRAUD since the aggregated OOS-LLH score for EFRAUD is 364.6 for model M1 and 264.0 for model M3. Furthermore, BDSF is barely affected by the in- or exclusion of economic factors since OOS-LLHs are within a small range of 95.2 - 100.0 for all four models. With regard to the time-varying OOS-LLH scores, we observe that model M3 is superior in time periods in which the number of losses unexpected in- or decrease. The OOS-LLH score, aggregated over all risk categories, in 2006 Q4 is 204.9, 223.5, 96.2, 120.8 for models M1 - M4, respectively. Lastly, we are surprised that most firm-specific factors are unidentified. The number of losses ought to be partially driven by firm-specific factors because latent factors, such as internal compliance, bonus structure, general level of risk-aversion of employees and the organization's culture, are probable to affect the likelihood of losses occurring. Perhaps these factors could be explained by economic factors. Another theory is that both economic and latent firm-specific factors have the same underlying, which is the condition of the business cycle.

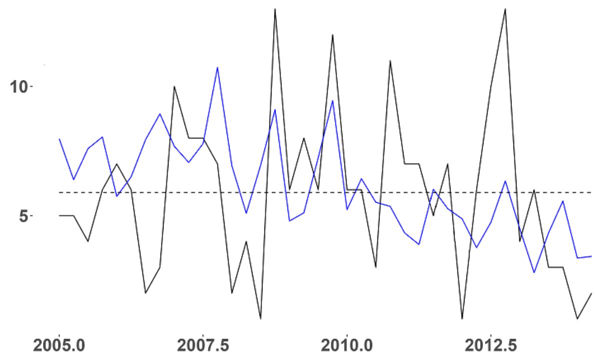
Additionally, we find that the sensitivity of the economic factors with respect to risk cate-



