

Understanding the Economic Determinants of the Severity of Operational Losses: A Regularized Generalized Pareto Regression Approach

J. Hambuckers^{1,2,†}, A. Groll^{1,3} and T. Kneib¹

Abstract

We investigate a novel database of 10,217 extreme operational losses from the Italian bank UniCredit. Our goal is to shed light on the dependence between the severity distribution of these losses and a set of a set of macroeconomic, financial and firm-specific factors. To do so, we use Generalized Pareto regression techniques, where both the scale and shape parameters are assumed to be functions of these explanatory variables. We perform the selection of the relevant covariates with a state-of-the-art penalized-likelihood estimation procedure relying on L_1 -penalty terms. A simulation study indicates that this approach efficiently selects covariates of interest and tackles spurious regression issues encountered when dealing with integrated time series. Lastly, we illustrate the impact of different economic scenarios on the requested capital for operational risk. Our results have important implications in terms of risk management and regulatory policy.

Keywords: Operational loss; Generalized Pareto distribution; Penalized-likelihood; LASSO; Variable selection.

¹ Georg-August-Universität Göttingen, Chair of Statistics, Humboldtallee 3, 37073 Göttingen, Germany.

² University of Liege, HEC Liege, Center for Quantitative Methods and Operation management, 14 rue Louvrex, 4000 Liège, Belgium.

³ Technische Universität Dortmund, Faculty of Statistics, 44221 Dortmund, Germany.

† Corresponding author: jhambuc@uni-goettingen.de.

1 Introduction

Understanding the relationship between the severity distribution of operational losses and economic variables is particularly important for the banking industry, especially for risk management and regulatory purposes. Indeed, if we identify economic factors that explain variations in the severity distribution of these losses, we could improve the estimation of the associated risk measures and requested capital charge.

Operational losses are defined by the Basel Committee for Banking Supervision (BCBS) as “*direct or indirect losses resulting from inadequate or failed internal processes, people and systems or from external events*” [Basel Committee on Banking Supervision (BCBS), 2004]. In the Advanced Measurement Approach (AMA) proposed by the BCBS, the total operational loss L_t over the time period $]t - 1, t]$ is given by

$$L_t = \sum_{i=1}^{N_t} Z_{t,i}, \quad (1)$$

where N_t is the number of losses¹ and $Z_{t,i}$ the size of the i^{th} loss during the t^{th} period. From this representation, the operational risk capital charge is established as a function of the 99.9% quantile of L_t [Basel Committee on Banking Supervision (BCBS), 2010], denoted $Q_{0.999}(L_t)$ and defined such that $\mathbb{P}(L_t \leq Q_{0.999}(L_t)) = 0.999$. This quantity heavily depends on the distribution of $Z_{t,i}$ (i.e. the distribution of the loss size), referred under the term *severity distribution* in the banking literature. Hence, if we link properly this distribution with the economic context, we could obtain better estimates of $Q_{0.999}(L_t)$. Moreover, we would be able to assess the impact of adverse scenarios on expected capital requirements [Petrella and Resti, 2013]. Instead of using over-the-cycle risk measures, we could compute point-in-time or stressed risk measures related to specific economic conditions and construct more realistic forecasts.

Starting from the representation of the operational loss phenomenon given by equation (1), we study in detail the severity distribution of a novel sample of 10,217 extreme operational losses from the Italian bank UniCredit, over a ten-year period (2005Q1-2014Q2) and across seven event types². This is a unique feature of the present paper, as such a huge amount of private data is unusual and extremely difficult to obtain for academics [Chavez-Demoulin et al., 2016]. The lack and the poor quality of data have been obstacles in most empirical studies looking at the dependence between operational losses and economic factors. Indeed, confidentiality, heterogeneity and small sizes of public databases make it difficult to associate covariates to losses, and to analyse the data properly. To the best of our knowledge, the only exception is the research of Cope et al. [2012] who studied a database of 57,000 operational events across many firms and countries. Here, we focus on a single bank due to data availability. This is clearly a limitation that does not allow us to generalize our conclusions to the whole banking sector. However, our analysis

¹Often assumed to be larger than a given threshold.

²The latter referring to an internal classification of the losses based on their physical process.

brings many insights on the relationship between economic factors and the loss generation process in a typical financial institution.

We focus on the distribution of large values of $Z_{t,i}$ since these extreme events are the main drivers of $Q_{0.999}(L_t)$ ³. As covariates, we use a mix of firm-specific, macroeconomic and financial indicators, in addition to the event type. The choice of these variables is motivated by a number of theoretical reasons discussed in Chernobai et al. [2011], Cope et al. [2012] and Wang and Hsu [2013]. The main argument is that the sizes of operational losses are by nature subject to the economic environment: on one side, economic conditions influence the physical process of the losses because of the incentives they create. On the other side, the economic conditions influence the allocation of resources inside banks, which has then a direct impact on the way risks are measured, detected and mitigated.

This paper fills an important gap since no studies investigate the dependence between the severity distribution of large losses and economic factors. Early studies [Moscadelli, 2004, Chapelle et al., 2008, Soprano et al., 2009] focus on modelling the operational losses *independently from the economic conditions* with a model given by equation (1). More recently, the attention of researchers shifts towards the *conditional distribution* (i.e. conditional on the economic situation) of operational losses [Chernobai et al., 2011, Cope et al., 2012, Wang and Hsu, 2013, Chavez-Demoulin et al., 2016]. However, Chernobai et al. [2011] and Wang and Hsu [2013] focus on the frequency process only (i.e. the distribution of N_t), using Poisson and logistic regressions. Cope et al. [2012] study the dependence between the expected values of the loss size and macroeconomic covariates using a log-normal regression approach, but they do not focus on extreme events. Chavez-Demoulin et al. [2016] model the severity and frequency distribution of large losses but only using the time and the event type as covariates.

We conduct our analysis using Generalized Pareto (GP) regression techniques, in the idea of the Extreme Value Theory (EVT) and Peak-over-Threshold (POT) approach. More precisely, we model the distribution of the loss size $Z_{t,i}$ above some threshold τ , given a vector of covariates $\mathbf{x}_{t,i}$. We assume that the cumulative distribution function (cdf) of $Z_{t,i}$ given $\mathbf{x}_{t,i}$ is a Generalized Pareto distribution (GPD). To introduce a dependence with $\mathbf{x}_{t,i}$, we assume the scale and shape parameters of the GPD to be additive functions of $\mathbf{x}_{t,i}$. As a consequence, the scale and shape parameters are also time-varying. This modeling approach can be related to the so-called Generalized Additive Model for Location, Scale and Shape (GAMLSS) introduced in Rigby and Stasinopoulos [2005]. Chavez-Demoulin et al. [2016] use a similar approach, considering both fully parametric and semiparametric models. In the present paper, we restrict our attention to the parametric case. An alternative way to introduce time-variation is through the score mechanism of Creal et al. [2012] and Harvey [2013]. This procedure is used, e.g. in Massacci [2017] to model the tail dynamics of stock returns. However, Massacci [2017] only considers dynamic equations that are pure autoregressive processes and no economic variables are included. At the contrary, our approach has the advantage to allow identifying the economic sources of

³In Appendix A, the interested reader can find a short simulation study supporting this claim.

time-variation.

In this complex distributional regression framework, an important issue is to select the relevant covariates among a large number of potential predictors. However, testing and comparing all combinations of covariates is rapidly computationally infeasible, whereas stepwise regression is not particularly efficient in finding the optimal set of predictors [Tibshirani, 1996]. To overcome this issue, we develop a state-of-the-art penalized-likelihood estimation procedure adapted to the GP regression case and inspired by the local-quadratic approximations presented in Oelker and Tutz [2017]. We focus our attention on penalties based on the L_1 -norm of the regression coefficients.

In a simulation study, we investigate the ability of this technique to select the true set of covariates when performing the regression on both distribution parameters. We find that our approach identifies well informative and uninformative covariates and is particularly useful to escape the spurious regression trap [Granger and Newbold, 1974] when some covariates are nonstationary integrated time series. Applying this technique to UniCredit's data, our results show that only a limited number of economic variables have a significant association with the severity distribution. These results are important contributions of the present paper, since apart from stepwise regression and boosting [Mayr et al., 2012], no automatic variable selection procedures have been investigated in GP regression problems. The development and the use of a variable selection procedure are also novelties compared with previous empirical studies on operational losses, where this question is neglected [Chernobai et al., 2011, Wang and Hsu, 2013, Cope et al., 2012].

Our main findings regarding the determinants of operational loss severity are the following: a high GDP growth rate in the European Union, a low Italian unemployment rate and a high value of the VIX are associated with an increased likelihood of extreme losses at UniCredit. In addition, we find that imminent deleveraging periods are associated with a decrease in this likelihood. Using a model with interaction variables, we also find that the severity distribution of specific event types is impacted by changes in housing price and deposit growth rate. In terms of regulatory policy, these results indicate heavily that there is a need for setting a level of regulatory capital in accordance with the economic conditions.

Lastly, we illustrate the effects of different economic scenarios, and their consequences in terms of requested capital. To this end, we fit an additional inhomogeneous Poisson process on the frequency data and fix the relevant covariates to some values reflecting plausible economic scenarios. The resulting estimated parameters are used to simulate the total loss distribution and to compute estimates of the 99.9% quantile. Subsequently, we compare and discuss the values obtained in the different scenarios.

The rest of the paper is organized as follows: in Section 2, we describe the loss data and the covariates. Then, in Section 3 we introduce the models used for the study of these extreme losses. Practical considerations like estimation, inference and model selection are discussed in Sections 3.1 and 3.2. In Section 4, we study the finite sample properties of the proposed regularization method. We present the results of the regression analyses in

Section 5. In Section 6, we illustrate the effects of different economic scenarios on the requested capital charge obtained with the proposed approach. Lastly, we discuss and conclude in Section 7.

2 Data

In this section, we discuss the loss data as well as the economic covariates used in the regression analysis.

2.1 Description of the losses

We have initially at our disposal 41,871 operational losses provided by UniCredit operational risk department. Due to data collection procedures, only losses above 2,000€ are available. For anonymity reasons, the losses are scaled by an unknown factor. Even though scaling prevents us from any conclusion regarding the level of the losses, relationships with potential variables are not affected. UniCredit is one of the largest European commercial banks, and suffered heavily from the subprime and sovereign crises. It operates in 17 countries with over 143,000 employees. Around 50% of UniCredit's revenues come from its Italian activities.

As in Cope et al. [2012] and Chavez-Demoulin et al. [2016], losses are adjusted by means of the Italian consumer price index (CPI). The losses are dispatched into seven risk categories - called *event types* and referring to the physical process causing the loss. Descriptions and abbreviations are given in Table 1. The collection period ranges from January 2005 to June 2014. We have at our disposal the exact registration date of the event but we work on a quarterly basis since many macroeconomic variables are only available at this frequency. Figure 1 shows the raw data. Figure 2 shows the total loss and the number of losses, over time and across event types. The number of losses varies substantially for some categories, e.g. for the CPBP category.

Table 2 gives descriptive statistics of the losses per categories. Losses belong mainly to the CPBP and EDPM categories, as well as to the fraud categories (IFRAUD and EFRAUD). CPBP and IFRAUD losses exhibit a mean, median, standard deviation and third quartile particularly larger than the other categories. Looking at the values of these statistics over time, we observe several huge variations, indicating potential changes in the severity distribution over time (see, e.g. Figure 14 in Appendix B for the standard deviation).

To further investigate this supposition, we compute pairwise bootstrap Kolmogorov-Smirnov (KS) two-sample statistics between time periods. We reject the null hypothesis that the observed samples at each time period are drawn from the same distribution, either splitting the losses per event types or pooling all losses. Since in our empirical analysis we only study a sample of losses above a given threshold, we also apply the KS test to subsamples of the losses. We consider losses larger than the median, the third quartile and the 85% quantile, respectively. It reveals several significant variations in the

severity distribution over time. The detailed results of these tests and additional technical information can be found in Appendix B. The choice of our final threshold is discussed in the next section.

Lastly, we use a bootstrap likelihood ratio test to investigate the homogeneity of the frequency process. This assumption is rejected for all categories and all thresholds, indicating that the frequency rates also change over time. The results of the tests as well as some technical details can be found in Appendix C. These results indicate that if one wants to model the total loss distribution (as done in Section 6), inhomogeneity in the frequency process should be taken into account as well.

Both results are in line with the findings of Cope et al. [2012] for the severity distribution, and Chernobai et al. [2011] for the frequency process. They motivate us to investigate if changes in economic conditions could be associated with changes in the severity distribution.

2.2 Threshold selection

We focus our analysis on losses larger than some threshold τ , in the framework of the POT. Several authors proposed automatic threshold selection techniques [see Wadsworth and Tawn, 2012, for a review], but not in a regression context. Thus, these techniques are not applicable in our framework.

Here, we use instead a threshold derived from quantiles of the unconditional loss distribution, as proposed, e.g. in Chavez-Demoulin et al. [2014] or Chavez-Demoulin et al. [2016]. To ensure a good balance between correct specification of the GPD and low variance of the estimates, we set τ equal to the third quartile, giving us a final sample of 10,217 observations. To avoid creating a bias towards event types with fatter tails, we use event type-specific thresholds, namely the empirical third quartiles for losses from the same event type. Only losses exceeding these thresholds are kept for the regression analysis. This relatively low threshold is motivated by evidence from our simulation study, suggesting that a large sample size is needed to ensure a good model selection⁴. An additional motivation is that, in terms of total loss, this threshold allows us to consider losses that amount to a significant proportion of the quarterly total loss (on average, 87%). Therefore, this threshold allows us to incorporate most of the useful data when modelling the total loss distribution in Section 6. Figure 1 displays the thresholds and the losses included in the final sample.

This relatively low threshold might be responsible for detecting associations in our empirical analysis that do not relate clearly to extreme losses. To stress this issue, we also conducted an analysis using a threshold based on the 90% quantile (see supplementary material) and found similar results, at least regarding the main effects of the regression

⁴Even with samples of size $n = 4,000$ in the simulations, we tend to miss important explanatory variables when nonstationary variables are used in the regression. Conducting additional simulations, presented in the supplementary material, we found that a sample of size $n = 10,000$ reduces this issue.

analysis⁵. In Section 5, we only report and discuss results with τ chosen as the third quartile.

2.3 Description of the explanatory variables

As explanatory variables we use a set of 20 economic indicators computed at a quarterly frequency. We only consider lagged values of the variables to reduce potential two-way causality issues and to ease prediction. Since operational losses registration suffers regularly from a reporting delay, this set-up also avoids using variables that followed rather than preceded the losses. Extended descriptions can be found in Appendix D.

We consider three types of explanatory variables: firm-specific, macroeconomic and financial ones. The complete list can be found in Table 3 whereas time series are displayed in Figure 3. Firm-specific variables consist mostly in the event type and performance ratios. They bring information regarding the health of the financial institution, the profit generation as well as the size of the activity. Macroeconomic variables are related to GDP, unemployment rate, money supply and interest rates. They capture the effect of the general state of the economy. Lastly, financial variables measure the uncertainty, volatility and stock log-return. They account for the specific state of financial markets. We use a mix of indicators at the Italian, European and global level. In addition, we consider event type-specific effects, meaning that we use as covariates interactions between the binary event type variables and the economic factors.

Notice that we do not investigate the possible nonstationarity of the considered time series. Indeed, those are short (38 time periods) and would not allow us to conduct unit root tests with a sufficient power. Non-stationarity of the variables is, however, well handled by means of the regularization procedure as shown by the simulation study.

3 Methodology

In this section, we present first the regression model used to conduct our analysis, before detailing the estimation and regularization procedures. Lastly, we discuss penalty parameter selection and inference.

3.1 Model

We define here a regression model for the distribution of extreme losses, i.e. losses larger than a threshold τ . Indeed, in practice small losses are neglected. Consequently, N_t in equation (1) refers to the number of losses $Z_{t,i}$ larger than τ . We denote by n_t the observed value of N_t . Similarly to Beirlant and Goegebeur [2004], let us define the exceedances

⁵Additionally, since the data already received a POT treatment via the collection threshold, we might also follow Guillou et al. [2015] and use the smallest observed loss as threshold. We conducted a regression analysis using this strategy, but final QQ-plots suggested a misspecification problem in the lower tail, indicating that a higher threshold should be used. Thus, we settled for the third quartile. Associated results can be found in the supplementary material.

$Y_{t,i} = Z_{t,i} - \tau$ given that $Z_{t,i} \geq \tau$ and $i = 1, \dots, n_t$. Thus, the index i refers to the index of the i^{th} exceedance⁶. Relying on the EVT and the POT approach, we use the GPD to model the distribution of $Y_{t,i}$. The cumulative distribution function (cdf) of the GPD is given by

$$GPD(y; \gamma, \sigma) = \begin{cases} 1 - \left(1 + \gamma \frac{y}{\sigma}\right)^{-1/\gamma}, & \gamma \neq 0 \\ 1 - e^{-y/\sigma}, & \gamma = 0 \end{cases} \quad (2)$$

with $y \geq 0$. $\gamma \in \mathbb{R}$ and $\sigma > 0$ are the shape and scale parameters, respectively. For $\gamma < 0$, then $0 < y < -\sigma/\gamma$ and y is bounded. In the case of $\gamma = 0$, we observe an exponential decay. We restrict our attention here to the case where $\gamma > 0$, i.e. the heavy-tail case. Indeed, this assumption is in accordance with most of operational loss data [see, e.g. Moscadelli, 2004, Chavez-Demoulin et al., 2016]. This GPD approximation stems from fundamental results in extreme value analysis, known under the names Gnedenko and Pickands-Balkema-De Haan theorems [Gnedenko, 1943, Balkema and de Haan, 1974, Pickands, 1975]. The GPD approximation holds when the cdf of Z , $F(z)$, is of the type

$$F(z) = (z)^{-1/\gamma} G(z), \quad (3)$$

for some measurable, slowly varying function $G : (0, \infty) \rightarrow (0, \infty)$ so that

$$\lim_{z \rightarrow \infty} \frac{G(\omega z)}{G(z)} = 1, \quad (4)$$

for $\omega > 0$. This condition, known as *the regular variation* property, implies that the tail of the loss distribution decays at a power rate of Z and that the GPD is the limiting distribution of the exceedances [see, e.g. Embrechts et al., 1997, Beirlant et al., 2005, Davison and Smith, 1990, for theoretical details]. Therefore, in practice, it is usually assumed that the severity distribution of these large losses is effectively GPD and the regression analysis is performed on $Y_{t,i}$ [Aue and Kalkbrener, 2006, Moscadelli, 2004, Dutta and Perry, 2006, Chapelle et al., 2008, Soprano et al., 2009, Chavez-Demoulin et al., 2016]. In addition, notice that for $\gamma \in (0, 1)$, it can be shown that Y has a finite first moment [Chavez-Demoulin et al., 2016]. This is often a desirable property from a methodological perspective (e.g. for moment-based inference) as well as for applications involving the computation of a conditional expectation. Indeed, point estimates of γ that are larger than one are often discarded by practitioners. Here, we don't impose this additional restriction and discuss estimated values of γ with respect to their associated confidence intervals. More details are given at the end of Section 5.4.