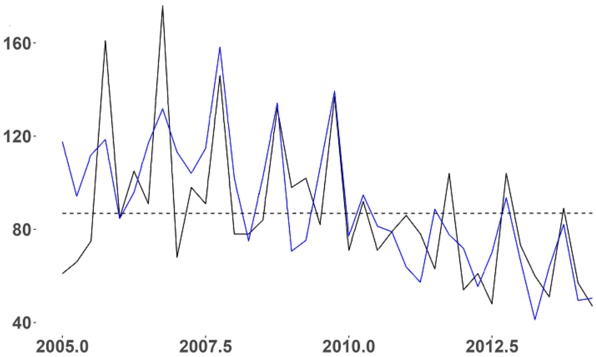
(a) BDSF



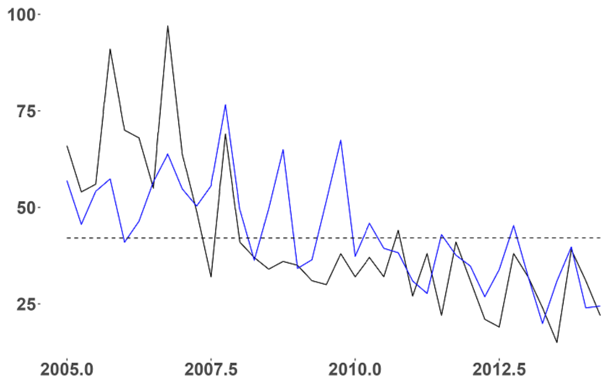
(b) CPBP



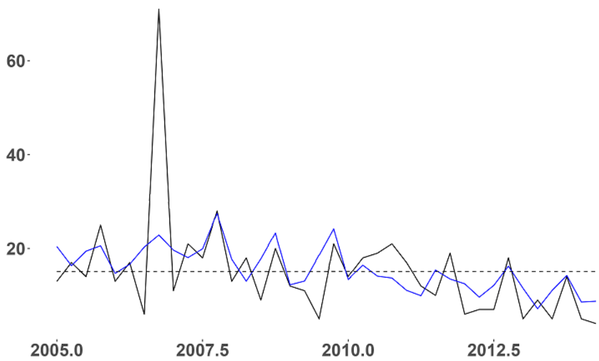
(c) DPA



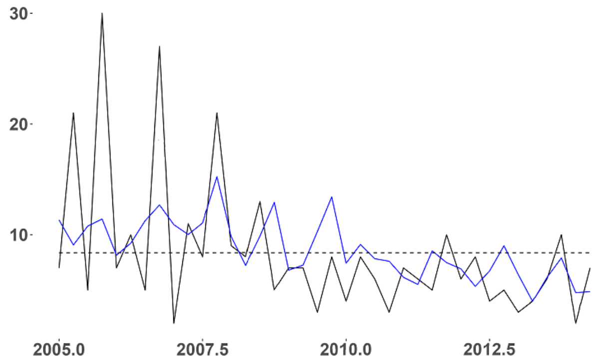
(d) EDPM



(e) EFRAUD



(f) EPWS



(g) IFRAUD

Figure 10: Number of losses (y-axis) over the 38 quarterly time periods (x-axis). The estimated (black) and true (blue) parameters are solid. Estimated parameters are obtained using model M3. The long-term average number of losses is the horizontal dotted black line.

gories differs. The number of CPBP and EDPM losses are highly sensitive to economic conditions since estimated LFEs substantially change between time periods. This change is frequently positive, which means that the estimates LFEs moves towards the true LFEs for a given time period. Risk categories with fewer losses, such as BDSF and DPA are not able to anticipate as abruptly. All in all, we conclude that our LFE models are able to adequately predict the number of losses for risk categories that have sufficient losses. All risk-category-specific estimated and true losses are in Table 28.

Table 24: OOS-LLH scores of the LFEs in the validation set using model M1. The scores are calculated for individual and aggregated risk categories for individual and aggregated time periods, whereby "NA" values mean that the particular risk category has no observation in that given time period.

Time period	EFRAUD	IFRAUD	EPWS	CPBP	EDPM	DPA	BDSF	All
2005 Q1	10.8	7.5	2.4	2	9.2	20.6	1.8	54.3
2005 Q2	4	1.8	9.7	4.6	3.3	2.5	5.9	31.7
2005 Q3	3.6	2.6	4	5.2	2.3	1.9	2	21.5
2005 Q4	22.3	30.4	20.9	5.4	26.1	2.4	1.8	109.2
2006 Q1	11.3	2	2.4	6.8	3.2	2	NA	27.7
2006 Q2	3.3	5.1	10.2	2.5	2.2	2.9	1.8	28
2006 Q3	3.3	11.7	2.6	4.8	5.5	1.9	3.1	33
2006 Q4	63.1	45.8	16.7	41.2	2.4	31.3	4.4	204.9
2007 Q1	3.6	5.4	8.2	2.8	4.9	4.4	3.3	32.6
2007 Q2	3.9	2.5	5	3.4	3.5	2	2.3	22.7
2007 Q3	4.7	2	3.3	2.7	4	2.3	1.7	20.7
2007 Q4	21.2	9.7	61.6	10.7	7.3	2.9	2	115.4
2008 Q1	5.2	2.1	3.6	2.4	2.8	2.4	3.1	21.5
2008 Q2	10.8	3.6	2	3.1	2.7	2.4	2	26.5
2008 Q3	3.2	3.4	3.6	3.8	3.5	1.7	4.2	23.4
2008 Q4	14.1	2.6	3.8	3.2	3.2	5.3	5.6	37.7
2009 Q1	3.9	9.2	2	2.5	2.9	3.4	1.8	25.8
2009 Q2	4.3	5.8	4.5	2	3.3	2.8	2.3	25.1
2009 Q3	3.9	6	3.3	4.6	6.5	1.8	1.9	28
2009 Q4	16.4	16.4	2	3.5	4.8	3	3.3	49.5
2010 Q1	4.7	3.9	3.1	2.3	4	1.8	2.4	22.2
2010 Q2	3.3	3.4	2	3.1	2.7	2.4	1.8	18.7
2010 Q3	4.7	6.8	4	2.2	2.9	2.4	2.4	25.5
2010 Q4	5.8	3.5	3.5	3.9	2.9	4	2.4	26
2011 Q1	3.7	3.2	2	2.5	5.8	2	NA	19.2
2011 Q2	3.6	2.2	7.7	2	2.5	3	1.9	22.9
2011 Q3	17.9	6.8	2.6	8.5	3.1	2.4	1.8	43.2
2011 Q4	3.3	2.2	4.9	2	2.9	2.8	NA	18.1
2012 Q1	10.5	2.2	27.2	5.5	4.3	1.9	4.2	55.9
2012 Q2	7.5	6.8	2	1.8	4.7	9.2	2.4	34.5
2012 Q3	11.3	3.1	13.7	3.3	4.7	10.7	3.3	50.2
2012 Q4	2.7	2.6	6.7	4.9	1.7	3	5.6	27.1
2013 Q1	22.2	4.3	3.9	6.5	4	2.4	2	45.3
2013 Q2	8.4	7.9	1.8	3.1	7.3	1.7	3.5	33.8
2013 Q3	18.6	2.2	6.5	12	14.3	2.4	2.4	58.5
2013 Q4	3.2	2.2	5.2	2.9	1.7	2.3	2.4	19.9
2014 Q1	14.3	4.9	9.1	1.9	4.3	6.5	4.2	45.3
2014 Q2	2	14.4	11.3	7.6	8.5	3.1	2.4	49.4
All	364.6	258.2	288.9	193.3	186.2	163.7	100	1554.9