Applying the POT method in a nonstationary context, we assume that the severity distribution of the exceedances is GPD, but with γ and σ depending on economic factors. Thus, for the i^{th} loss larger than τ taking place over the time period $]t - 1, t]$, we assume that

$$Y_{t,i} \sim GPD(Y_{t,i}; \gamma(\mathbf{x}_{t,i}^\gamma), \sigma(\mathbf{x}_{t,i}^\sigma)), \quad (5)$$

⁶Notice that due to the threshold condition, the losses come from a larger set. Therefore, the index related to this larger set - left undefined here for simplicity - and index i can differ for a given loss.

with $Y_{t,i} \geq 0$, $\gamma(\mathbf{x}_{t,i}^\gamma) > 0$, $\sigma(\mathbf{x}_{t,i}^\sigma) > 0$, and where these parameters can be characterized by structures of the type

$$\log(\gamma(\mathbf{x}_{t,i}^\gamma)) = \alpha_0^\gamma + \sum_{l=1}^{p_\gamma} \alpha_l^\gamma x_{t,i}^\gamma(l), \quad (6)$$

$$\log(\sigma(\mathbf{x}_{t,i}^\sigma)) = \alpha_0^\sigma + \sum_{l=1}^{p_\sigma} \alpha_l^\sigma x_{t,i}^\sigma(l), \quad (7)$$

where $x_{t,i}^\theta(l)$ denotes the l^{th} component of the vector of covariates $\mathbf{x}_{t,i}^\theta$ associated to $Y_{t,i}$ for $\theta \in \{\gamma, \sigma\}$, and p_γ (resp. p_σ) denoting the number of covariates for γ (resp. σ). In the rest of the paper, boldface refers to vectors. We use a logarithmic link function to ensure positivity of the parameters. Such a model is categorized as a parametric GAMLSS under the particular case of a GPD response distribution [Rigby and Stasinopoulos, 2005]. Parameters of this model are consistently estimated with a likelihood maximization procedure. We denote an observed sample of losses over T time periods by $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_T)$. The observed number of exceedances at each period is given by the vector (n_1, \dots, n_T) , such that the total sample size is given by $n = \sum_{t=1}^T n_t$. Then, under the assumptions of the conditional GPD model, the likelihood function is given by

$$L(\mathbf{y}; \Theta, \mathbf{x}) = \prod_{t=1}^T \prod_{i=1}^{n_t} \text{gpd}(y_{t,i}; \gamma(\mathbf{x}_{t,i}^\gamma), \sigma(\mathbf{x}_{t,i}^\sigma)), \quad (8)$$

where *gpd* denotes the probability density function (pdf) of the GPD, Θ is the set of all parameters of the model; \mathbf{x} a design matrix of all observed covariates, and $\gamma(\mathbf{x}_{t,i}^\gamma)$ and $\sigma(\mathbf{x}_{t,i}^\sigma)$ are the shape and scale parameters for the severity distribution of the i^{th} exceedance in period t . The log-likelihood function is given by

$$\mathcal{L}(\mathbf{y}; \Theta, \mathbf{x}) = \log(L(\mathbf{y}; \Theta, \mathbf{x})) = \sum_{t=1}^T \sum_{i=1}^{n_t} \log(\text{gpd}(y_{t,i}; \gamma(\mathbf{x}_{t,i}^\gamma), \sigma(\mathbf{x}_{t,i}^\sigma))). \quad (9)$$

An estimator $\hat{\Theta}$ is obtained by maximizing equation (9) with respect to Θ :

$$\hat{\Theta} = \arg \max_{\Theta} \{\mathcal{L}(\mathbf{y}; \Theta, \mathbf{x})\}. \quad (10)$$

Notice that we wrote equations (6) and (7) in a very general way, not explicitly accounting that we use lagged values of the covariates. Such a feature is included in a straightforward way by defining $\mathbf{x}_{t,i} = \tilde{\mathbf{x}}_{t-1,i}$, where $\tilde{\mathbf{x}}_{t,i}$ denotes the observed time series.

3.2 Regularization procedure

As discussed in Section 1, a recurrent issue when conducting GP regression is to identify the optimal subset of influential predictor variables (i.e. the variables that effectively enter the true model). We perform this task via a penalized-likelihood approach. In this approach, estimation and variable selection are performed simultaneously since the

penalization shrinks some coefficients down to zero, thus excluding some covariates from the final model. In particular, Tibshirani [1996] suggests the use of the L_1 -norm of the coefficients as regularization penalty, leading to what is known as the LASSO (least absolute shrinkage and selection operator) regression (see Efron et al. [2004], Zou [2006], Park and Hastie [2007] and Goeman [2010] for theoretical considerations and Fan and Tang [2012] for a broad overview). However, these studies mostly stayed in the framework of exponential family response distributions.

In the present paper, we rely on the same concept but in the GP regression case. More precisely, we consider the following penalized log-likelihood function:

$$\mathcal{L}_{pen}(\mathbf{y}; \Theta, \mathbf{x}) = \mathcal{L}(\mathbf{y}; \Theta, \mathbf{x}) - n\mathcal{P}_\nu(\Theta), \quad (11)$$

where $\mathcal{P}_\nu(\Theta)$ is the penalty with vector of penalty parameters $\boldsymbol{\nu} = (\nu_\sigma, \nu_\gamma)$. In particular, we use the following penalty:

$$\mathcal{P}_\nu(\Theta) = \nu_\sigma \sum_{l=1}^{p_\sigma} a_l^\sigma |\theta_l^\sigma| + \nu_\gamma \sum_{l=1}^{p_\gamma} a_l^\gamma |\theta_l^\gamma|, \quad (12)$$

where θ_l^σ (resp. θ_l^γ), $l = 1, \dots, p_\sigma$ (resp. p_γ) consists of the l^{th} parameter associated to the equation of σ (resp. γ). ν_σ and ν_γ are the penalty parameters associated to the coefficients of the predictors for σ and γ , respectively. In the classical LASSO case, $a_l^\sigma \equiv a_l^\gamma \equiv 1$ if corresponding coefficients are regularized, 0 otherwise. In the adaptive LASSO case (adLASSO), these weights are based on the inverse of the unpenalized estimates, such that $a_l^\sigma = 1/|\hat{\theta}_l^\sigma|$ (resp. $a_l^\gamma = 1/|\hat{\theta}_l^\gamma|$) if regularized, 0 otherwise. This weighted penalty has been found to provide more stable model selection in a variety of cases [Zou, 2006]. The final penalized-likelihood estimator is given by

$$\hat{\Theta}_{pen} = \arg \max_{\Theta} \{\mathcal{L}_{pen}(\mathbf{y}; \Theta, \mathbf{x})\}. \quad (13)$$

In the present case, no analytical solution to this problem exists and (13) needs to be solved numerically. An additional difficulty arises due to the non-differentiability of the penalty term. To overcome this issue, Oelker and Tutz [2017] developed a general framework for the approximation of different penalty types ensuring continuity, differentiability as well as sparsity of the final solution. They only considered the exponential family case, though.

Here, we follow Zou and Li [2008] and Oelker and Tutz [2017], and use a local quadratic approximation of the penalty terms within the estimation procedure. Relying on this approximation, the maximization problem can be linearized and solved with usual Newton methods. Technical details of our implementations are deferred to Appendix E. Additional theoretical considerations regarding the approximation of the penalized gradient and Hessian can be found in Oelker and Tutz [2017].

In our application, we consider the particular case of interactions of continuous and binary covariates. For dummy-encoded categorical covariates usually the group LASSO

[Meier et al., 2008] is used, which penalizes the whole group of the corresponding coefficients. In this way, either all dummies are included in the model or not. However, in our special situation, the categorical covariate reflects the event type (ET) and is a central aspect of the analysis: here, for each ET category it should be decided separately which corresponding interactions are selected. Therefore, we restrict ourselves to the conventional LASSO case only⁷.

For the selection of the vector of penalty parameters $\boldsymbol{\nu}$, usually AIC or BIC are used. In several studies [see, for example Fan and Tang, 2012, Groll and Tutz, 2014], the BIC criterion turned out to be more effective than AIC in selecting adequate penalty parameters and in ensuring a better sparsity of the final solution. Since we have both a scale and a shape parameter, we use a different penalty parameter for their respective subsets of covariates. The search is performed over a two-dimensional grid. The degrees of freedom (DF) can be estimated either by the trace of the product of the information matrix for the unpenalized likelihood function and the inverse of the information matrix for the penalized likelihood function [Gray, 1992], or by the number of non-zero coefficients in the final solution [Zou et al., 2007]. The BIC criterion, for a given vector $\boldsymbol{\nu}$ of tuning parameters, is given by

$$BIC(\hat{\Theta}_{pen}; \boldsymbol{\nu}) = -2 \mathcal{L}(\mathbf{y}; \hat{\Theta}_{pen}, \mathbf{x}) + \log(n) df(\boldsymbol{\nu}), \quad (14)$$

where n is the sample size and $df(\boldsymbol{\nu})$ are the estimated DF for the estimated model. For AIC, simply replace the factor $\log(n)$ by 2. The optimal $\boldsymbol{\nu}$ then minimizes (14). Alternatively, one could use cross-validation (CV) techniques, but these are computationally expensive. In Section 4, we show that BIC with estimated DF chosen as in Zou et al. [2007] is more suited than AIC to enable a good selection of the penalty parameters. In particular, it ensures simultaneously a low selection rate of uninformative covariates and a good selection rate of the informative ones.

To build confidence intervals, we have two approaches at our disposal. On the one side, we can use a bootstrap procedure but this is computationally expensive. On the other side, as an alternative, researchers proposed to use the regularization procedure as a covariate selection tool only. Then, the model corresponding to the non-zero coefficients is re-estimated without penalization (see, for example, Groll and Tutz [2014]). In that framework, asymptotic normality should hold and confidence intervals can be built using the inverse of the Fisher information matrix. Another advantage is that the resulting estimates are less biased, which is beneficial with regard to prediction. Hence, in the empirical analysis, we choose to follow this approach to obtain our final estimates of the coefficients, as well as for statistical inference⁸.

⁷It should also be noted that the presented approach could be extended to the case of the elastic net penalty [Zou and Hastie, 2005]. This would be particularly preferable in situations with highly correlated features where LASSO might behave erratically. Nevertheless, it implies selecting two additional penalty parameters - namely one per distribution parameter - which increases drastically the computational complexity of the problem and renders this task extremely challenging in practice. For this reason, we did not follow this path in the present paper.

⁸However, it has been argued in Goeman [2010] that the computed p -values tend to underestimate

4 Simulation

In this section, we study the finite sample properties of the proposed regularization procedure, both with LASSO and adLASSO penalties. We investigate the quality of the estimated parameters in terms of MSE and correct classification rate. In addition, since this is a critical step in the application, we consider different criteria for the selection of the penalty parameter.

4.1 Simulation with i.i.d. $N(0, 1)$ covariates

Firstly, we generate $N = 200$ samples of size $n = \{1000; 2000; 4000\}$ and $n = \sum_{t=1}^T n_t$ with $n_t = 1, \forall t$, leading to $T = n$. All covariates are assumed to follow independent standard Gaussian distributions $N(0, 1)$. We use $p = \{15, 25, 50\}$ covariates for both distribution parameters (i.e. overall 30, 50 and 100 covariates plus two constants). The number of active covariates is set to 3 for each corresponding linear predictor. We investigate the case where the covariate selection is performed on both linear predictors simultaneously. The selection of the penalty parameters is performed over a two-dimensional grid, using either BIC or AIC as defined in Section 3.2 with the degrees of freedom obtained either from Gray [1992] (subscript 1) or from Zou et al. [2007] (subscript 2) formulas. As a penalty, we use equation (12) with LASSO or adLASSO weights. The grids for the selection of the penalty parameters are defined between $[0.0025; 0.045]$ for LASSO, and $[10^{-5}; 0.009]$ for adLASSO. The values of the regression parameters for the active covariates are the following: $\boldsymbol{\alpha}^\gamma = (-.9, -.3, .2, .2)$, $\boldsymbol{\alpha}^\sigma = (4, .6, .4, -.3)$. To assess the quality of the estimates, we compute the ratio of mean squared error (RMSE) between penalized and unpenalized estimates, obtained with the different selection methods for the penalty parameters (a ratio below one is in favour of the penalized approach). We also compute the RMSE obtained with an additional unpenalized estimation step (superscript +). In that case, the reference MSE is obtained from the model with the covariates for which the p -values are lower than 0.05. To assess the ability of the regularization procedure to discriminate between informative and non-informative covariates, we compute three measures: the false positive rate (f.p.) that denotes the average proportion of selected uninformative covariates; the true positive rate (t.p.) that denotes the average proportion of selected informative covariates; and eventually the correct classification rate (CCR) that gives the global proportion of covariates correctly classified as informative or non-informative.

Detailed results are displayed in Tables 4 to 6. We focus on penalized LASSO estimates. Our results indicate that AIC_1 and AIC_2 ensure a low RMSE. However, it comes at the price of a poor regularization: we select on average between 45% and 56% of the uninformative covariates for AIC_1 , as indicated by f.p. Hence, the low RMSE of AIC_1

the overall level of uncertainty in the estimates as the uncertainty coming from the foregone selection procedure is ignored. We totally agree and stress the fact that, indeed, these p -values should be regarded with caution. Nevertheless, in a simulation set-up at least, our results suggest that this technique helps to decrease the mean squared error (MSE) of the final solution compared to the direct LASSO solution.

is achieved by shrinking the coefficients that should be zero but it is not sufficient to ensure a sparse solution. On the contrary, the method based on BIC_2 provides a better regularization procedure: we select on average between 0.4% to 7.1% of the uninformative covariates. BIC_2 also allows a good selection of informative covariates, as indicated by t.p.: it reaches 100% constantly for covariates associated to σ whereas it varies between 46.5% and 99.8% for covariates associated to γ . When n increases, we reach very high selection rates comparable to the other approaches. Moreover, the CCR obtained with BIC_2 varies between 91.5% and 98.9%, which is better than all other methods, except for $p = 15$ and $n = 1,000$. Hence, the important shrinkage leads to a good regularization but also to a large bias of the estimated coefficients, reflected in the relatively high RMSE. However, performing the additional estimation step reduces the RMSE (column RMSE(+)). This approach becomes particularly useful when the number of covariates is very large. In that case, the best RMSE(+) obtained with BIC_2 is between 50% and 80% lower than the one obtained with $Unpen.^+$, and is reduced compared to RMSE. The additional estimation step corrects for the bias of the active covariates, which in turn decreases the RMSE.

The results associated with adLASSO are qualitatively similar: whereas BIC_1 provides good estimates in terms of RMSE, selection rates improve with BIC_2 , and the additional estimation step reduces the RMSE. Compared with the LASSO penalty, RMSE and correct classification rates are slightly better.

4.2 Simulation with alternative models

To assess the proposed procedure in more complex frameworks, we consider four realistic extensions of the present set-up: firstly, we assume that the same covariates enter the structural equations of γ and σ , introducing collinearity. Secondly, we relax the Gaussian assumption of the covariates and use instead a t -distribution with the degrees of freedom equal to 5. Thirdly, we relax the independence assumption on the realizations of a given covariate in assuming that all covariates follow stationary AR(1) processes, with an autoregressive parameter of 0.7. Lastly, we relax the stationarity assumption of the covariates. We assume either that only the uninformative ones are nonstationary (specification denoted RW_1), or that all covariates follow a random walk (specification denoted RW_2). Associated results can be found in Table 7. For space considerations, we only present results for $n = 4,000$ and $p = 50$. We see that the results obtained with these more realistic models confirm the superiority of BIC_2 over the unpenalized approach to obtain good final estimates. Interestingly, when we add non-informative nonstationary predictors (panels RW_1), the penalization procedure tackles efficiently the classical spurious regression issue [Granger and Newbold, 1974]: we select very few uninformative covariates, whereas p -values constantly select a lot of them. A similar result is observed when the informative predictors are also nonstationary time series (panel RW_2), even though in this configuration we observe a lower CCR for γ . Thus, not being able to take care *a priori* of stationarity issues does not impact much the final conclusions when using the regularization procedures. Again, we observe qualitatively similar results with adLASSO

penalties, and slightly better results in terms of CCR and RMSE, but also lower t.p. for γ (panel RW_2). Nevertheless, even if the use of adLASSO might improve a bit the RMSE, it comes at the price of potentially worse t.p. rates when non-stationary covariates are included.

Finally, we investigate if higher levels of collinearity impact our results. We consider a last set-up where $n_t \neq 1$ and T is small. To generate the covariates, we use different multivariate time series models, including stationary, non-stationary and non-Gaussian effects with a dependence structure in the error terms. Indeed, in practice we observe several losses during a given time period and only short time series. Therefore, we assume $n_t \in \{80, 200\}, \forall t$, and $T = 50$, such that we obtain final sample sizes $n \in \{4,000; 10,000\}$ similar to the one in our empirical application. Results are presented in Table 8. Conclusions regarding the p -value-based technique are similar. In this configuration, it is clear that LASSO provides equivalent or better results than adLASSO in terms of RMSE, CCR and t.p. rates. We make the hypothesis that in this configuration the nonstationarity and collinearity of the covariates lead to highly variable weights $1/|\hat{\theta}|$ that deteriorate the adLASSO solution⁹.

Overall, these results suggest that, selecting the penalty parameters with BIC_2 and performing the additional re-estimation step, one achieves a good trade-off between sparsity, adequacy of the selected variables and MSE of the final solution. These findings are consistent with previous works that emphasize the good performance of BIC over AIC [Fan and Tang, 2012, Groll and Tutz, 2014]. We see that, even in more realistic frameworks, this technique allows to exclude the non-informative covariates. The counterpart is that we tend to regularize too much the informative covariates for γ when the collinearity level is high (especially for adLASSO). Thus, we can reasonably assume in our empirical applications that if a variable is included in the optimal model, the likelihood of a false detection is very low. We might not be able to identify all important variables if their effect is low, but at least we bear a very low risk of false associations.

For these reasons, we focus in the empirical analysis on the LASSO solution. The results of the empirical analysis obtained with adLASSO can be found in the supplementary material.

5 Empirical analysis

In this section we present the results of the regression analyses. As explained in Section 2.2, we only consider losses larger than the third empirical quartile, conditional on the event type. All covariates are standardized. We consider three different regression models:

$$\mathbf{Model\ 1:} \log(\theta(\mathbf{x}_t^\theta)) = \alpha_0^\theta + \sum_{c=1}^{C-1} \alpha_c^\theta ET(c),$$

⁹In the supplementary material, the interested reader finds technical details related to this simulation, as well as additional simulations using a mix of the different time series model considered here with $n_t = 1$.

$$\textbf{Model 2: } \log(\theta(\mathbf{x}_t^\theta)) = \alpha_0^\theta + \sum_{c=1}^{C-1} \alpha_c^\theta ET(c) + \sum_{l=1}^{p_\theta} \beta_l^\theta x_t^\theta(l),$$

$$\textbf{Model 3: } \log(\theta(\mathbf{x}_t^\theta)) = \alpha_0^\theta + \sum_{c=1}^{C-1} \alpha_c^\theta ET(c) + \sum_{l=1}^{p_\theta} \beta_l^\theta x_t^\theta(l) + \sum_{s=1}^q \delta_s^\theta INT(s),$$

for $\theta \in \{\sigma, \gamma\}$, where $ET(c)$ denotes the c^{th} binary variable coded from the C modalities of the event type variable, $x_t^\theta(l)$ is the l^{th} economic covariate, $l = 1, \dots, p_\theta$ and $INT(s)$ the s^{th} interaction variable, $s = 1, \dots, q$. $INT(s)$ is the interaction between category and continuous covariates¹⁰.

5.1 Model estimation

Table 17 in Appendix F displays the results of the unregularized regressions for Model 1 and Model 2. For each model, we find that the event type is a strongly significant predictor for both parameters. For Model 2, almost all economic covariates exhibit significant effects: we find 42 out of 54 parameters to be significant at the 5% test level. From a theoretical perspective, these results are hardly interpretable and we would prefer a more parsimonious model. Moreover, as shown in our simulation study, potential nonstationarity and collinearity can distort the size of the tests¹¹.

Table 9 shows the estimated coefficients obtained with the regularized approach based on LASSO for Model 2. Most variables are not included in the final model. For σ , beside the event type, only the Italian GDP growth rate and the PRF are selected. For γ , we find that the Italian unemployment rate, the VIX and the leverage ratio are selected¹².

The results obtained with the penalization method for Model 2 are appealing but rely on the over-simplistic assumption of similar effects across event types. To investigate potential interactions between event types and other economic factors, we fit Model 3 to the data. In the rest of this section we focus on this model. Overall, we have 294 parameters (147 for each distribution parameter) including 1 constant, 6 categories, 20 main effects and 120 category-specific effects. Relying on the unpenalized approach (Tables 18 and 19 in Appendix F), most interactions exhibit coefficients significantly different from zero, making this model impossible to interpret since 282 parameters are found to be significant. At the contrary, using the penalized approach, we obtain a sparser model where only a limited number of interactions are included.

Table 10 displays the estimated coefficients. For the clarity of the presentation, only interaction variables with non-zero coefficients have been included. For σ , we see that the main effect of the PRF disappears. In addition, two interaction variables enter the final model: one variable related to CPBP¹³ and the EU GDP growth rate, and the other one

¹⁰For clarity, we use different notations for the coefficients of the different types of variables: α for categorical, β for continuous and δ for interactions.

¹¹As additional evidence, we re-run the regression adding 5 uninformative time series of i.i.d. T(5) random variables to the equations of the parameters, and repeat the operation 500 times. We see that 67% of the time, we select between 1 and 5 of these variables (results are available upon demand).

¹²Using BIC_1 instead of BIC_2 , we select the same model plus the LT rate in the equation of γ .

¹³Customer, Product and Business Practices

related to EPWS¹⁴ and the housing price index growth rate. For γ , we see that only one new variable enters the final model: the interaction variable between EFRAUD¹⁵ and the deposit growth rate (DGR). This model is way more parsimonious and interpretable than the unpenalized one.

As shown in the simulation study, the regularization procedure based on BIC_2 might be too stringent. Using BIC_1 instead might provide insights on important covariates that are potentially omitted. Table 11 displays these results. Several new interactions enter the final model, suggesting additional dependencies with the short term rates, MIB returns, Italian GDP and VFTSE index. However, almost all of these new variables exhibit very small regression coefficients (close to the rounding level), suggesting that some of them might be excluded from the final model with few consequences. Similarly, the selection of highly correlated covariates indicates that there is redundant information. For example, VIX and VFTSE are both included in this model, whereas their correlation coefficient is .956.

5.2 Model comparison

To compare the different models, we first look at the generalized information score (IS) proposed by Nishii [1984]:

$$IS(k) = -2 \mathcal{L}(\mathbf{y}; \hat{\Theta}, \mathbf{x}) + k df(\boldsymbol{\nu}), \quad (15)$$

with $k \in [2; 10]$. For $k = 2$ and $k = \log(n) \approx 9.29$, we obtain the AIC and the BIC, respectively. Secondly, we look at the in-sample predictive ability of the selected subsets of variables by computing cross-validated likelihood scores. We do that in a time series fashion by setting aside all the losses from a given quarter (the validation set) and using the rest of the data (the training set) to estimate the parameters of the GPD models with the different subsets of covariates. Then, we compute the sum of the log-likelihood (LLF) and censored log-likelihood score (CLS) across validation sets. CLS has been proposed by Diks et al. [2011] to compare the goodness-of-fit in the tail when models have different tail probabilities. For a given validation set t , the CLS is given by

$$CLS_t = - \sum_{i=1}^{n_t} \left\{ \mathbb{1}(y_{t,i} > \kappa_i) \log(gpd(y_{t,i}; \hat{\gamma}(\mathbf{x}_{t,i}^\gamma), \hat{\sigma}(\mathbf{x}_{t,i}^\sigma))) \right. \\ \left. + \mathbb{1}(y_{t,i} \leq \kappa_i) \log(GPD(\kappa_i; \hat{\gamma}(\mathbf{x}_{t,i}^\gamma), \hat{\sigma}(\mathbf{x}_{t,i}^\sigma))) \right\}, \quad (16)$$

with $\mathbb{1}(\cdot)$ being an indicator function, and $\hat{\gamma}(x_{t,i}^\gamma)$ and $\hat{\sigma}(x_{t,i}^\sigma)$ are the distribution parameters estimated from the training set. As threshold κ we use the empirical quantile at levels between .5 and .99, for a given event type. This operation is repeated for the 38 time periods and we report the total predicted negative log-likelihood and censored log-likelihood over the validation samples. These quantities are denoted $CV(LLF)$ and $CV(CLS)$.

¹⁴Employee practices and Workplace Safety

¹⁵External Fraud

Our reference model is Model 3 based on BIC_2 . We compare it to models involving the additional estimation step and covariates being selected either by means of p -values or relying on the penalization approach. According to the IS (Figure 4), Model 3 based on BIC_2 is the best for $k \in [3.5, 8]$. Thus, our reference model seems to achieve a good trade-off between sparsity and complexity. At the contrary, using either the log-likelihood, the AIC or the BIC (Table 12), we end up selecting Model 3 based on p -values, Model 2 based on p -values or Model 1, respectively. However, these three models have important shortcomings, namely over-parametrization, absence of interactions and misspecified time-homogeneity. Their selection indicates that these information criteria penalize either too few or too much the complexity of the models.

Eventually, considering cross-validated likelihood and censored likelihood scores, the reference model performs best at all levels (Table 12). $CV(LLF)$ and $CV(CLS)$ are the smallest among the seven models. Figures 5 and 6 show score ratios and score differences based on different censoring thresholds. In particular, the reference model performs way better than Model 3 based on p -values, and reasonably better than Models 1 and 2. These results indicate also that the fit in the tail of the predicted density is the best with our reference model¹⁶.

Thus, in light of these indicators, we can safely conclude that Model 3 based on BIC_2 combines adequately sparsity and a good fit. This result holds both globally and in the far end of the tail. In the next sections, we focus on this reference model. In Section 5.3 we discuss its economic interpretation and subsequently illustrate the effect of different economic scenarios on the distribution of L_t in Section 6.

5.3 Economic interpretation of the dependence structure

We draw several economic interpretations from the signs of the estimated regression coefficients. Firstly, the signs of the coefficients indicate that an increase in the EU GDP growth rate is associated with an increase in σ . It suggests that in good economic times the likelihood of large losses increases. We observe the same effect for γ : a decrease in the Italian unemployment rate (i.e. an improvement of the general state of the economy) is associated with an increase in γ . These observations are explained by the fact that in a booming economy, the size of the transactions increases and so does the potential amount of money to be lost in case of failure of these transactions. Considering losses related to court settlement, it has been observed that the amounts of fines and compensation claims in lawsuits tend also to increase in good economic times [Cope et al., 2012]. In addition, booming economic conditions may also create incentives to commit frauds [Povel et al., 2007], increasing the likelihood of large losses related to fraud events. Overall, our findings are in line with the ones made by Cope et al. [2012].

Secondly, a positive regression coefficient related to the VIX implies that increasing values of the VIX are associated with larger values of γ , hence with an increased likelihood of

¹⁶Results for the adLASSO models can be found in the supplementary material. For a censoring level above 0.8, Model 3 obtained with LASSO also exhibits the best fit.

extreme severities. The VIX is considered to convey information on the expected stock market volatility (i.e. the uncertainty around stock returns) as well as on risk aversion through a variance premium component [Bekaert et al., 2013]. As shown in Bekaert and Hoerova [2014], the VIX has a high predictive power on the future level of financial instability through its uncertainty component. In particular, a high level of the VIX is indicative of a future high level of financial instability. Since most operational losses are related to financial transactions, it appears coherent that their severity distribution is impacted by financial instability. For example, losses related to execution, delivery and process management (EDPM) or to clients, products and business practices (CPBP) consist mostly of losses associated to derivatives and other financial products. A high future financial instability on the financial markets might be associated with potentially large unexpected variations of the price of these products. In case of failures related to delivery, pricing, selling or IT systems, it might easily lead to huge losses. Nevertheless, high values of the uncertainty component of the VIX are also indicative of a future decrease in economic activity [Bekaert and Hoerova, 2014] and, thus, might be indicative of a decrease in the likelihood of extreme events. However, the effect of the economic activity is already captured through the GDP and the unemployment rate. Therefore, the observed positive regression coefficient might reflect mostly the predictive content of the VIX for financial instability. Lastly, Bekaert et al. [2013] notice that high uncertainty and high risk aversion lead to a looser monetary policy stance as measured by lower real interest rates. Such a policy has been shown to incentivize financial actors towards more volatile investment strategies, opening the door for increasingly costly operational mistakes (see, e.g. Delis and Kouretas [2011] and Boubaker et al. [2017]). Indeed, when market volatility is high, huge adverse market movements tend to be more likely. Consequently, operational events generating market timing issues for both the bank and its customers are aggravated by these sudden market changes and translate into larger losses.

Thirdly, a negative regression coefficient related to the leverage ratio suggests that an increase in the leverage ratio during one quarter is associated with a decrease in γ the next quarter. This result seems counter-intuitive since an increase in leverage is usually associated with more risks and more risky behaviors of a bank [Bhagat et al., 2015]. However, there is evidence in the literature that leverage ratios are often mean-reverting [Löffler and Maurer, 2011]. Indeed, banks exhibit a long-run constant leverage ratio, sometimes altered by structural breaks [Koch, 2014]. Examining UniCredit's time series of the leverage ratio, we find evidence of a mean-reverting process. Fitting an AR(1) model to this time series, we find a significant autoregressive parameter, suggesting a long-term expected value around 18, which corresponds to a Tier-I capital ratio - with respect to the total assets - of 5.5%. This dynamic indicates that high values of the leverage ratio at the end of a period are likely followed by deleveraging actions during the next period, which corresponds to a less risky situation. Conversely, a low leverage ratio indicates an important likelihood of increasing the leverage the next period, which corresponds to a riskier situation in the future. In addition, it has been shown that banks suffering

from a huge degree of uncertainty regarding future losses tend to self-insure by holding more capital, i.e. decreasing their leverage ratio [Valencia, 2016]. Hence, a decrease in the leverage ratio seems to be indicative of a higher probability of large losses in the future, which is consistent with the negative regression coefficient observed¹⁷.

Looking at interactions, we observe first an additional effect of the EU GDP on losses belonging to the CPBP event type. CPBP losses are directly related to the commercial activities of the bank, especially the ones concerned with derivatives and other financial instruments. Thus, this additional sensitivity of CPBP losses towards the economic conditions is likely due to a stronger relationship between the size of the commercial transactions and the severity of the losses. Secondly, we observe a positive effect of the HPI growth rate on the likelihood of extreme EPWS losses. Those losses are mostly concerned with employee litigations. As pointed out by Cope et al. [2012], severance payments awarded to employees during lawsuits often depend on the economic conditions, the general level of the wages and past pay level. In periods of growth of the HPI, these three factors are particularly good in financial institutions and might increase the likelihood of large EPWS losses.

Lastly, we observe a large positive regression coefficient related to the deposit growth rate for losses related to EFRAUD (i.e. frauds committed by outsiders, like hacking or credit card frauds). The sign of this coefficient indicates that when the size of the transactions (here flows of deposits) increases, the size of the related losses increases as well, in particular the extreme ones. A similar effect has been discovered for internal frauds by Povel et al. [2007]. Our result is also consistent with Cope et al. [2012] who found a positive effect of GDP per capita on expected size of EFRAUD losses¹⁸.

5.4 Time series perspective

Figures 8 and 9 show the estimated shape and scale parameters, as well as the associated confidence intervals obtained with parametric bootstrap procedures. We see that σ decreases drastically in all categories between 2008Q2 and 2009Q2, which coincides with

¹⁷Notice also that for Model 2 and some of the alternative models studied in the supplementary material, we observe that an increase in the PRF is associated with increasing values of σ . This result can be related to findings in the banking literature [Williams, 2016, Laeven and Levine, 2007], indicating that banks relying heavily on non-interest incomes are more likely to suffer from agency problems. In particular, Laeven and Levine [2007] noticed that *“insiders may expand the range of financial activities if this diversification enhances their ability to extract private benefits from the financial institution”*, thus leading to an increased likelihood of large fraud losses from the said insiders. However, when including interactions, this is not found to be significant anymore.

¹⁸In the supplementary material, the interested reader finds the results obtained when using the empirical 90% quantile as the threshold. With this smaller sample, the model is heavily regularized and few covariates are selected, which could indicate a lack of power as suggested by the simulations. Essentially, we select the main effect of the VIX for γ and three interactions variables for σ , all related to the EFRAUD event type (among them DGR x EFRAUD). Using adLASSO, we select similar variables (VFTSE and Italian unemployment rate for γ , EU GDP x CPBP for σ) but also new ones (unemp. EU x BDSF and VFTSE x CPBP for γ).

the peak of the financial crisis. For this time period, we observe a highly negative EU GDP growth rate. Later, σ increases briefly, in parallel with the economic recovery. Then, the decrease in the GDP during the Great Recession is synonym of a second important decrease in σ . We also observe noticeable differences for EPWS and CPBP event type, due to the additional dependencies with the HPI and the GDP. Thus, during the crisis we observe a larger decrease in σ , whereas after the crisis we observe smaller increases in σ for these two event types.

Considering γ , we observe an increase in this distribution parameter during the financial and Euro crises (from 2008Q1 to 2011Q2), indicating a larger probability of extreme severity (all other things equal). This increase is linked to the the enormous surge of the VIX (whose value was multiplied between two and four times in three years) and to a decrease in the leverage ratio. Afterwards, when the VIX decreases to a more reasonable level, we observe a steady decrease in γ due to high unemployment rates. The large drop in γ at the end of 2007 is associated to a large increase in the leverage ratio following Uni-Credit's merger with Capitalia. Moreover, we also see a different pattern for the EFRAUD event type, driven by important variations in the deposit growth rate.

In light of these results, γ and σ seem to follow opposite paths, especially during the financial and Euro crises: σ exhibits first a huge decrease followed by an increase. At the contrary, γ displays first a large increase followed by a steady decrease.