Table 25: OOS-LLH scores of the LFEs in the validation set using model M2. The scores are calculated for individual and aggregated risk categories for individual and aggregated time periods, whereby "NA" values mean that the particular risk category has no observation in that given time period.

Time period	EFRAUD	IFRAUD	EPWS	CPBP	EDPM	DPA	BDSF	All
2005 Q1	53.2	39.5	6.5	4.5	3.3	11.6	3.5	122.2
2005 Q2	5.1	1.9	8.8	3.8	3.6	2.4	7.6	33.2
2005 Q3	14.4	4.4	14.4	3	3.7	2.9	3.1	45.8
2005 Q4	7.9	13.2	15.2	3.3	14.1	2	2	57.7
2006 Q1	22	2	2.4	3.1	6.4	2.6	NA	38.4
2006 Q2	3.4	6.7	12.2	2.6	2.4	3.2	1.9	32.3
2006 Q3	10.9	40.9	5	3.6	11.2	3.2	5.5	80.4
2006 Q4	66.3	51.8	17.8	46.3	2.2	34.5	4.7	223.5
2007 Q1	5.5	13.9	3.9	4.5	7	3.2	2.4	40.3
2007 Q2	3.3	2.2	3.3	2.9	2.8	1.8	2.1	18.4
2007 Q3	8.6	3	10.8	3	13	2	2.2	42.7
2007 Q4	3.4	4.1	11.6	3	2.8	2	2.4	29.3
2008 Q1	3.4	2.1	6.1	2.9	3.5	2.1	3.8	23.8
2008 Q2	64.8	15.8	3.5	8.4	8.1	5.1	1.8	107.7
2008 Q3	8.1	2.4	7.9	5.8	6.1	1.9	6	38.2
2008 Q4	17.7	9.9	49	30.1	6.4	2.1	2.2	117.5
2009 Q1	31.4	7.3	2.8	3.5	6.4	7.2	3.2	61.9
2009 Q2	5.1	46.3	35	2.8	1.5	3	5.2	99
2009 Q3	4.7	3.8	4.7	6.6	7.9	1.9	2.2	31.7
2009 Q4	5.3	4.2	3.9	3.1	2.3	13.9	6.6	39.2
2010 Q1	3.1	7.4	2.4	2.3	2.8	1.9	2	22
2010 Q2	3.3	3.4	2	3.2	2.6	2.4	1.8	18.6
2010 Q3	5.5	8.1	4.5	2.3	2.7	2.5	2.5	28.2
2010 Q4	6.6	3.8	3.3	4.1	2.8	3.8	2.5	27
2011 Q1	5.7	7.3	2	3.9	2.7	2.7	NA	24.4
2011 Q2	14.2	2.2	50.2	3.9	3.3	8.2	1.6	83.5
2011 Q3	39.6	19	4.1	17.2	5.5	3.5	2.4	91.3
2011 Q4	21.5	4.6	29.8	3.7	7.8	8.9	NA	76.2
2012 Q1	5	1.9	15.7	3.8	2.8	1.6	3.3	34.1
2012 Q2	6.7	10	3.5	3.2	1.9	2.5	1.3	29
2012 Q3	4.6	2.2	6.4	2.5	3.1	6.1	4.5	29.4
2012 Q4	3.2	2.2	11.4	8.1	1.8	2.8	6.7	36.1
2013 Q1	16.6	3.2	3.3	5.4	3.2	2.1	1.8	35.7
2013 Q2	10.3	8.3	3.5	1.6	3.2	3.2	2.4	32.6
2013 Q3	8.8	1.9	4.4	5.5	8.8	1.8	1.8	33
2013 Q4	3.8	2.4	7.6	2.8	1.8	2.2	2.1	22.7
2014 Q1	3	2.3	3.6	1.5	3.7	2.6	2.1	19
2014 Q2	2.5	2.8	3.9	3	2.5	1.6	1.4	17.7
All	508.4	368.3	386.3	224.7	178	169.1	108.7	1943.4

Table 26: OOS-LLH scores of the LFEs in the validation set using model M3. The scores are calculated for individual and aggregated risk categories for individual and aggregated time periods, whereby "NA" values mean that the particular risk category has no observation in that given time period.

Time period	EFRAUD	IFRAUD	EPWS	CPBP	EDPM	DPA	BDSF	All
2005 Q1	25.3	18.1	3.5	2.7	4.2	16.2	2.3	72.4
2005 Q2	4.6	1.9	9.1	4.1	3.5	2.4	6.9	32.4
2005 Q3	8.4	3.5	8.9	3.1	2.9	2.4	2.6	31.8
2005 Q4	7.5	12.6	14.9	3.2	13.6	1.9	2.1	55.9
2006 Q1	13.3	2	2.3	5.2	3.2	2.1	NA	28.1
2006 Q2	3.6	4.2	8.7	2.4	2.2	2.7	1.8	25.5
2006 Q3	5.4	26.3	3.8	2.9	8.6	2.6	4.4	54
2006 Q4	41.6	12.8	9.9	12.5	4.1	12.6	2.7	96.2
2007 Q1	4.8	12.3	4.2	4.2	6.7	3.3	2.5	38.1
2007 Q2	3.2	2.2	3.4	2.9	2.8	1.9	2.1	18.5
2007 Q3	3.9	2.3	5.1	2.4	7.9	2	1.8	25.5
2007 Q4	3.6	3.8	9	3.2	2.6	2	2.6	26.7
2008 Q1	3.5	2	5.4	2.8	3.3	2.1	3.7	22.8
2008 Q2	20.8	3.4	2.1	2.8	3.5	2.9	1.7	37.1
2008 Q3	4.3	2.8	5	3.4	4.5	1.8	4.9	26.7
2008 Q4	3.4	4.8	9.7	9.7	2.6	3.2	3.2	36.5
2009 Q1	9.4	3.7	1.9	2.2	3.9	2.8	2	25.9
2009 Q2	2.9	12	8.9	1.9	2.6	2.2	2.8	33.4
2009 Q3	4.9	3.4	5.5	7.4	8.5	1.9	2.3	33.9
2009 Q4	3.5	3.4	3.1	2.6	2.6	9.7	5.6	30.5
2010 Q1	3.1	7	2.5	2.3	2.9	1.9	2	21.7
2010 Q2	3.2	3.3	2	3.4	2.5	2.5	1.8	18.7
2010 Q3	3.4	4.5	3.1	2	3.4	2.1	2.1	20.7
2010 Q4	3.6	3.2	4.5	3.2	3.5	4.8	2	24.8
2011 Q1	6.1	7.8	2	4	2.7	2.7	NA	25.3
2011 Q2	7.6	1.9	32.7	3.1	2.5	5	1.5	54.4
2011 Q3	16.9	6.3	2.5	8.1	3	2.3	1.8	41
2011 Q4	4.8	2.6	8.7	2.3	3.7	3.1	NA	25.1
2012 Q1	4.8	1.9	15	3.7	2.7	1.6	3.2	32.9
2012 Q2	3.5	4.8	2.7	2.5	2.2	2.9	1.5	20.1
2012 Q3	4.4	2.2	6	2.4	3	5.8	4.6	28.6
2012 Q4	2.6	2.7	5.8	4.4	1.7	3.1	5.4	25.7
2013 Q1	8.8	3.7	2.5	3.9	2.7	1.7	1.6	25
2013 Q2	9.7	7.8	3.4	1.6	3.1	3.1	2.4	31.2
2013 Q3	6	1.8	3.7	3.9	6.9	1.6	1.7	25.6
2013 Q4	3.8	2.4	7.8	2.8	1.8	2.2	2.1	23
2014 Q1	3.1	2.2	3.9	1.5	3.9	2.5	2.1	19.3
2014 Q2	2.4	2.9	3.7	3.1	2.5	1.6	1.4	17.6
All	275.7	206.7	236.9	139.7	148.7	129.5	95.2	1232.4