To investigate how the severity distribution, as a whole, is influenced far in the tail by these variations, we plot the estimated 99% quantile over time (Figure 10). For comparison purposes, we also plot the quantile obtained from Model 1. However, since the losses have been scaled by an unknown factor, the absolute amount cannot be interpreted and only relative comparisons are relevant. We see that the IFRAUD event type is particularly risky compared with other categories. These results also indicate that variations can be important from one period to another compared to what is implied by Model 1. In particular, during the financial and the Euro crisis (period 2007Q4-2012Q4), Model 3 exhibits substantially larger estimated quantiles than Model 1, except for EFRAUD: on average, the differences range between 12% and 27%. These differences highlight the importance of including dependence features in operational loss models. Variations from one period to another are, however, limited. We observe median (resp. mean) absolute variation of the estimated quantile, from one period to another, between 6% and 11% (resp. 10% and 23%). This relative stability over time is appealing from a practical point of view to establish the requested operational risk capital¹⁹.

Lastly, we see that several estimated values for γ are larger than 1 (especially for the IFRAUD, EDPM and DPA classes). Practitioners might often consider these results implausible, since it implies an infinite expectation for the losses exceeding τ . However, those values are point estimates and need to be associated with confidence intervals. For IFRAUD, BDSF and DPA at least, the lower bounds of the confidence intervals are almost always lower than 1, so that we cannot firmly conclude that $\gamma > 1$ (see Figure 9). For

¹⁹The interested reader can find estimated 95% quantiles in the supplementary material

EDPM, losses often relate to pricing and delivery issues of derivatives and complex financial instruments. A value of $\gamma > 1$ reflects therefore that the expected loss in this category is likely to be extremely large, which has been the case in practice. In addition, whereas Model 1 suggests that γ is constantly larger than 1 for IFRAUD, EDPM and DPA, our conditional approach suggests that this is the case only in transitory crisis episodes. Overall, values of $\gamma > 1$ must be read as additional warnings that extreme losses might cause the bankruptcy of the bank. Given the systemic importance of UniCredit (it was ranked 8th at the global level, in February 2017), the associated global cost of a bankruptcy tends surely to infinity, as demonstrated by Lehman’s bankruptcy²⁰.

6 Scenario analysis

Lastly, we investigate the implications of using Model 3 as a conditional severity model, to establish the total operational risk capital charge. Regulatory capital for period t is defined as $Q_{.999}(L_t; \mathbf{x}_t)$ s.t. $\mathbb{P}(L_t \leq Q_{.999}(L_t; \mathbf{x}_t)) = .999$, i.e. the quantile at the 99.9% level of the total loss distribution given some explanatory variables \mathbf{x}_t . Here, \mathbf{x}_t encompasses all covariates related to both the severity and the frequency distributions. To obtain estimates of this quantity, we make the additional assumption that the frequency process of the exceedances follows an inhomogeneous Poisson process [Embrechts et al., 1997, Chernobai et al., 2011, Chavez-Demoulin et al., 2016]. More formally, the number of exceedances N_t over period $]t - 1, t]$, conditional on a set of covariates \mathbf{x}_t^λ , has the following distribution:

$$N_t | \mathbf{x}_t^\lambda \sim \text{Poisson}(\lambda(\mathbf{x}_t^\lambda)), \quad (17)$$

where $\lambda(\mathbf{x}_t^\lambda)$ is the expectation of N_t given \mathbf{x}_t^λ . We assume the following dependence structure between $\lambda(\mathbf{x}_t^\lambda)$ and \mathbf{x}_t^λ :

$$\log(\lambda(\mathbf{x}_t^\lambda)) = \alpha_0^\lambda + \sum_l^{\text{p}_\lambda} \alpha_l^\lambda x_t^\lambda(l). \quad (18)$$

Using the variables found to be relevant in Model 3 (Italian unemployment rate, EU GDP growth rate, VIX, leverage ratio, DGR and HPI), we estimate the parameters in equations (17)-(18) with the penalization procedure. Estimated parameters can be found in Table 13. Estimated values of λ over time are displayed in Figure 11. We see an important negative effect of the unemployment rate and the VIX on the frequency rate, coherent with previous observations. A high DGR is associated with a high frequency of operational events, especially for IFRAUD and CPBP event types. We also observe an important impact of the HPI across event types. Over time, the frequency rate displays high values before the crisis with a surge at the end of 2007, before decreasing until 2014.

²⁰Interestingly, if we look at the Marginal Expected Shortfall of UniCredit over time, e.g. on V-Lab website, <https://vlab.stern.nyu.edu>, we find a correspondence between periods of extremely large MES values and $\gamma > 1$. An interesting discussion related to this idea and referring to the dismal theorem of Weitzman [2009] can be found in Chavez-Demoulin et al. [2016].

With this full model at hand, we are now able to simulate $N = 10^6$ realizations of L_t , using equation (1). Then, our estimator $\hat{Q}_{.999}(L_t; \mathbf{x}_t)$ is given by the empirical quantile of these realizations. The simulation is repeated 100 times for each event type to quantify the uncertainty around these estimates. Notice that to make comparisons across categories, we add to the severity of each loss its corresponding threshold τ . We also obtain estimated quantiles using Model 1 and refer to these estimates as the *constant scenario*.

For the conditional model, we consider the following values of the explanatory variables reflecting plausible economic scenarios:

Crisis scenario: 15% Italian unemployment rate, -4% GDP EU growth rate, VIX at 37.5, leverage ratio at 12, HPI growth rate at -5% and a deposit growth rate of -15%.

Moderate scenario: 10% Italian unemployment rate, 1.5% GDP EU growth rate, VIX at 15, leverage ratio at 20, HPI growth rate at 1% and a 5% deposit growth rate.

Expansion scenario: 7% Italian unemployment rate, 3.5% GDP EU growth rate, VIX at 20, leverage ratio at 25, HPI growth rate at 10% and a 20% deposit growth rate.

Table 14 gives estimated values of σ , γ and λ associated with these scenarios. We see that Model 1 provides larger values of γ for all event types, and larger values of λ for IFRAUD and EDPM compared with Model 3.

Figure 12 shows boxplots of the estimated quantiles for the different configurations. We see that the constant scenario gives the largest estimates. This is due to γ and σ taking simultaneously large values. At the contrary, our regression framework allows accounting for the opposite effects of the financial instability on the one side (responsible for an increase in the risk) and bad macroeconomic conditions on the other side (responsible for a decrease in the risk). Discrepancies with respect to the expansion scenario vary from 5% (for EFRAUD) to 350% (for EDPM). Pooling all event types, the expansion scenario implies a 70% reduction in requested capital compared with the constant approach.

Overall, IFRAUD, EDPM and CPBP appear to be the main contributors to the total risk. The risk increases for all event types shifting from the crisis scenario to the expansion scenario. For IFRAUD, a large quantile of the total loss is due to the severity distribution exhibiting a particularly fat tail (γ close to 1, σ between 2 to 4 times larger than the other event types). It leads to a huge requested capital, despite a low frequency of events (between 2 and 10 events per quarter, on average). EDPM associates a high γ with a high frequency rate, whereas CPBP combines a high σ with even higher frequency rates. Across scenarios, a shift from the crisis scenario to the moderate scenario leads to a 50% increase in the total requested capital, whereas a shift from the moderate scenario to the expansion scenario doubles the requested capital.

These results suggest that the conditional approach achieves more reasonable capital charges than an approach based on the event type only. It also indicates that in good

economic times, a company should set aside more capital to cover itself, especially against operational losses related to selling and producing financial products (i.e. EDPM and CPBP event types), and to losses related to internal frauds. In times of crisis, however, even if IFRAUD losses have a severity distribution exhibiting a larger γ , their scale parameter and expected frequency rate are drastically reduced, leading to a smaller capital charge. Thus, this conditional framework avoids procyclicality in the considered crisis scenario.

7 Conclusion

In this paper, we detail a complete statistical procedure to study and model the severity distribution of extreme operational losses, given a particular economic context. To do so, we develop a Generalized Pareto (GP) regression framework, where relevant variables are automatically selected by means of a penalized-likelihood approach. The important choice of the penalty parameter is investigated in a simulation study, which suggests that the BIC is an adequate criterion to perform this selection. It also reveals that the penalized approach prevents us from spurious findings when some covariates are integrated time series. Subsequently, this procedure is applied to a novel dataset of operational losses from the bank UniCredit, with the objective to shed light on potential economic determinants of the severity distribution.

Our empirical study highlights interesting associations between the parameters of the severity distribution and several economic factors. Thanks to the proposed regularization procedure, our analysis selects a limited number of predictors and facilitates the interpretation of the results. We observe a strong association with the VIX, the Italian unemployment rate, the EU GDP growth rate and the leverage ratio. In addition, the deposit growth rate and the change in EU residential housing prices were found relevant for *external frauds* (EFRAUD) and *employee practices and workplace safety* (EPWS) event types, respectively. In view of our findings, higher unemployment rate and lower GDP growth rate imply a lower likelihood of extreme severities. At the contrary, higher uncertainty, risk aversion and future financial instability on the financial markets are synonyms of a higher likelihood of extreme losses. Higher deposit growth rates as well as higher housing prices are also associated with a larger likelihood of extreme losses for EFRAUD and EPWS event types. Eventually, we face the puzzling discovery that a low leverage ratio at the beginning of a period is associated with a high likelihood of extreme losses during that period. To explain this relation, we suggest that a low leverage ratio indicates an imminent leveraging phase or a self-insurance mechanism against risk taking, both situations leading to larger losses.

A cross-validation procedure suggests that selecting this small - but adequate - number of predictors provides the best fit. Focusing on the tail of the distribution with the help of a censored likelihood score [Diks et al., 2011], we find that our model with interactions is superior to the considered alternatives. In terms of risk management, it suggests that our

model should deliver a superior performance to compute good predicted risk measures, like quantiles far in the tail.

To study the total loss distribution, we fit an additional inhomogeneous Poisson model to the frequency process. This model is combined with the severity model to simulate the total loss distribution via Monte Carlo simulations, allowing us to compute the requested operational risk capital for three economic scenarios. This exercise suggests that IFRAUD, CPBP and EDPM event types are by far the biggest consumers of regulatory capital. We notice that taking into account the heterogeneity in the data reduces the capital charge compared to a model with constant parameters. It also highlights the fact that in good economic times (e.g. times exhibiting high GDP growth rate, low unemployment rate, strong deposit growth and an increase in housing prices), the implied capital charge increases drastically, especially for event types related to the commercialization of financial products and to internal frauds. In terms of regulatory consequences, this framework avoids unnecessary procyclicality and might be used to streamline the Advanced Measurement Approach, an important topic in the current operational risk literature [Peters et al., 2016].

Regarding extensions of the present work, the combination of the proposed procedure with copulas as in Brechmann et al. [2014] to take into account the intrinsic correlation across event types might provide a more general set-up. In addition, investigating a larger dataset covering several banks might also be of interest to generalize our empirical findings. Then, considering other types of penalties (e.g. fused LASSO or elastic net) in the regularization procedure might allow studying more complex regression frameworks, especially regarding interactions. Lastly, notice that the detailed econometric procedure might be used for analyzing different data with a GP regression approach. Among others, applications in climatology [Kharin et al., 2013], medicine [Das et al., 2010] or scientometrics [van Zyl, 2013] may benefit from the proposed technique.

Acknowledgement

The authors warmly thank two anonymous referees and the Editor (E. Ghysels) for detailed comments and suggestions. These have led to improve and to drastically streamline an earlier version of this paper. J. Hambuckers acknowledges the financial support of the Research Training Group 1644 *Scaling Problems in Statistics* funded by the German Research Foundation (DFG). T. Kneib acknowledges the financial support of DFG, grant KN 922/4-2. The authors thank F. Piacenza and A. Nassig for providing them with the operational losses data and some of the related covariates. J. Hambuckers thanks V. Chavez-Demoulin as well as the participants of HEC Lausanne seminars and the 10th EVA conference in Delft for the fruitful discussions that helped improving earlier versions of the present manuscript.

Appendix A Sensitivity analysis

This appendix illustrates the claim made in Section 1 regarding the importance of γ . To do so, we perform a small simulation study to show the effect of a variation in the frequency parameter (λ) or in the severity parameters (γ and σ) on the 99.9% quantile of L_t as given by equation (1). To do so, we generate 100,000 realizations of L_t with $N_t \stackrel{iid}{\sim} Poisson(\lambda)$ and $Z_{t,i} \stackrel{iid}{\sim} GPD(\gamma, \sigma)$ with $\lambda = 10$, $\gamma = .35$ and $\sigma = 5$ (configuration I). We take the 99.9% empirical quantile as a quantile estimate. This operation is repeated 30 times. A similar operation is performed in changing the parameters in the following way: $\lambda = 15$ (50% increase), other parameters stay the same (configuration II), and $\gamma = .42$, $\sigma = 6$ (20% increase in both parameters), λ stays equal to 10 (configuration III). Figure 13 shows the boxplot of the estimated quantiles in the three different configurations. Clearly, 20% variations in the severity parameters cause way bigger changes compared to a 50% increase in the quantile's value. This suggests that the severity parameters are the main drivers of the total loss distribution.

Appendix B Testing for nonstationarity of the losses

In this Appendix, we discuss the nonparametric bootstrap approach used in Section 2 to test if T samples come from the same distribution. Our goal is to avoid relying on existing tests that make particular assumptions on the conditional distributions of the considered variable (e.g. normality, equality of variance, symmetry), as well as to correct the test level for the multiple comparison issue. Our reasoning is the following:

- (I) Under the null hypothesis of no difference among conditional distributions over time, the loss sizes in each group $t = 1, \dots, T$ are independently drawn from a single unknown distribution F . Consequently, the expected value of the Kolmogorov-Smirnoff (KS) statistics should be close to zero.
- (II) Under assumption (I), the empirical distribution obtained by pooling together losses from different quarters is a good estimator \hat{F} of F .
- (III) The number of losses in each group is assumed constant and equal to n_t , for $t = 1, \dots, T$.

Thus, under these conditions, a suitable strategy to test if at least one group has a different distribution than the $T - 1$ other groups is the following:

1. For $t = 1, \dots, T$, draw a bootstrap sample \mathbf{y}_t^* of size n_t from $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_T)$.
2. Compute the bootstrap vector $\delta^* = (\delta_1^*, \dots, \delta_{T!/(2!(T-2)!)}^*)$ of all pairwise KS statistics.
3. Repeat steps 1 and 2 a large number of times B (e.g. $B = 2,000$) to obtain the bootstrap multivariate distribution of the vector δ (i.e. the vector of true pairwise KS statistics under the null of no differences among conditional distributions).
4. Compute a bootstrap estimator \hat{V}^* of $\text{Var}(\delta)$, and use its square root to standardized the B vectors δ^* .

5. Obtain $m^* = \max_i \delta_i^* / \sqrt{\hat{V}^*}$, i.e. the largest difference among all pairwise standardized statistics, for all bootstrap samples.
6. We reject the null hypothesis of no difference with a test level α if at least one element of $\hat{\delta}$ (i.e. the vector of observed differences from \hat{V} from the original sample) is not in $[0; q_{1-\alpha}^* \cdot \hat{V}^*]$, where $q_{1-\alpha}^*$ is the $1 - \alpha$ quantile of the bootstrap distribution of m^* .

This procedure takes into account the fact that the null hypothesis is a multiple hypothesis. References about the least favourable alternative principle are Scholz and Stephens [1987], White [2000] and Saadoghi-Alvandi and Malekzadeh [2014]. The results of this test can be found in Table 15. Notice that this test is likely to exhibit a low power, due to the use of the least favourable alternative. Power issues are also frequent with tests based on the KS test statistic [White, 2000]. Another possibility to test for a constant distribution over time consists in using a moment-based hypothesis test. As an example, we compute 95% confidence intervals over time, for the null hypothesis of a constant variance (Figure 14). This assumption is rejected for all event types.

Appendix C Frequency process

In this Appendix, we detail first the likelihood ratio test used to test the homogeneity hypothesis of the Poisson process in Section 2. More formally, the null hypothesis of the test is the following:

$$H_0 : N_t \sim \text{Poisson}(\lambda_t), \lambda_1 = \lambda_2 = \dots = \lambda_T = \lambda^+,$$

for $t = 1, \dots, T$. We have at our disposal a sample $\mathbf{N} = (n_1, \dots, n_T)$. Following Feng et al. [2012], under the null hypothesis the maximum likelihood estimator (MLE) $\hat{\lambda}^+$ of λ^+ is given by the sample mean $\bar{\Lambda}$, whereas under the alternative, the MLEs ($\hat{\lambda}_t, i = 1, \dots, T$) are given by \mathbf{N} . Then, the likelihood ratio test statistics is given by

$$LRT = 2 \sum_{t=1}^T n_t \log \left(\frac{n_t}{\bar{\Lambda}} \right).$$

We use a parametric bootstrap procedure to obtain the rejection region. We use the following algorithm:

1. Compute $\bar{\Lambda}$.
2. For $b = 1, \dots, B$, with B sufficiently large (e.g. 2,000), draw a bootstrap sample Λ_b^* of size T from a $\text{Poisson}(\bar{\Lambda})$ distribution.
3. For $b = 1, \dots, B$, compute $\bar{\Lambda}_b^*$ and LRT_b^* , i.e. the MLE under the null and the corresponding likelihood ratio statistics in the bootstrap world, respectively.
4. We reject the null hypothesis if LRT is greater than the $1 - \alpha$ empirical quantile of $\bar{\Lambda}_b^*$.

With this framework, we take into account the size of our sample ($T = 38$), as well as the estimation error on λ^+ . Results of this test can be found in Table 16, with different thresholds defining different realizations of N_t .

Second, we provide the detailed numerical results of the regression analysis for the frequency process done in Section 6. Table 13 provides the estimated coefficients whereas Figure 11 displays the estimated λ over time and across event types.