Table 27: OOS-LLH scores of the LFEs in the validation set using model M4. The scores are calculated for individual and aggregated risk categories for individual and aggregated time periods, whereby "NA" values mean that the particular risk category has no observation in that given time period.

Time period	EFRAUD	IFRAUD	EPWS	CPBP	EDPM	DPA	BDSF	All
2005 Q1	21.3	15.1	3.2	2.5	5	17	2.1	66.2
2005 Q2	4.6	1.9	9.1	4.1	3.5	2.4	6.9	32.5
2005 Q3	5.1	3	5.6	4	2.5	2.1	2.2	24.4
2005 Q4	11.4	17.7	16.9	3.8	17.4	2	1.9	71.2
2006 Q1	8	2.3	2.8	11.1	4.1	1.9	0	30.2
2006 Q2	4.3	3.6	7.4	2.3	2.1	2.6	1.9	24.1
2006 Q3	4.1	21.2	3.4	3.1	7.5	2.4	4	45.6
2006 Q4	47.2	19.8	11.6	18.7	3.5	16.9	3	120.8
2007 Q1	4.9	12.6	4.2	4.2	6.7	3.2	2.5	38.3
2007 Q2	3.9	2.1	3.7	3	2.5	1.8	2	19.1
2007 Q3	3.6	2.3	4.6	2.4	7.4	2	1.8	24.2
2007 Q4	4.6	5.1	19.5	3.4	3.3	2	2.1	40.1
2008 Q1	3.9	2.2	9.4	3.5	4.7	2	4.4	30.1
2008 Q2	11.1	3.5	2	3	2.7	2.4	2	26.7
2008 Q3	4.6	2.8	5.2	3.5	4.6	1.8	5	27.4
2008 Q4	7.8	3.2	3.6	4.8	2.6	4.3	4.5	30.8
2009 Q1	8	4.3	1.9	2.2	3.7	2.7	2	24.7
2009 Q2	4.4	5.7	4.5	2	3.4	2.8	2.3	25
2009 Q3	5.9	3.9	9	10.4	10.3	2.1	2.7	44.2
2009 Q4	5.7	6.2	2.4	2.5	3.4	5.6	4.5	30.3
2010 Q1	3.1	8.8	2.3	2.4	2.7	2	1.9	23.1
2010 Q2	3.2	3.3	2	3.5	2.5	2.6	1.8	18.9
2010 Q3	3.1	3.9	2.9	1.9	3.7	2	2	19.6
2010 Q4	5.9	3.5	3.5	3.9	2.8	4	2.4	26.1
2011 Q1	3.2	3.3	1.9	2.7	4.7	2.1	0	18
2011 Q2	3.5	2.2	8	2	2.5	2.9	1.9	23.1
2011 Q3	17.9	6.8	2.6	8.5	3.1	2.4	1.8	43.1
2011 Q4	5.3	2.7	9.4	2.3	3.9	3.2	0	26.8
2012 Q1	5.8	1.9	17.5	4.1	2.9	1.7	3.4	37.3
2012 Q2	3	3.2	2.3	2.2	2.7	4.2	1.7	19.2
2012 Q3	3.2	1.9	4.2	2.2	2.5	4.4	5.3	23.8
2012 Q4	3.4	2.1	13.1	9.3	1.8	2.8	7	39.6
2013 Q1	6.4	4.9	2.2	3.3	2.9	1.6	1.6	23.1
2013 Q2	7.5	6.1	3.1	1.6	2.8	2.9	2.2	26.2
2013 Q3	3.7	1.9	2.9	2.9	4.9	1.5	1.5	19.4
2013 Q4	8.7	3.3	18.1	4.1	2.1	2.7	1.7	40.7
2014 Q1	3.1	2.3	3.8	1.5	3.9	2.5	2.1	19.2
2014 Q2	2.3	3.1	3.2	3.5	2.7	1.7	1.5	17.9
All	266.7	203.6	233	156.3	156.3	127.3	97.9	1241

Table 28: Time-varying true ("True") parameters and parameter estimates using model M3 ("Est.").

	BDSF		CPBP		DPA		EDPM		EFRAUD		EPWS		IFRAUD	
Time period	Est.	True	Est.	True	Est.	True	Est.	True	Est.	True	Est.	True	Est.	True
2005 Q1	6.02	22	143.75	69	7.98	5	117.63	61	56.93	66	20.41	13	11.33	7
2005 Q2	4.82	1	115.13	94	6.39	5	94.22	66	45.6	54	16.35	17	9.07	21
2005 Q3	5.73	3	136.78	97	7.59	4	111.93	75	54.17	56	19.42	14	10.78	5
2005 Q4	6.07	7	144.88	173	8.04	6	118.57	161	57.38	91	20.57	25	11.42	30
2006 Q1	4.33	0	103.46	80	5.74	7	84.67	85	40.97	70	14.69	13	8.15	7
2006 Q2	4.91	8	117.17	104	6.5	6	95.88	105	46.4	68	16.64	17	9.23	10
2006 Q3	5.99	3	142.99	67	7.94	2	117.01	91	56.63	55	20.31	6	11.27	5
2006 Q4	6.74	10	161.04	210	8.94	3	131.79	176	63.78	97	22.87	71	12.69	27
2007 Q1	5.8	10	138.38	114	7.68	10	113.24	68	54.8	64	19.65	11	10.91	2
2007 Q2	5.32	6	127.12	125	7.06	8	104.03	98	50.35	49	18.05	21	10.02	11
2007 Q3	5.88	5	140.32	123	7.79	8	114.83	91	55.57	32	19.93	18	11.06	8
2007 Q4	8.1	8	193.37	230	10.73	7	158.25	146	76.58	69	27.46	28	15.24	21
2008 Q1	5.23	7	124.77	126	6.93	2	102.1	78	49.41	41	17.72	13	9.83	9
2008 Q2	3.84	7	91.71	147	5.09	4	75.05	78	36.32	37	13.02	18	7.23	8
2008 Q3	5.25	5	125.3	116	6.96	1	102.54	84	49.63	34	17.79	9	9.88	13
2008 Q4	6.87	11	163.98	116	9.1	13	134.19	132	64.94	36	23.29	20	12.92	5
2009 Q1	3.62	8	86.32	73	4.79	6	70.64	98	34.19	35	12.26	12	6.8	7
2009 Q2	3.85	1	92	129	5.11	8	75.29	102	36.44	31	13.07	11	7.25	7
2009 Q3	5.47	3	130.51	130	7.25	6	106.8	82	51.69	30	18.53	5	10.29	3
2009 Q4	7.13	1	170.23	161	9.45	12	139.31	137	67.42	38	24.17	21	13.42	8
2010 Q1	3.95	2	94.24	117	5.23	6	77.12	71	37.32	32	13.38	14	7.43	4
2010 Q2	4.85	2	115.78	112	6.43	6	94.75	92	45.85	37	16.44	18	9.12	8
2010 Q3	4.16	2	99.38	80	5.52	3	81.33	71	39.36	32	14.11	19	7.83	6
2010 Q4	4.04	2	96.51	84	5.36	11	78.98	79	38.22	44	13.71	21	7.61	3
2011 Q1	3.27	0	78.04	96	4.33	7	63.86	86	30.91	27	11.08	17	6.15	7
2011 Q2	2.93	3	69.92	137	3.88	7	57.22	78	27.69	38	9.93	12	5.51	6
2011 Q3	4.54	2	108.35	56	6.02	5	88.67	63	42.91	22	15.39	10	8.54	5
2011 Q4	3.98	0	94.92	108	5.27	7	77.68	104	37.59	41	13.48	19	7.48	10
2012 Q1	3.68	3	87.8	44	4.87	1	71.85	54	34.77	31	12.47	6	6.92	6
2012 Q2	2.84	2	67.76	80	3.76	6	55.45	61	26.84	21	9.62	7	5.34	8
2012 Q3	3.58	1	85.46	68	4.74	10	69.93	48	33.84	19	12.14	7	6.73	4
2012 Q4	4.78	5	114.23	133	6.34	13	93.48	104	45.24	38	16.22	18	9	5
2013 Q1	3.4	2	81.06	50	4.5	4	66.33	73	32.1	32	11.51	5	6.39	3
2013 Q2	2.11	5	50.31	75	2.79	6	41.17	60	19.92	24	7.14	9	3.96	4
2013 Q3	3.26	2	77.94	55	4.33	3	63.79	51	30.87	15	11.07	5	6.14	6
2013 Q4	4.2	5	100.25	126	5.57	3	82.03	89	39.7	39	14.24	14	7.9	10
2014 Q1	2.53	3	60.47	62	3.36	1	49.48	57	23.95	31	8.59	5	4.77	2
2014 Q2	2.58	2	61.66	68	3.42	2	50.46	47	24.42	22	8.76	4	4.86	7