Appendix D Explanatory variables

In this section, we provide detailed descriptions of the covariates introduced in Section 2. Firm-specific variables are internal ratios and have been provided by the operational risk department of UniCredit. Macroeconomic and financial variables have been downloaded from Thomson Reuters Datastream for S&P500, TRSI, MIB, VIX, VFTSE and the unemployment rate, from EUROSTAT website for the GDP growth rates (<http://ec.europa.eu/eurostat/data/database>), from FRED, the database of the Federal Reserve Bank of St. Louis (<https://fred.stlouisfed.org>) for the housing price index and the Italian interest rates (notice that FRED also aggregates data from the OECD website - <http://stats.oecd.org> - where detailed definitions of the time series are provided). Last, the monetary aggregate M1 and the consumer loans rates have been downloaded from ECB database.

Event type (ET): internal classification of the recorded event, indicating a difference in terms of the physical process of the loss formation. There are 7 different types of events, based on the Basel classification [Basel Committee on Banking Supervision (BCBS), 2016]. Regulators explicitly expect banks to model the operational losses according to this internal classification. The event type is coded into 6 binary variables and the reference category is DPA.

Tier-I capital ratio (TCR): ratio between the core capital of the bank (equity capital, disclosed reserves, preferred stocks) and the total asset. The Basel framework focuses on this ratio, with the goal to ensure that the risk exposure of the bank is backed by sufficient capital [Basel Committee on Banking Supervision (BCBS), 2010].

Leverage ratio (LR): total value of the assets divided by the value of the equity. A low leverage ratio indicates a balanced investment policy, relying on internal funds. At the contrary, a highly leveraged bank relies heavily on external debts. The revised Basel III framework plans a mandatory leverage ratio for the 2018 horizon, with the goal to limit the probability of a bank run and to compensate for imperfect information about the asset's quality of a bank [Dermine, 2015].

Percentage of the revenue coming from fees (PRF): proportion of non-interest income. This quantity conveys information about the economic well-being and the risk-taking behavior in the bank, especially related to agency issues [Williams, 2016].

Deposit growth rate (DGR): proportional variation of total deposits from one quarter to another (in %). This quantity can be seen as a measure of the bank's funding dynamic.

Unemployment rate (UR IT and UR EU): proportion of people searching for work among the active population. This quantity provides information about the general state of the economy. We consider both the unemployment rate in Italy and in the European Union (EU).

Gross domestic product (GDP) growth rate: measure of the variation of wealth, over time (in yearly

%). It can be a determinant of fine amounts in lawsuits [Cope et al., 2012], beside measuring also the health of an economy. The data have been adjusted for seasonality. As for the unemployment rate, we consider the GDP in Italy and in the EU.

Housing price (HPI) growth rate: growth rate of residential property prices index for the Euro area, in yearly %. It measures the dynamics of the prices on the housing market, mortgage lending being an important business for banks.

Monetary aggregate (M1) growth rate: Average monthly growth rate (in yearly %) over the quarter of Euro M1 aggregate (defined as currency, i.e. banknotes and coins, as well as balances which can immediately be converted into currency or used for cashless payments, i.e. overnight deposits, see <https://www.ecb.europa.eu>). It provides information regarding money supply dynamics and monetary policy.

3-month Italian interbank rate: short term interest rate for Italy, computed by OECD. This rate refers to the three month interbank offer rate attaching to loans given and taken amongst banks for any excess or shortage of liquidity over several months. This quantity indicates the price of its short-term borrowing. It conveys information about the liquidity and the general state of the financial system.

10-year Italian government bond yield: long-term interest rate for Italy, computed by OECD. As defined by the OECD website, such interest rates are the ones implied by the prices at which the government bonds are traded on financial markets. Long-term interest rates influence investment expenses. Low long-term interest rates encourage investment in new equipment and high interest rates discourage it.

Consumption loan rate (LOR): Floating rate or initial rate fixation of up to one year, for Euro-denominated consumption loans made by Euro-area households (in yearly percentage, on new business). It provides information regarding the cost of consumption goods. As for GDP and UR, we use LOR at the Italian and the EU level.

VIX: Volatility index, computed from a panel of S&P500 options prices. Also known as a *fear index*, it was shown to be a good proxy for risk aversion [Bekaert et al., 2013]. It also reflects the uncertainty on stock markets [Bekaert and Hoerova, 2014].

VFTSE: Volatility index, based on options of the FTSE100 index. Contrary to the VIX based on American stocks, VFTSE is based on UK stocks and reflects the uncertainty on European markets. Similarly to the VIX, it might also convey information regarding risk aversion. VFTSE has been preferred to VSTOXX due to data accessibility reasons and its important use among practitioners.

S&P500 returns: Quarterly log-returns of S&P500 stock index, based on 500 large capitalizations in the U.S, used as a measure of the overall dynamic of financial markets.

Thomson Reuters European stock index (TRSI) log-returns: Quarterly returns of TRSI index, based on around 2,000 capitalisations in Europe, used as a measure of the dynamic of European financial markets.

Milano Italia Borsa (MIB) returns: Quarterly log-returns of MIB index, based on the largest 40 capitalisations on Milan stock exchange, and used as a measure of the dynamic of Italian financial markets.

UniCredit stock returns (SR): Quarterly log-returns of UniCredit stock price at Milan stock exchange. It measures the evolution of UniCredit's value in the eyes of the financial market.

Appendix E Regularization procedure

In the present situation of conventional LASSO, suitable approximations for the penalty and its first derivative are the following:

$$|\xi| \approx \sqrt{\xi^2 + c}, \quad (19)$$

$$\partial|\xi|/\partial\xi \approx (\xi^2 + c)^{-1/2}\xi, \quad (20)$$

where c is a small constant²¹. Throughout the paper, we use $c = 10^{-7}$ (Oelker and Tutz [2017] recommend to use at least $c = 10^{-5}$). Then, the linearized problem

$$\mathbf{s}_{pen}(\Theta_{(k)}) + \mathbf{H}_{pen}(\Theta_{(k)})(\Theta - \Theta_{(k)}) = 0 \quad (21)$$

can be solved linearly, where \mathbf{s}_{pen} and \mathbf{H}_{pen} denote the gradient and Hessian of the penalized-likelihood function. It consists in computing the sequence

$$\hat{\Theta}_{(k+1)} = \hat{\Theta}_{(k)} - \mathbf{H}_{pen}(\hat{\Theta}_{(k)})^{-1}\mathbf{s}_{pen}(\hat{\Theta}_{(k)}) \quad (22)$$

until convergence²². Starting values are obtained, e.g. from unpenalized regression. For more technical details, especially how \mathbf{H}_{pen} and \mathbf{s}_{pen} can be approximated, see Oelker and Tutz [2017]. Note that due to the imposed penalty from equation (12) the inversion of the Hessian in equation (22) is unproblematic as long as ν_σ and ν_γ are chosen large enough.

Appendix F Unpenalized estimation

In this section, we provide the estimated parameters of the unpenalized estimates, for Models 1 to 3 (Tables 17 to 19).

²¹Note that due to the quadratic approximation of the penalty in (19) and (20), coefficient estimates cannot be set exactly to zero. Instead, coefficients that should be zero differ from zero in some rear decimals. For this reason, one usually uses rounded coefficients with the consequence that estimates very close to zero are simply set to zero and the corresponding covariates are excluded from the model (see, for example, Table 1 in Schauburger and Tutz [2017]). In the present work, we rounded coefficient estimates to the third decimal, see Table 9.

²²In terms of stopping conditions, we perform up to 200 iterations or stop the algorithm if we observe proportional variations lower than 10^{-8} for the vector of parameters and the objective function.

Figures

Main text

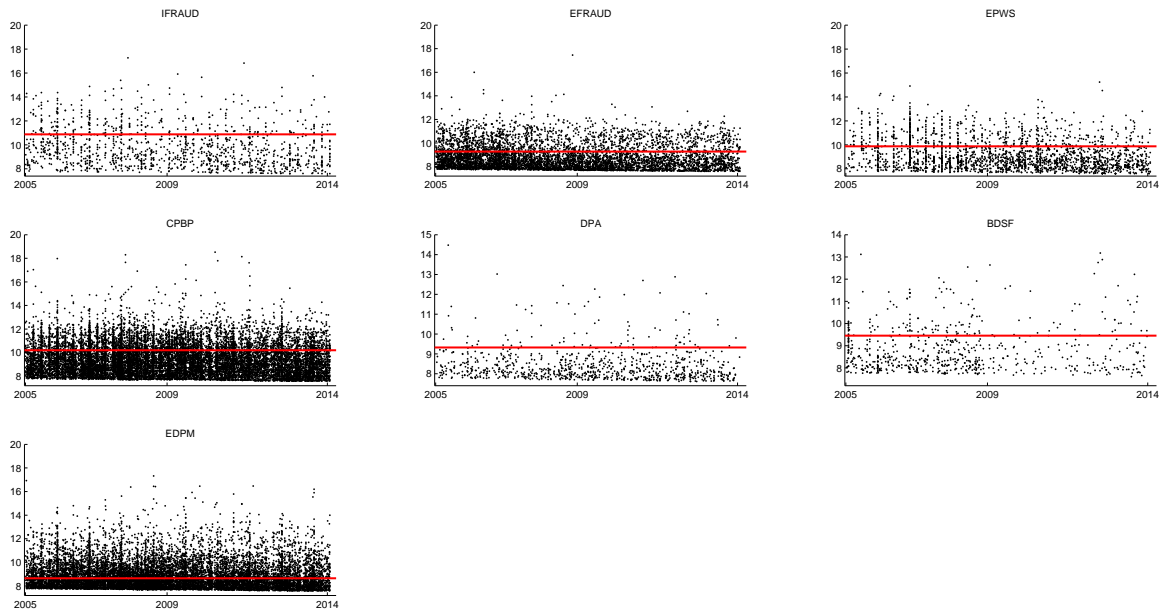


Figure 1: For each event type, registered (log) operational losses over time. Red: third quartile.

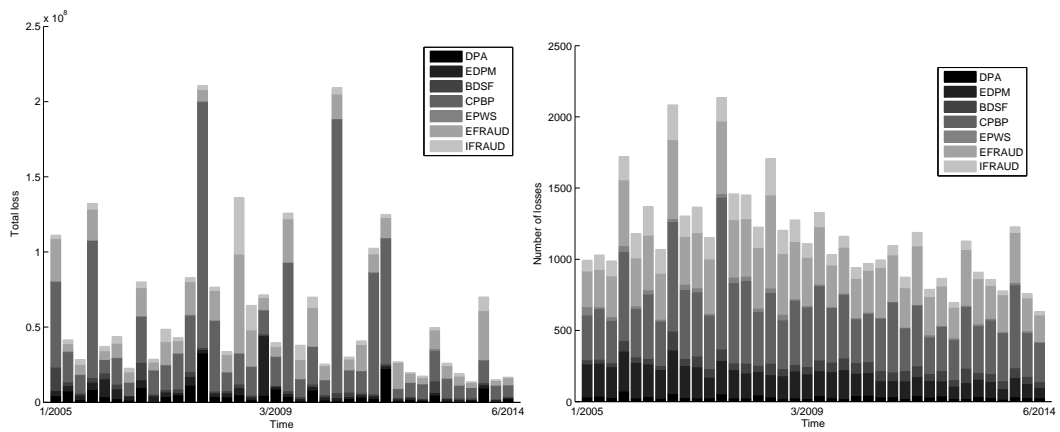


Figure 2: Quarterly sum of the operational losses in € (left) and number of operational losses (right) across all event types, over the period 01/2005 - 06/2014.

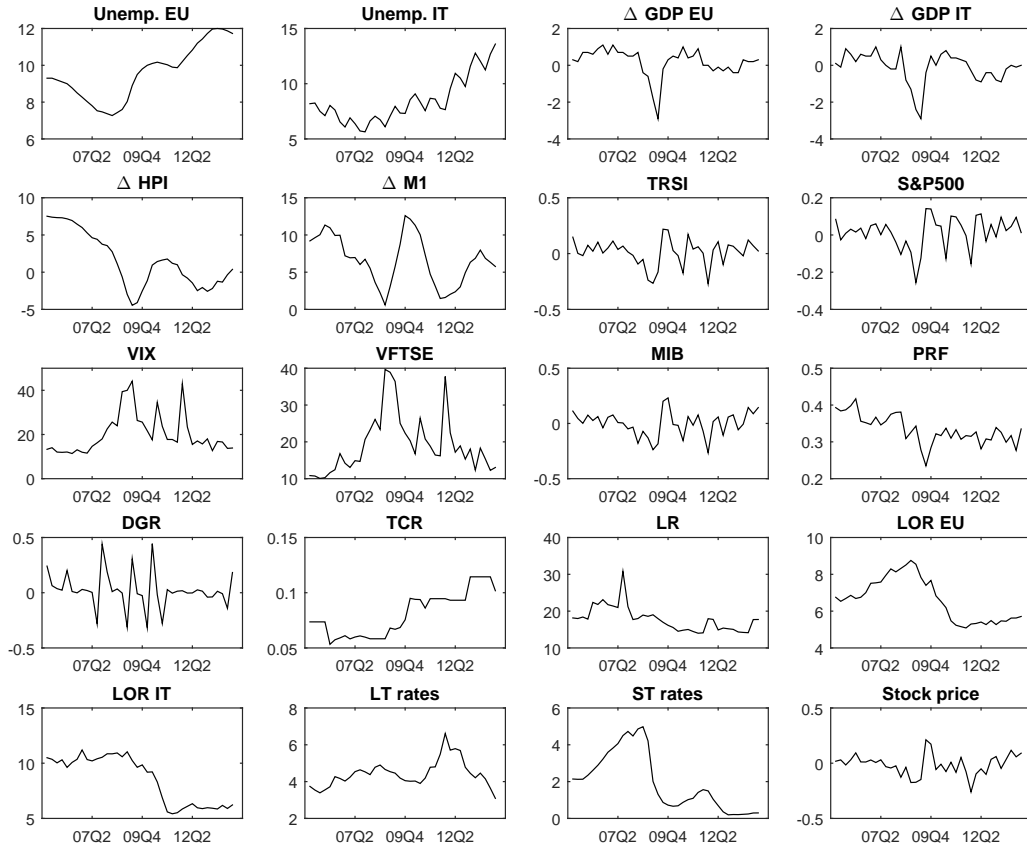


Figure 3: Time series of the considered economic factors, between January 2005 and June 2014. Rates are expressed in %.

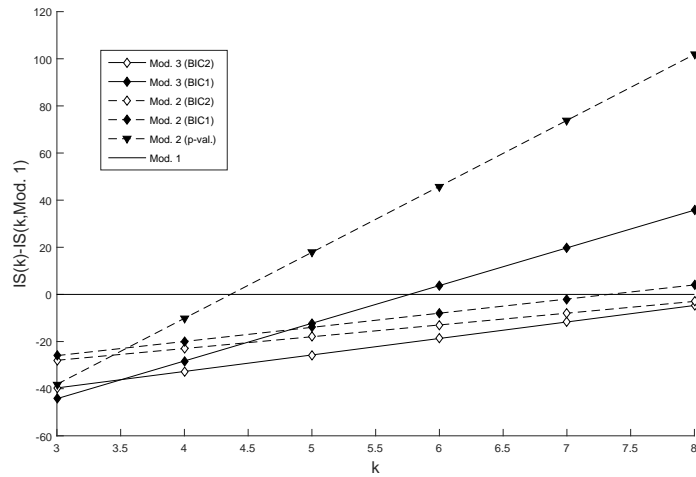


Figure 4: Information score $IS(k)$ with different values for k . Y-axis: difference with Model 1 (values smaller than 0 indicate a better fit of the alternative model). Solid: Model 3. Dashed: Model 2. \diamond (resp. \blacklozenge): selection based on BIC_2 (resp. BIC_1). \blacktriangledown : selection based on p -values. Model 3 based on p -values has been omitted for clarity, its values ranging from 350 to 1693.

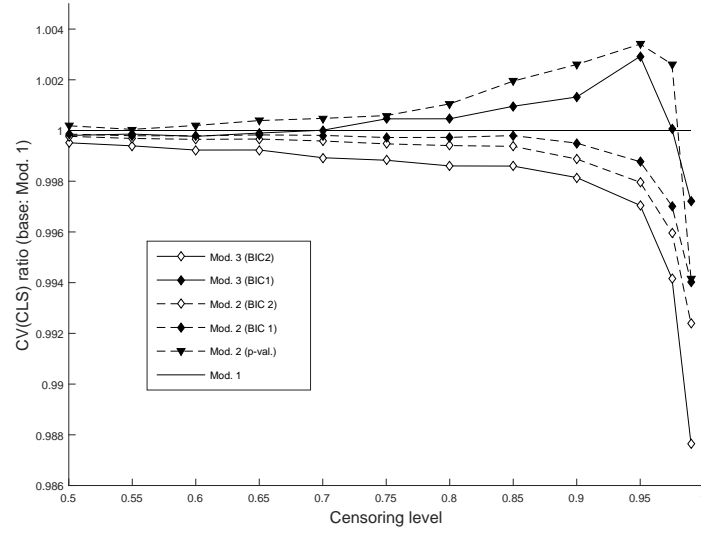


Figure 5: Cross-validated CLS for different values of κ , based on empirical quantiles. X-axis: quantile level used for the threshold. Y-axis: ratio between $CV(CLS)$ of a given model and $CV(CLS)$ of Model 1, such that values smaller than 1 are in favor of the alternative model. Solid: Model 3. Dashed: Model 2. \diamond (resp. \blacklozenge): selection based on BIC_2 (resp. BIC_1). \blacktriangledown : selection based on p -values. Model 3 based on p -values has been omitted for clarity, its values ranging from 1.033 to 1.40.