A.12. Real-world application, scenario analysis

Table 29 presents the economic factors in this scenario analysis, where we again use economic reasoning to determine predetermined values for the economic factors. For this analysis, we fix LOR EU to the long-term average since Table 3 presents that the realized and expected number of losses deviate from what we expect using economic reasoning. After conducting the scenario analysis, LF parameter estimates in Table 30 are not aligned with our central argument that expansions drive extreme losses. This can be explained by the fact that model M3-met includes too many factors and hence the presence of multicollinearity yields extreme economic factor weights with contemporary signs for factors that exhibit high cross-correlations. We present the scenario analysis results for LFE-model M3 and further elaborate on the issues at hand in Appendix A.12. The reason is that we earlier found that "unemployment rate" has a net positive and strong impact for similar levels of UR EU and UR IT on extreme OpRisk losses. This is the case, because the sum of UR EU and UR IT coefficients of model M3 is positive, and both levels are high, as illustrated by Table 11. Since both UR EU and UR IT are used in the scenario analysis, the impact of the joint unemployment rate levels therefore is both positive and substantial. Consequently, the estimated number of losses is lower for favourable economic conditions.

Table 29: LFE; original values and normalized equivalents of economic factors using model M3 that represent scenarios SC0 - SC3.

		Constant	Recession		Intermediate		Expansion	
Factor	Impact	Value	Value	Norm. value	Value	Norm. value	Value	Norm. value
DGR	+	NA	-0.15	-1.08	0.03	-0.04	0.20	0.95
LOR EU	-	NA	6.8	-0.04	6.8	-0.04	6.80	-0.04
MIB IT	+	NA	-0.20	-1.81	0	-0.03	0.20	1.76
SIR IT	-	NA	5	1.74	2.50	0.17	0.5	-1.09
SP USA	+	NA	-0.12	-1.61	0	-0.17	0.10	1.04
UCSR	+	NA	-0.23	-2.31	0	0.12	0.15	1.7
UR IT	-	NA	13	2.67	9	0.54	5	-1.59
UR EU	-	NA	12	1.92	10	0.50	7	-1.62
VIX USA	*	NA	35	1.62	15	-0.55	20	0.00

Table 30: Scenario analysis; $\hat{\boldsymbol{\kappa}}$ estimates using the normalized economic factors in Table 29 as well as the M3-met model coefficients in Table 11.

	Constant	Crisis	Intermediate	Expansion
BDSF	4.45	4.28	4.93	3.21
CPBP	106.18	102.21	117.73	76.65
DPA	5.89	5.67	6.54	4.26
EDPM	86.89	83.64	96.35	62.73
EFRAUD	42.05	40.48	46.63	30.36
EPWS	15.08	14.51	16.72	10.89
IFRAUD	8.37	8.06	9.28	6.04
All	268.92	258.86	298.17	194.13

A.13. R program and associated functions

We hereby present an overview of the core functions used in this research. These functions can be called by the main file: "*Main.R*". We implemented all instances in a branch of the *BAMLSS* R package (Umlauf et al. (2019)). Our code can be viewed and downloaded in our own fork of the official *BAMLSS* GitHub page.[6] The functions start with "C." since our functions are added to the set of current *BAMLSS* functions.

```
#Simulation study, main functions.
#WARNING: these functions are computationally expensive
C.SimStudyFormulae()
C.SimStudySimulateData()
C.SimStudyMainUnpenalized(N = 5000, p_unf = 4, c = 1, nu = 0.05,
maxit = 100000, samples = 10)
C.SimStudyMainPenalizedP4(c = 1, samples = 10)
C.SimStudyMainPenalizedP11(c = 1, sample = 1)

#Simulation study, evaluation;
#WARNING: these functions are computationally expensive
C.SimStudyEvaluateAllModels(c = 1, s = 1)
C.SimStudyEvaluationOverAllIter()
C.SimStudyEvaluateP11()

#Simulation study, plots
C.SimStudyPlotAIC()
C.SimStudyPlotBIC()
C.SimStudyPlotEDF()
C.SimStudyPlotLLHInsample()
C.SimStudyPlotFNRandFPR()
C.SimStudyPlotMaxiter()
C.SimStudyPlotBetaCoef()
C.SimStudyPlotEtaCoef()
C.SimStudyPlotLLHOutofSample()

#Simulation study, tables
C.SimStudyEvaluateAllModelsTable()
C.SimStudyTableNumberofUpdates()
C.SimStudyTablePrepP4()
```

[6] https://github.com/ConstantijndeJonge/bamlss

```
C.SimStudyTablePrepP11()
C.SimStudyTableP4()
C.SimStudyTableP4L1norms()

#Simulation study, miscellaneous
#WARNING: These functions are computationally expensive
C.SimStudyCreateSimulatedCovaritates()
C.SimStudyDGP()
C.SimStudyMiscConvertRDAtoRDS_p4_B1()
C.SimStudyMiscConvertRDAtoRDS_p4_B2()
C.SimStudyMiscConvertRDAtoRDS_p11_B1()
C.SimStudyMiscStoreNumberofIterationsP4()
C.SimStudyMiscStoreNumberofIterationsP11()

#Real-world analysis, models
C.UniCreditFormulae()
C.UniCreditLFMain()
C.UniCreditLSEUnpen()
C.UniCreditLSEPen()

#Real-world analysis, scenario analysis & estimation of capital charges
C.UniCreditScenarioAnalysis()
C.UniCreditCapitalChargeEstimates()

#Real-world analysis, data preparation, tables, visualization
C.UniCreditDataPreparation()
C.UniCreditDataSummary()
C.UniCreditLLHContTable()
C.UniCreditLoadData()
C.UniCreditLSEEvalBoostingModels()
C.UniCreditLSEPenModelsTable()
C.UniCreditPlotLF()
C.UniCreditPlotLossesOverTime()
C.UniCreditCapitalChargePlots()

#Real-world analysis, miscellaneous
C.UniCreditEconomicFactorsCategorical()
C.UniCreditEconomicFactorsImpact()
C.UniCreditEconomicFactorsIntervalLimits()
```

```
C.UniCreditEconomicFactorsMetric()
C.UniCreditReadDataFile()
C.UniCreditRiskCategories()
C.UniCreditScenarios()
C.UniCreditTimePeriods()
C.UniCreditTimePeriodToInteger()
```