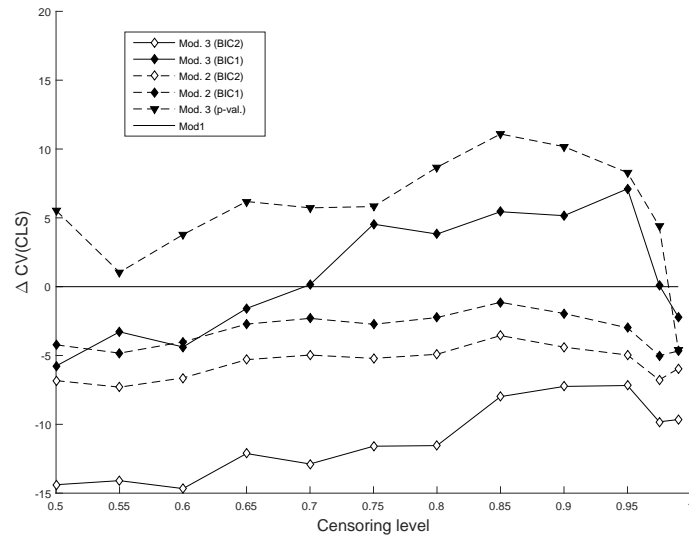


Figure 6: Difference in cross-validated CLS for different values of κ , based on empirical quantiles. X-axis: quantile level used for the threshold. Y-axis: Difference between $CV(CLS)$ of a given model and $CV(CLS)$ of Model 1, such that values smaller than 0 are in favor of the alternative model. Solid: Model 3. Dashed: Model 2. \diamond (resp. \blacklozenge): selection based on BIC_2 (resp. BIC_1). \blacktriangledown : selection based on p -values. Model 3 based on p -values has been omitted for clarity.

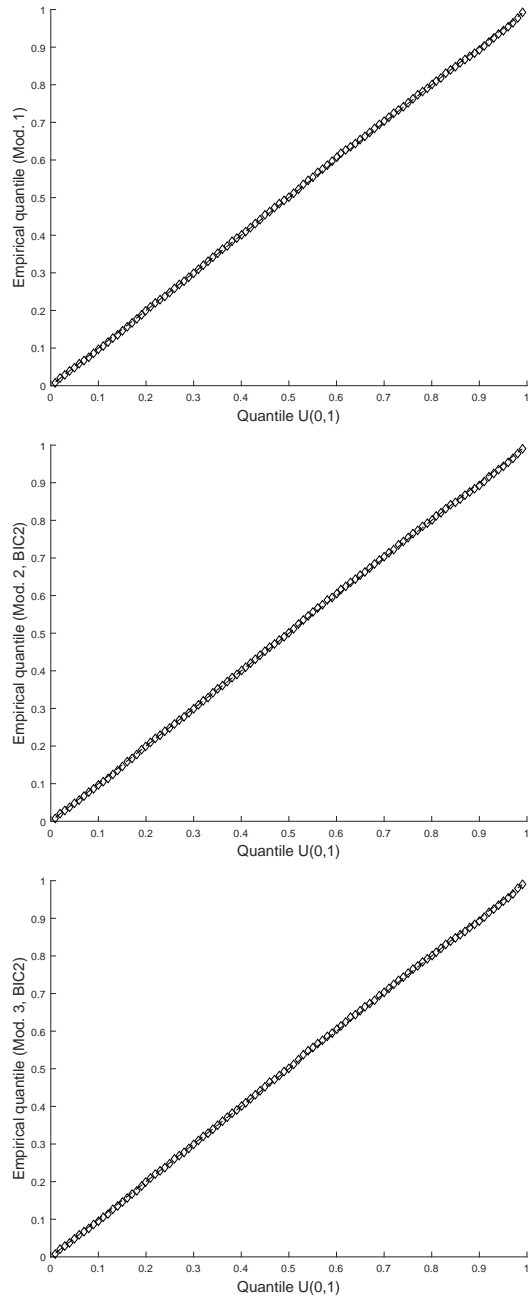


Figure 7: QQ-plots of the (pseudo-)residuals for Model 1, and penalized Models 2 and 3 with BIC_2 . Dashed: quantiles of $U(0,1)$; Diamond: quantiles of the residuals. The residuals are obtained applying the probability integral transform to the exceedances.

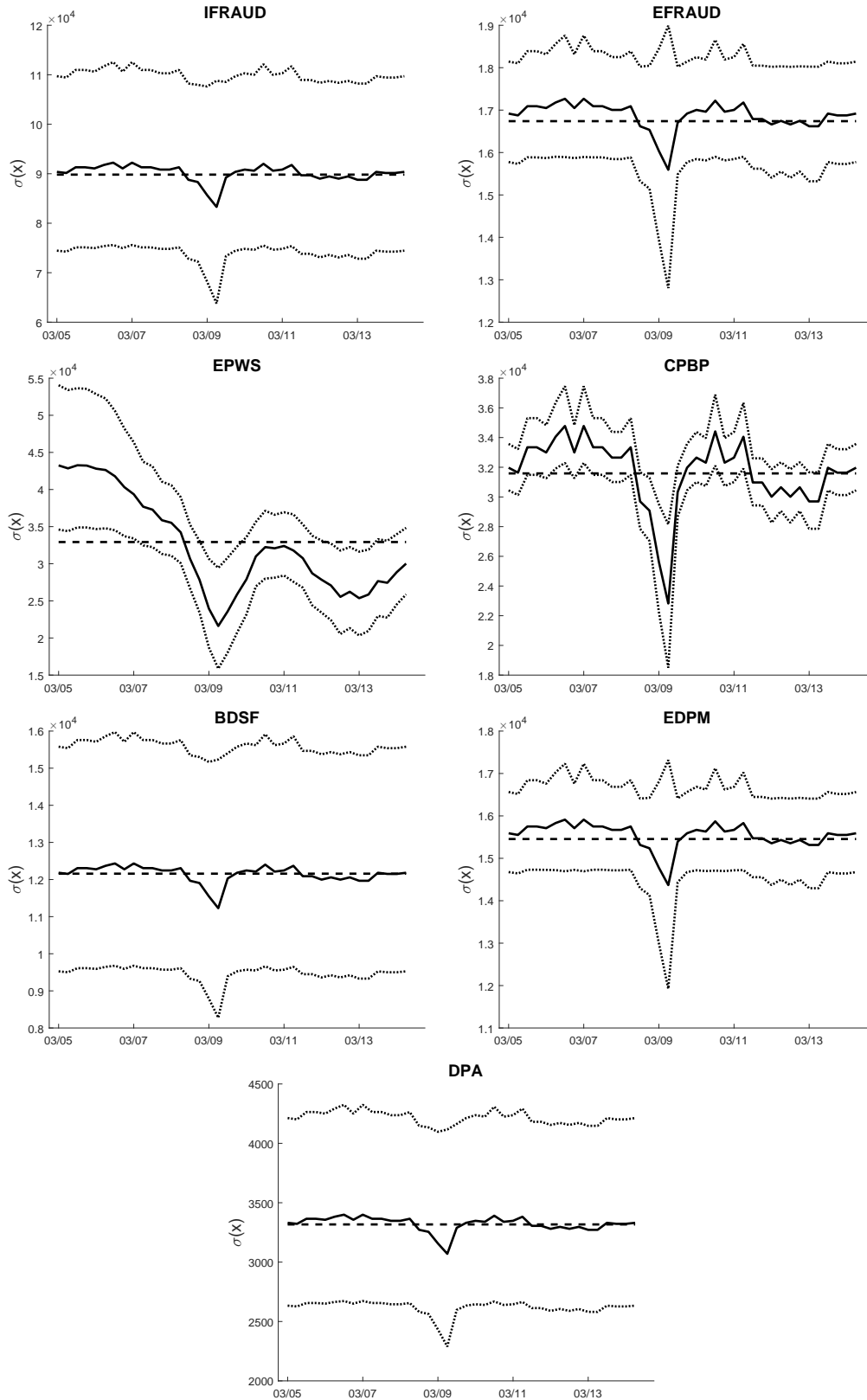


Figure 8: Solid: estimated scale parameter σ over time and across categories, for Model 3 with penalization approach and re-estimation step. Dotted: 95% confidence intervals. Dashed: estimation from Model 1.

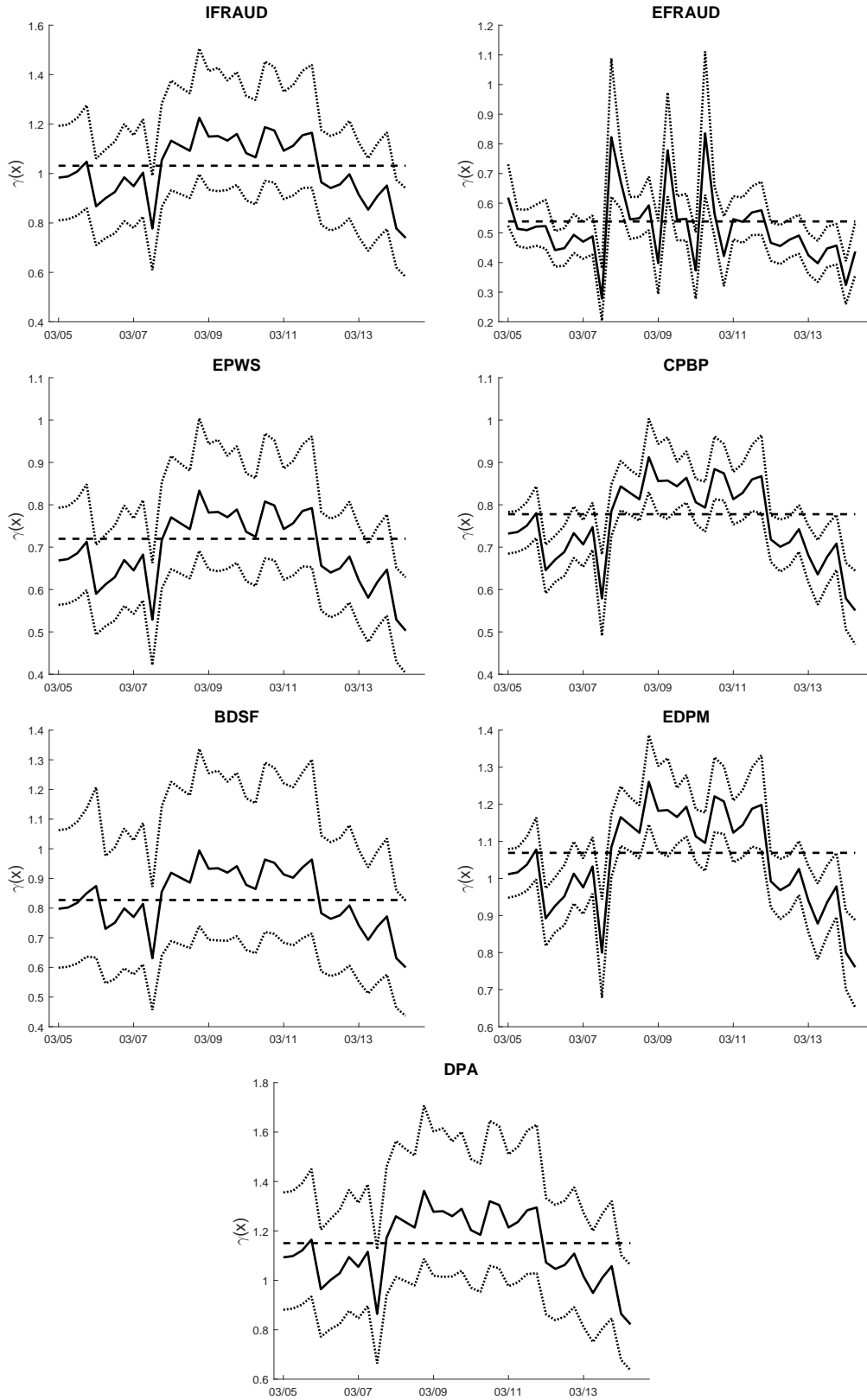


Figure 9: Solid: estimated shape parameter γ over time and across categories, for Model 3 with penalization approach and re-estimation step. Dotted: 95% confidence intervals. Dashed: estimation from Model 1.

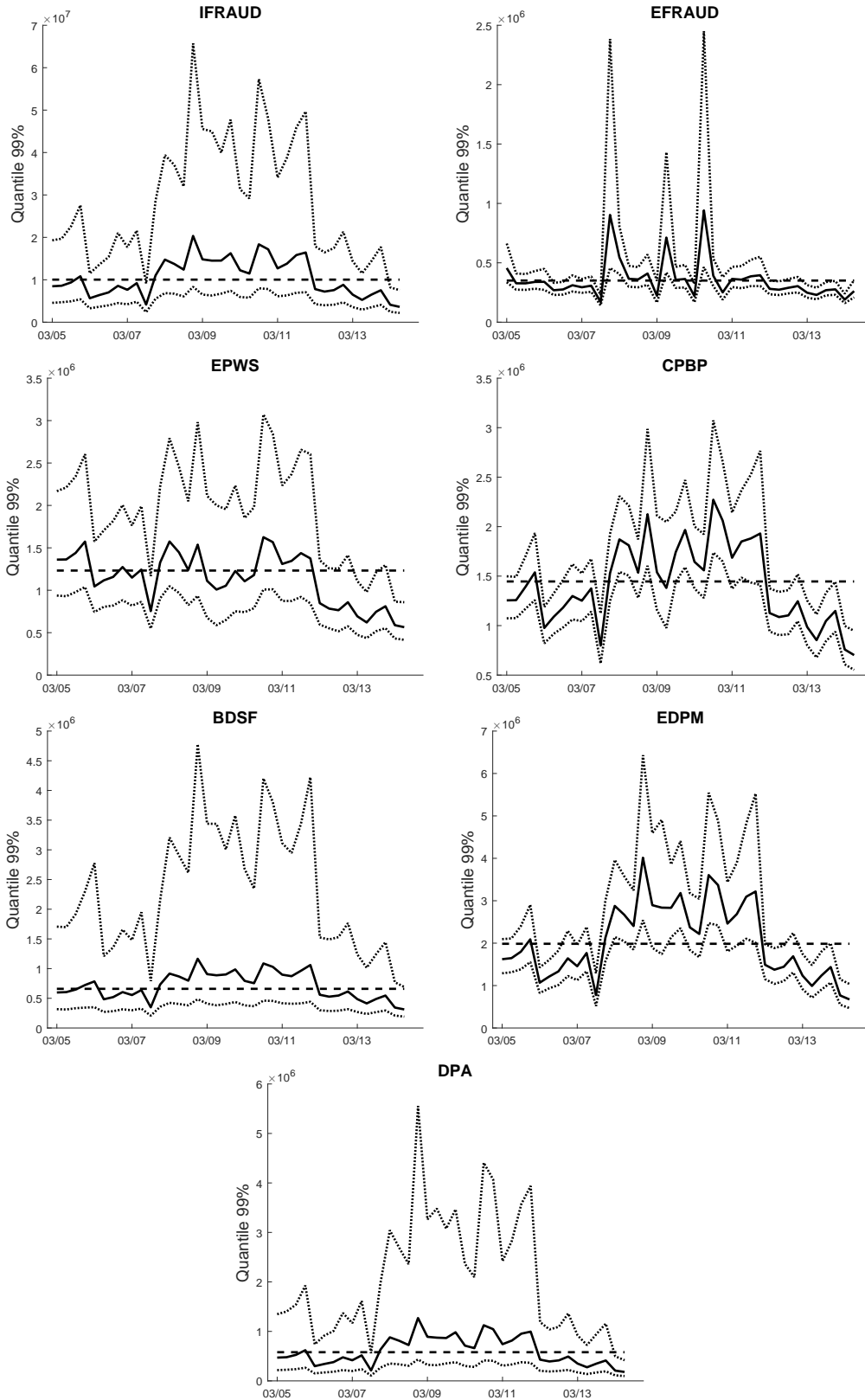


Figure 10: Solid: estimated quantile at level 99% of the severity distribution, over time, for Model 3 (penalized estimate based on BIC_2 with re-estimation step). Dotted: 95% confidence interval obtained with parametric bootstrap ($B = 5,000$). Dashed: estimated quantile for Model 1.

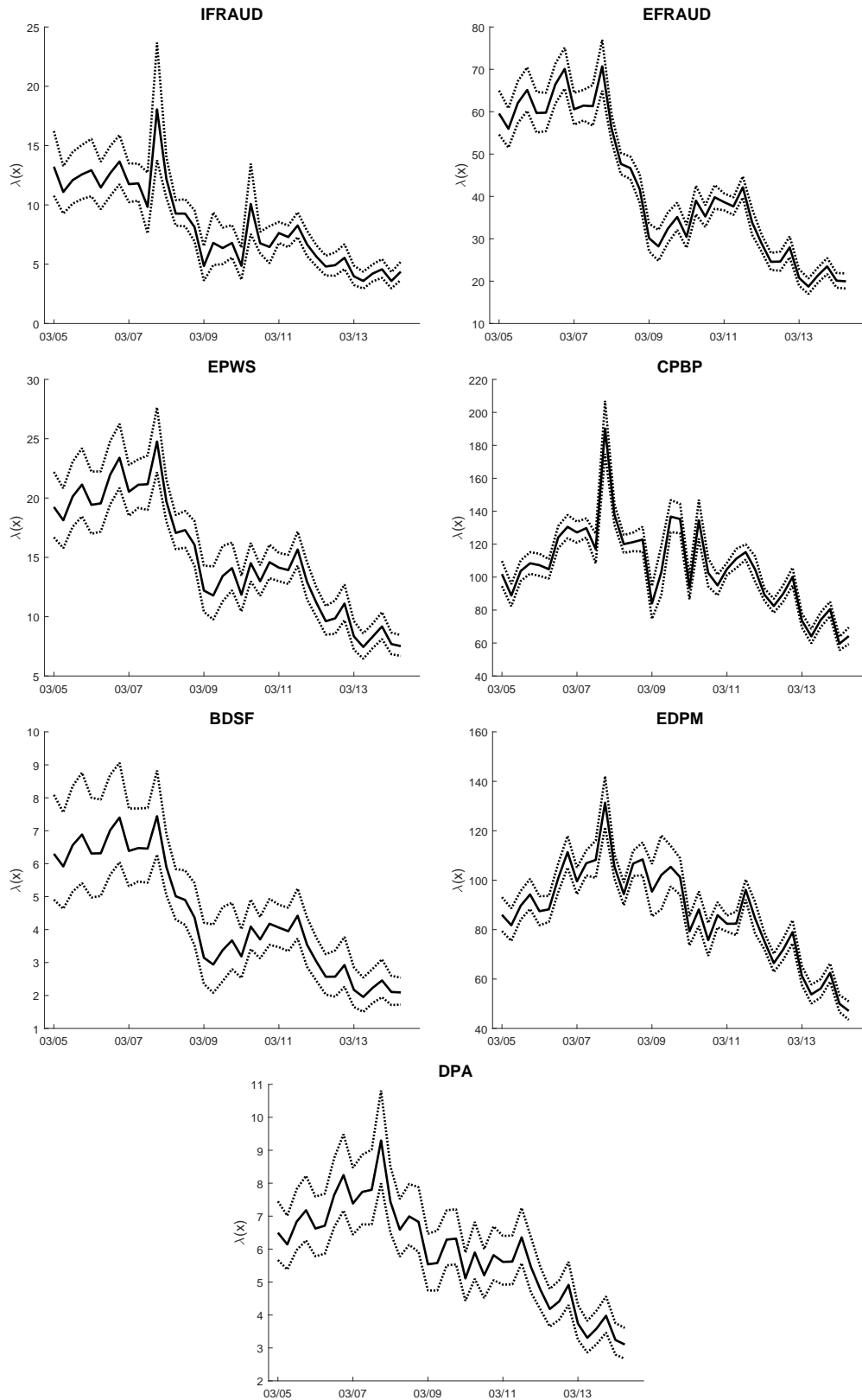


Figure 11: . Solid: Estimated frequency parameter $\lambda(x)$ over time, for Model 3. Dotted: 95% bootstrap confidence interval.

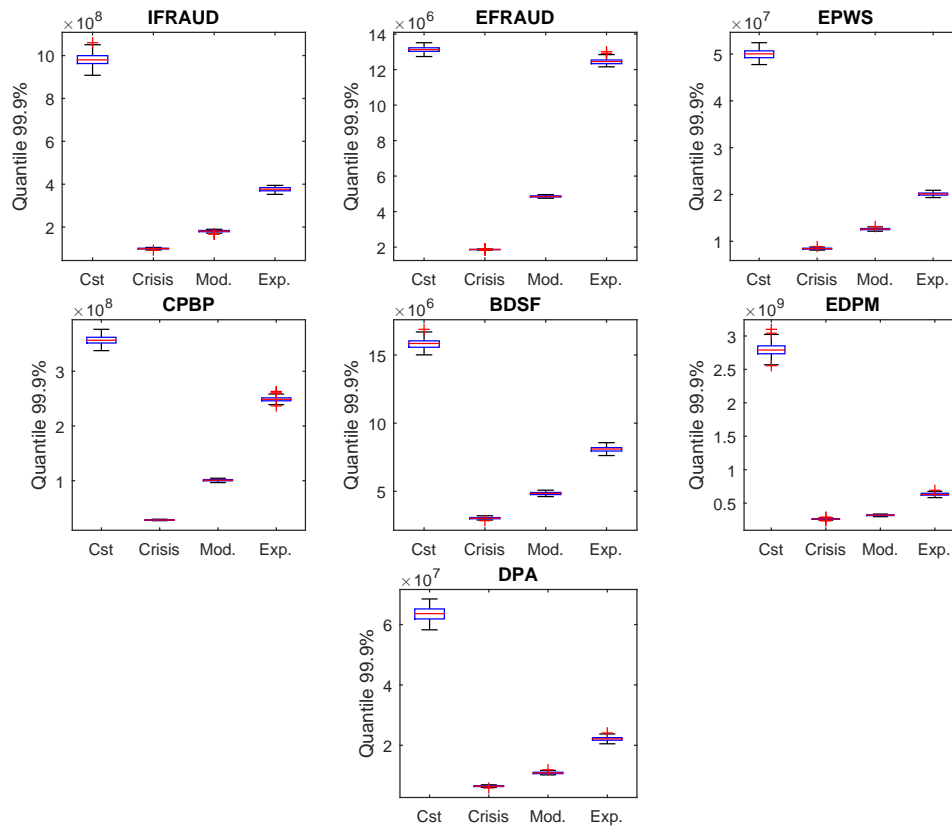


Figure 12: Boxplot of the Monte Carlo estimates of $\hat{Q}_{0.999}(L_t; \mathbf{x}_t)$, for the different scenarios and the different event types. From left to right: constant, crisis, moderate and expansion scenarios.

Appendix

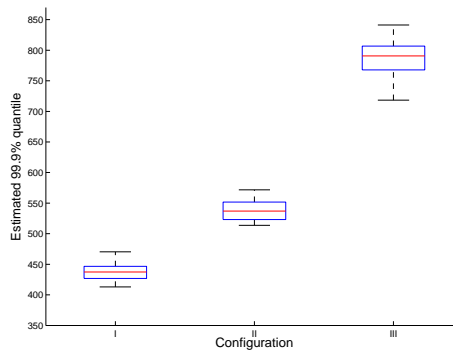


Figure 13: Boxplot of the quantile estimates for the three different configurations.

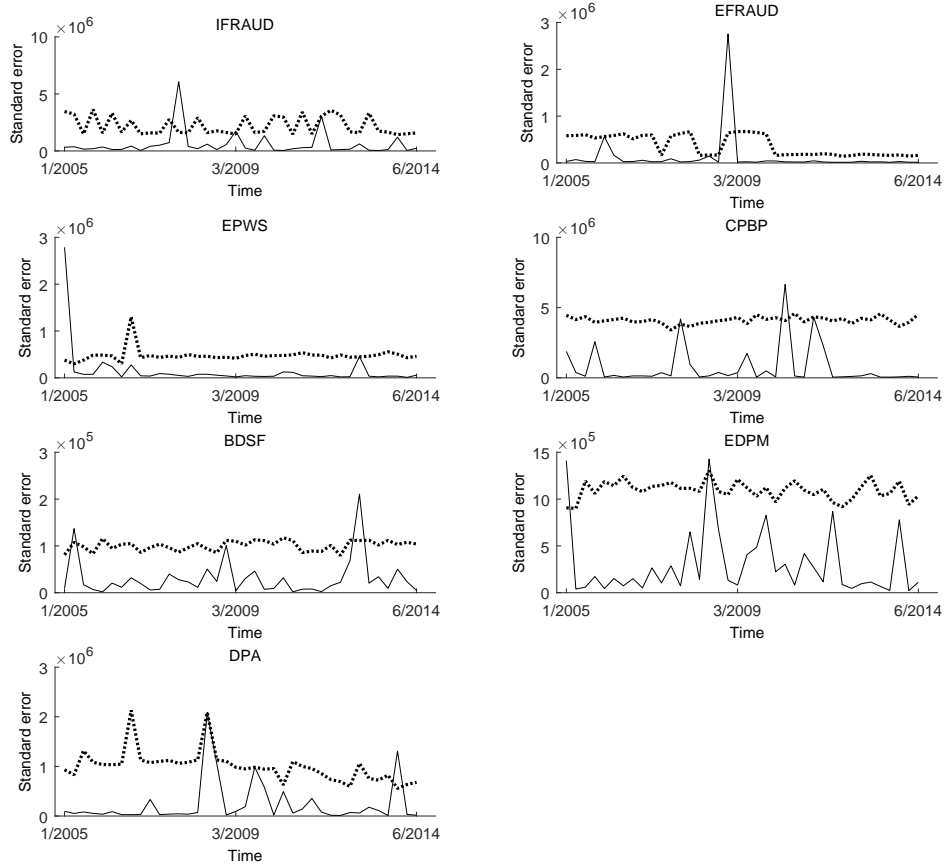


Figure 14: Solid line: standard deviation over time and per categories. Dotted: bootstrap 95% confidence interval under the null of a constant variance, obtained with $B = 2,000$.

Tables

Main text

Table 1: Abbreviations for the different event types.

Event type	Description
IFRAUD	Internal frauds
EFRAUD	External frauds, related to payments and others
EPWS	Employment practices and workplace safety
CPBP	Clients, products and business practices, related to derivatives, financial instruments and others
DPA	Damages to physical assets
BDSF	Business disruptions and system failures
EDPM	Execution, delivery and process management, related to financial instruments, payments and others

Table 2: Descriptive statistics of the losses (adjusted for inflation), per event type and pooled together. Quantities other than skewness and number of losses are in 1,000€. Upper panel: complete dataset. Lower panel: losses larger than the third quartile only.

No threshold						
ET	Mean	Std. Dev.	Median	3 rd quartile	Skewness	# losses
IFRAUD	151.83	1,143.28	13.64	53.03	21.28	1,271
EFRAUD	21.93	490.36	4.77	10.59	73.51	6,391
EPWS	39.01	343.46	6.99	19.48	37.57	2,292
CPBP	77.53	1,587.18	10.36	27.11	48.71	16,138
BDSF	15.73	42.99	4.83	11.2	7.66	674
EDPM	40.31	495.05	5.43	12.74	40.51	13,209
DPA	10.94	70.50	3.47	5.75	23.35	896
All	54.48	1,076.86	6.75	18.63	64.85	40,871
$\tau = Q(0.75)$						
ET	Mean	Std. Dev.	Median	3 rd quartile	Skewness	# losses
IFRAUD	568.05	223.7	143.48	350.62	10.83	318
EFRAUD	74.62	978.98	25.87	46.54	36.81	1,598
EPWS	136.66	678.02	49.23	100.41	19.14	573
CPBP	282.99	3,165.96	56.5	105.55	24.36	4,034
BDSF	49.06	77.09	21.51	51.58	3.99	168
EDPM	146.21	982.65	28.31	61.79	20.4	3,302
DPA	33.97	138.69	8.99	15.42	11.82	224
All	197.55	2,147.51	41.67	84.2	32.51	10,217

Table 3: Summary of the explanatory variables. See Appendix D for a full description.

Firm-specific	Macroeconomic	Financial
Event type (ET)	Italian unemployment rate (UR IT)	S&P 500 returns
Leverage ratio (LR)	EU unemployment rate (UR EU)	TR EU Stock Index returns (TRSI)
Tier-I capital ratio (TCR)	Italian GDP growth rate (GDP IT)	FTSE MIB index returns (MIB)
% revenue coming from fees (PRF)	EU GDP growth rate (GDP EU)	VIX
Deposit growth rate (DGR)	EU housing price growth rate (HPI)	VFTSE
UniCredit stock returns (SR)	Monetary aggregate M1 growth rate (M1)	3-month Italian interbank rate
	Consumer loans rate < 1 year in Italy (LOR IT)	10-year Italian government bond yield
	Consumer loans rate < 1 year in EU (LOR EU)	

Table 4: Ratio of Mean Squared error (RMSE) between penalized and unpenalized estimates, false positive rate (f.p.), true positive rate (t.p.) and correct classification rate (CCR). Predictor variables are i.i.d. $N(0,1)$ distributed, $p = 15$ and the penalty is either LASSO or adLASSO. The lines p -val. relate to f.p. and t.p. when the selection is performed with Wald tests at the 5% test level. CCR is defined as the number of active covariates selected plus the number of uninformative covariates not selected, divided by the total number of covariates.

iid $N(0,1)$ predictors, $p = 15$, $\alpha^\gamma = \{-.9; -.3; .2; .2\}$, $\alpha^\sigma = \{4; .6; .4; -.3\}$										
Sample size	Penalty	Selection method	RMSE	RMSE(+)	$f.p.(\sigma)$	$t.p.(\sigma)$	$f.p.(\gamma)$	$t.p.(\gamma)$	CCR	
$n = 1000$	none	p -val.	1	0.50	0.066	1	0.068	0.763	0.928	
	LASSO	AIC_1	0.37	0.89	0.571	1	0.556	0.972	0.584	
		AIC_2	0.42	0.75	0.36	1	0.311	0.892	0.744	
		BIC_1	0.42	0.63	0.254	1	0.184	0.87	0.827	
		BIC_2	0.65	0.56	0.071	1	0.014	0.465	0.915	
	adLASSO	AIC_1	0.42	0.82	0.351	1	0.382	0.953	0.727	
		AIC_2	0.41	0.66	0.158	1	0.224	0.878	0.848	
		BIC_1	0.36	0.5	0.064	1	0.082	0.807	0.928	
		BIC_2	0.43	0.47	0.011	1	0.017	0.592	0.949	
	$n = 2000$	none	p -val.	1	0.51	0.062	1	0.07	0.943	0.946
		LASSO	AIC_1	0.49	0.93	0.562	1	0.565	0.998	0.587
			AIC_2	0.54	0.82	0.349	1	0.385	0.988	0.73
BIC_1			0.61	0.69	0.235	1	0.197	0.98	0.84	
BIC_2			0.91	0.62	0.045	1	0.035	0.805	0.951	
adLASSO		AIC_1	0.47	0.83	0.323	1	0.339	0.993	0.757	
		AIC_2	0.47	0.68	0.136	1	0.202	0.977	0.874	
		BIC_1	0.46	0.50	0.047	1	0.06	0.943	0.955	
		BIC_2	0.54	0.52	0.006	1	0.021	0.828	0.973	
$n = 4000$		none	p -val.	1	0.48	0.055	1	0.061	1	0.957
		LASSO	AIC_1	0.545	0.93	0.542	1	0.525	1	0.609
			AIC_2	0.591	0.86	0.379	1	0.42	1	0.707
	BIC_1		0.71	0.70	0.238	1	0.189	1	0.843	
	BIC_2		1	0.47	0.047	1	0.055	0.99	0.962	
	adLASSO	AIC_1	0.51	0.74	0.243	1	0.282	1	0.808	
		AIC_2	0.48	0.62	0.115	1	0.186	1	0.89	
		BIC_1	0.41	0.44	0.041	1	0.054	0.998	0.965	
		BIC_2	0.46	0.34	0.001	1	0.015	0.988	0.993	

Table 5: RMSE, f.p. rate, t.p. rate and CCR when predictor variables are i.i.d. $N(0,1)$ distributed, $p = 25$ and the penalty is either LASSO or adLASSO.

iid $N(0,1)$ predictors, $p = 25$, $\alpha^\gamma = \{-.9; -.3; .2; .2\}$, $\alpha^\sigma = \{4; .6; .4; -.3\}$										
Sample size	Penalty	Selection method	RMSE	RMSE(+)	$f.p.(\sigma)$	$t.p.(\sigma)$	$f.p.(\gamma)$	$t.p.(\gamma)$	CCR	
$n = 1000$	none	p -val.	1	0.4	0.079	1	0.081	0.765	0.919	
	LASSO	AIC_1	0.28	0.83	0.508	1	0.489	0.953	0.578	
		AIC_2	0.29	0.63	0.285	1	0.247	0.855	0.768	
		BIC_1	0.26	0.48	0.206	1	0.144	0.862	0.845	
		BIC_2	0.38	0.36	0.06	1	0.012	0.408	0.934	
	adLASSO	AIC_1	0.34	0.77	0.342	1	0.349	0.932	0.706	
		AIC_2	0.31	0.59	0.156	1	0.211	0.873	0.839	
		BIC_1	0.22	0.38	0.064	1	0.077	0.783	0.928	
		BIC_2	0.27	0.29	0.01	1	0.016	0.542	0.962	
	$n = 2000$	none	p -val.	1	0.4	0.071	1	0.063	0.937	0.94
		LASSO	AIC_1	0.34	0.89	0.505	1	0.498	0.998	0.579
			AIC_2	0.36	0.70	0.301	1	0.266	0.987	0.762
BIC_1			0.38	0.56	0.191	1	0.15	0.978	0.855	
BIC_2			0.66	0.43	0.03	1	0.012	0.718	0.965	
adLASSO		AIC_1	0.37	0.79	0.308	1	0.34	0.997	0.728	
		AIC_2	0.32	0.57	0.126	1	0.182	0.983	0.87	
		BIC_1	0.28	0.38	0.045	1	0.065	0.955	0.951	
		BIC_2	0.39	0.36	0.003	1	0.012	0.763	0.979	
$n = 4000$		none	p -val.	1	0.38	0.057	1	0.052	0.995	0.954
		LASSO	AIC_1	0.40	0.91	0.509	1	0.504	1	0.575
			AIC_2	0.45	0.73	0.29	1	0.299	0.998	0.753
	BIC_1		0.49	0.57	0.172	1	0.143	1	0.868	
	BIC_2		0.75	0.28	0.024	1	0.018	0.982	0.982	
	adLASSO	AIC_1	0.41	0.69	0.211	1	0.263	1	0.8	
		AIC_2	0.35	0.56	0.1	1	0.18	1	0.882	
		BIC_1	0.26	0.37	0.035	1	0.057	1	0.961	
		BIC_2	0.33	0.25	0.003	1	0.01	0.972	0.993	

Table 6: RMSE, f.p. rate, t.p. rate and CCR when predictor variables are i.i.d. $N(0,1)$ distributed, $p = 50$ and the penalty is either LASSO or adLASSO.

iid $N(0,1)$ predictors, $p = 50$, $\alpha^\gamma = \{-.9; -.3; .2; .2\}$, $\alpha^\sigma = \{4; .6; .4; -.3\}$										
Sample size	Penalty	Selection method	RMSE	RMSE(+)	$f.p.(\sigma)$	$t.p.(\sigma)$	$f.p.(\gamma)$	$t.p.(\gamma)$	CCR	
$n = 1000$	none	p -val.	1	0.25	0.116	1	0.131	0.752	0.879	
	LASSO	AIC_1	0.06	0.3	0.483	1	0.439	0.952	0.574	
		AIC_2	0.04	0.13	0.195	1	0.143	0.765	0.837	
		BIC_1	0.04	0.11	0.155	1	0.112	0.81	0.871	
		BIC_2	0.05	0.05	0.055	1	0.004	0.353	0.954	
	adLASSO	AIC_1	0.09	0.32	0.343	1	0.332	0.902	0.687	
		AIC_2	0.06	0.18	0.161	1	0.166	0.81	0.844	
		BIC_1	0.03	0.08	0.06	1	0.065	0.738	0.934	
		BIC_2	0.04	0.05	0.006	1	0.023	0.59	0.975	
	$n = 2000$	none	p -val.	1	0.29	0.074	1	0.071	0.927	0.931
		LASSO	AIC_1	0.21	0.79	0.468	1	0.449	0.993	0.578
			AIC_2	0.18	0.44	0.195	1	0.165	0.945	0.833
BIC_1			0.17	0.37	0.144	1	0.111	0.955	0.882	
BIC_2			0.29	0.19	0.012	1	0.008	0.687	0.981	
adLASSO		AIC_1	0.27	0.71	0.306	1	0.328	0.982	0.708	
		AIC_2	0.18	0.45	0.132	1	0.166	0.967	0.862	
		BIC_1	0.12	0.24	0.044	1	0.055	0.937	0.953	
		BIC_2	0.17	0.16	0.002	1	0.006	0.72	0.988	
$n = 4000$		none	p -val.	1	0.31	0.066	1	0.059	0.992	0.942
		LASSO	AIC_1	0.27	0.86	0.453	1	0.489	1	0.567
			AIC_2	0.24	0.54	0.222	1	0.181	0.995	0.815
	BIC_1		0.24	0.43	0.125	1	0.115	0.995	0.889	
	BIC_2		0.4	0.15	0.01	1	0.011	0.972	0.989	
	adLASSO	AIC_1	0.23	0.57	0.238	1	0.183	0.997	0.806	
		AIC_2	0.19	0.43	0.107	1	0.12	0.993	0.895	
		BIC_1	0.13	0.28	0.04	1	0.052	0.992	0.958	
		BIC_2	0.17	0.13	0.001	1	0.005	0.96	0.996	

Table 7: RMSE, f.p, t.p. and CCR when predictor variables are collinear, follow either a t-distribution with 5 degrees of freedom (T(5)), a Gaussian AR(1) process with parameter .7 and no constant, a mix between AR(1) and a Gaussian random walk (RW_1) or a Gaussian random walk only (RW_2). $n = 4000$ and $p = 50$. The penalty is either LASSO or adLASSO.

Alternative models, $n = 4000, p = 50, \alpha^\gamma = \{-.9; -.3; .2; .2\}, \alpha^\sigma = \{4; .6; .4; -.3\}$										
Model	Penalty	Selection method	RMSE	RMSE(+)	$f.p.(\sigma)$	$t.p.(\sigma)$	$f.p.(\gamma)$	$t.p.(\gamma)$	CCR	
Collinear	none	p -val.	1	0.36	0.120	1	0.116	0.995	0.891	
	LASSO	AIC_1	0.25	0.66	0.453	1	0.436	0.998	0.589	
		AIC_2	0.28	0.45	0.272	1	0.217	0.977	0.774	
		BIC_1	0.35	0.35	0.14	1	0.127	0.945	0.876	
		BIC_2	0.66	0.42	0.016	1	0.010	0.593	0.976	
	adLASSO	AIC_1	0.2	0.47	0.194	1	0.196	0.997	0.821	
		AIC_2	0.18	0.35	0.109	1	0.13	0.985	0.89	
		BIC_1	0.17	0.2	0.031	1	0.045	0.977	0.964	
		BIC_2	0.25	0.18	0.001	1	0.006	0.835	0.992	
	T(5)	none	p -val.	1	0.31	0.064	1	0.060	1	0.943
		LASSO	AIC_1	0.27	0.86	0.451	1	0.470	1	0.576
			AIC_2	0.24	0.53	0.207	1	0.183	1	0.821
BIC_1			0.23	0.42	0.121	1	0.108	1	0.895	
BIC_2			0.37	0.16	0.009	1	0.013	0.972	0.989	
adLASSO		AIC_1	0.24	0.57	0.217	1	0.192	1	0.812	
		AIC_2	0.18	0.42	0.094	1	0.114	1	0.904	
		BIC_1	0.13	0.28	0.038	1	0.051	0.997	0.959	
		BIC_2	0.17	0.13	0.001	1	0.007	0.97	0.995	
AR(1)		none	p -val.	1	0.32	0.062	1	0.061	0.993	0.943
		LASSO	AIC_1	0.27	0.84	0.435	1	0.443	1	0.596
			AIC_2	0.27	0.53	0.203	1	0.178	0.995	0.825
	BIC_1		0.27	0.44	0.124	1	0.116	0.997	0.890	
	BIC_2		0.45	0.17	0.012	1	0.010	0.953	0.989	
	adLASSO	AIC_1	0.24	0.56	0.208	1	0.182	0.997	0.821	
		AIC_2	0.21	0.44	0.098	1	0.123	0.993	0.898	
		BIC_1	0.15	0.29	0.038	1	0.053	0.992	0.958	
		BIC_2	0.2	0.14	0.001	1	0.005	0.952	0.996	
	RW_1	none	p -val.	1	0.96	0.710	1	0.710	0.997	0.347
		LASSO	AIC_1	0.01	0.075	0.315	1	0.235	1	0.747
			AIC_2	0.01	0.02	0.103	1	0.094	1	0.909
BIC_1			0.01	0.01	0.060	1	0.049	1	0.950	
BIC_2			0.01	0.00	0.012	1	0.008	0.997	0.991	
adLASSO		AIC_1	0.08	0.25	0.262	1	0.259	1	0.76	
		AIC_2	0.03	0.1	0.09	1	0.146	0.997	0.891	
		BIC_1	0.01	0.01	0.02	1	0.053	0.987	0.966	
		BIC_2	0.01	0.01	0.006	1	0.025	0.917	0.983	
RW_2		none	p -val.	1	0.96	0.717	1	0.729	0.847	0.330
		LASSO	AIC_1	0.04	0.1	0.364	1	0.237	0.730	0.715
			AIC_2	0.03	0.06	0.203	0.982	0.148	0.647	0.827
	BIC_1		0.04	0.04	0.150	0.963	0.105	0.613	0.870	
	BIC_2		0.04	0.04	0.109	0.938	0.040	0.380	0.911	
	adLASSO	AIC_1	0.122	0.3	0.265	0.988	0.279	0.652	0.739	
		AIC_2	0.09	0.2	0.145	0.982	0.193	0.6	0.832	
		BIC_1	0.06	0.05	0.05	0.975	0.078	0.488	0.925	
		BIC_2	0.06	0.04	0.021	0.963	0.039	0.353	0.952	

Table 8: RMSE between penalized and unpenalized estimates, f.p., t.p. and CCR when the penalty is either LASSO or adLASSO, for a mix of time series models, with $T = 50$ and $n_t \in \{80; 200\}$, leading to final sample size $n \in \{4000; 10000\}$. See supplementary materials for more details.

Alternative model (mixed), $p = 50$. $\alpha^\gamma = \{-.9; -.3; .2; .2\}$, $\alpha^\sigma = \{4; .6; .4; -.3\}$										
Sample size	Penalty	Selection method	RMSE	RMSE(+)	f.p.(σ)	t.p.(σ)	f.p.(γ)	t.p.(γ)	CCR	
$n = 4000$	none	<i>p</i> -val.	1	1.058	0.873	0.97	0.91	0.925	0.177	
		LASSO	AIC_1	0.004	0.009	0.34	1	0.312	0.875	0.693
			AIC_2	0.003	0.006	0.216	1	0.209	0.832	0.8
			BIC_1	0.003	0.004	0.142	1	0.145	0.783	0.862
			BIC_2	0.005	0.004	0.059	0.998	0.05	0.465	0.934
		adLASSO	AIC_1	0.009	0.019	0.321	0.975	0.319	0.695	0.695
			AIC_2	0.009	0.015	0.23	0.968	0.223	0.618	0.779
			BIC_1	0.006	0.007	0.131	0.942	0.13	0.508	0.864
			BIC_2	0.007	0.007	0.086	0.925	0.096	0.412	0.897
	$n = 10000$	none	<i>p</i> -val.	1	1.051	0.868	0.99	0.881	0.918	0.193
			LASSO	AIC_1	0.006	0.011	0.357	1	0.238	0.988
				AIC_2	0.006	0.009	0.231	1	0.205	0.985
			BIC_1	0.007	0.006	0.139	1	0.157	0.962	0.863
			BIC_2	0.009	0.006	0.072	1	0.099	0.848	0.917
		adLASSO	AIC_1	0.023	0.047	0.266	0.988	0.355	0.848	0.71
			AIC_2	0.028	0.048	0.173	0.988	0.304	0.837	0.775
			BIC_1	0.014	0.015	0.084	0.982	0.139	0.663	0.887
			BIC_2	0.015	0.016	0.053	0.977	0.096	0.573	0.918

Table 9: Results of the regularized regressions for Model 2, using LASSO penalties. Selection of the penalty parameters is performed over a grid ($[0.005; 0.01]$), using BIC_2 as criterion. Coefficients are rounded to the third decimal; those that equal zero after rounding are treated as exactly zero (see Footnote 21) and are marked by a “-”. $\hat{\alpha}^\sigma$ denotes the penalized estimates, whereas $\hat{\alpha}_\perp^\sigma$ denotes the solution obtained with the re-estimation step.

Pen.	Model 2, LASSO, BIC_2 ($\nu = (0.008; 0.0054)$)					
Covariate	$\hat{\alpha}^\sigma$	$\hat{\alpha}_\perp^\sigma$	(<i>p</i> -value)	$\hat{\alpha}^\gamma$	$\hat{\alpha}_\perp^\gamma$	(<i>p</i> -value)
(Intercept)	10.001	10.004	(0.00)	-0.199	-0.206	(0.00)
IFRAUD	0.573	0.572	(0.00)	-0.019	-0.019	(0.28)
EFRAUD	0.587	0.585	(0.00)	-0.274	-0.271	(0.00)
EPWS	0.528	0.527	(0.00)	-0.107	-0.106	(0.00)
CPBP	1.103	1.104	(0.00)	-0.195	-0.197	(0.00)
BDSF	0.165	0.163	(0.00)	-0.041	-0.038	(0.04)
EDPM	0.722	0.723	(0.00)	-0.037	-0.037	(0.03)
GDP EU	0.021	0.041	(0.01)	-	-	-
PRF	0.005	0.023	(0.16)	-	-	-
Unemp. IT	-	-	-	-0.053	-0.106	(0.00)
VIX	-	-	-	0.027	0.041	(0.02)
LR	-	-	-	-0.044	-0.105	(0.00)

Table 10: Results of the regularized regressions for Model 3, using LASSO penalty. Selection of the penalty parameters is performed over a grid using BIC_2 as criterion.

Model 3, LASSO, BIC_2 ($\nu = (0.0101, 0.0097)$)						
Covariate	$\hat{\alpha}^\sigma$	$\hat{\alpha}_+^\sigma$	(p-value)	$\hat{\alpha}^\gamma$	$\hat{\alpha}_+^\gamma$	(p-value)
(Intercept)	10.001	10.006	(0.00)	-0.198	-0.214	(0.00)
IFRAUD	0.574	0.574	(0.00)	-0.020	-0.019	(0.28)
EFRAUD	0.590	0.591	(0.00)	-0.283	-0.302	(0.00)
EPWS	0.522	0.502	(0.00)	-0.111	-0.113	(0.00)
CPBP	1.100	1.095	(0.00)	-0.195	-0.197	(0.00)
BDSF	0.166	0.165	(0.00)	-0.042	-0.040	(0.03)
EDPM	0.722	0.723	(0.00)	-0.036	-0.037	(0.04)
GDP EU	0.002	0.0192	(0.24)	-	-	-
Unemp. IT	-	-	-	-0.023	-0.105	(0.00)
VIX	-	-	-	0.023	0.041	(0.02)
LR	-	-	-	-0.014	-0.104	(0.00)
GDP EU x CPBP	0.0172	0.0394	(0.01)	-	-	-
HPI x EPWS	0.0137	0.0524	(0.00)	-	-	-
DGR x EFRAUD	-	-	-	0.019	0.0724	(0.00)

Table 11: Results of the regularized regressions for Model 3, using LASSO penalty. Selection of the penalty parameters is performed over a grid using BIC_1 as criterion.

Model 3, LASSO, BIC_1 ($\nu = (0.0109, 0.008)$)						
Covariate	$\hat{\alpha}^\sigma$	$\hat{\alpha}_+^\sigma$	(p-value)	$\hat{\alpha}^\gamma$	$\hat{\alpha}_+^\gamma$	(p-value)
(Intercept)	10.001	10.009	(0.00)	-0.200	-0.227	(0.00)
IFRAUD	0.574	0.576	(0.00)	-0.020	-0.029	(0.09)
EFRAUD	0.590	0.595	(0.00)	-0.286	-0.410	(0.00)
EPWS	0.525	0.468	(0.00)	-0.118	-0.152	(0.00)
CPBP	1.100	1.094	(0.00)	-0.195	-0.194	(0.00)
BDSF	0.166	0.166	(0.00)	-0.039	-0.022	(0.26)
EDPM	0.722	0.722	(0.00)	-0.036	-0.032	(0.07)
GDP EU	0.002	0.019	(0.23)	-	-	-
Unemp. IT	-	-	-	-0.037	-0.092	(0.00)
VIX	-	-	-	0.023	0.131	(0.00)
VFTSE	-	-	-	0.002	-0.081	(0.00)
LR	-	-	-	-0.029	-0.103	(0.00)
GDP EU x EFRAUD	-	-	-	-0.002	-0.044	(0.04)
GDP EU x CPBP	0.014	0.042	(0.01)	-	-	-
GDP IT x EDPM	-	-	-	-0.006	0.027	(0.12)
HPI x EPWS	0.004	-0.007	(0.66)	0.014	0.049	(0.01)
HPI x BDSF	-	-	-	-0.006	-0.048	(0.03)
ST rates x EFRAUD	-	-	-	0.001	0.141	(0.00)
ST rates x EPWS	0.014	0.081	(0.00)	-	-	-
DGR x IFRAUD	-	-	-	0.003	0.025	(0.13)
DGR x EFRAUD	-	-	-	0.027	0.053	(0.02)
MIB x EPWS	-	-	-	0.002	0.056	(0.02)

Table 12: $CV(LLF)$ consists in the sum of the cross-validated predicted negative log-likelihood, obtained from using sequentially the losses from one time period as validation set and the others as training set. $CV(CLS)$ denotes the cross-validated CLS using empirical 99% quantiles in each event type for κ .

	<u>p-val.</u>			<u>LASSO (BIC_2)</u>		<u>LASSO (BIC_1)</u>	
	Model 1	Model 2	Model 3	Model 2	Model 3	Model 2	Model 3
DF	14	42	282	19	21	20	30
LLF	121,050.7	120,989.6	120,825.0	121,029.2	121,020.4	121,028.7	121,004.6
AIC	242,129.4	242,063.3	242,214.1	242,096.4	242,082.7	242,097.4	242,069.2
BIC	242,230.7	242,367.0	244,253.5	242,233.9	242,234.6	242,242.1	242,286.1
$CV(LLF)$	121,073.4	121,072.5	122,463.8	121,070.7	121,061.2	121,073.3	121,072.9
$CV(CLS(0.99))$	781.7	777.2	1,097.2	775.8	772.1	777.0	779.5

Table 13: Results of the regularized regressions for λ , using the variables selected with LASSO for Model 3. $CV(LLF)$ has been obtained using the same partitioning procedure as for the analysis of the severity distribution.

Covariate	Model 1		Model 2 ($\nu = 0.0159$)		Model 3 ($\nu = 0.2857$)	
	$\hat{\lambda}_+$	(p -value)	$\hat{\lambda}_+$	(p -value)	$\hat{\lambda}_+$	(p -value)
(Intercept)	2.996	(0.00)	2.962	(0.00)	2.938	(0.00)
IFRAUD	0.123	(0.00)	0.123	(0.00)	0.073	(0.00)
EFRAUD	0.688	(0.00)	0.689	(0.00)	0.646	(0.00)
EPWS	0.329	(0.00)	0.329	(0.00)	0.306	(0.00)
CPBP	1.013	(0.00)	1.014	(0.00)	1.028	(0.00)
BDSF	-0.102	(0.00)	-0.101	(0.00)	-0.145	(0.00)
EDPM	0.943	(0.00)	0.943	(0.00)	0.959	(0.00)
GDP EU	-	-	-	-	-	-
Unemp. IT	-	-	-0.302	(0.00)	-0.259	(0.00)
HPI	-	-	-0.040	(0.00)	-	-
VIX	-	-	-0.085	(0.00)	-0.045	(0.00)
DGR	-	-	0.065	(0.00)	0.038	(0.00)
LR	-	-	-0.023	(0.01)	-	-
GDP EU x CPBP	-	-	-	-	0.025	(0.00)
GDP EU x EDPM	-	-	-	-	-0.006	(0.16)
HPI x IFRAUD	-	-	-	-	0.041	(0.00)
HPI x EFRAUD	-	-	-	-	0.072	(0.00)
HPI x EPWS	-	-	-	-	0.026	(0.00)
HPI x CPBP	-	-	-	-	-0.046	(0.00)
HPI x BDSF	-	-	-	-	0.007	(0.00)
HPI x EDPM	-	-	-	-	-0.020	(0.00)
DGR x IFRAUD	-	-	-	-	0.014	(0.04)
DGR x CPBP	-	-	-	-	0.023	(0.00)
AIC	3,037.08		2,405.45		2,278.07	
BIC	3,062.16		2,448.45		2,349.74	
$CV(LLF)$	1,567.15		1,312.98		1,256.84	

Table 14: Parameters of the total loss distribution for the different scenarios.

	Constant			Crisis scenario			Moderate scenario			Expansion scenario		
	λ_1	γ_1	σ_1	λ_3^+	γ_3^+	σ_3^+	λ_3^+	γ_3^+	σ_3^+	λ_3^+	γ_3^+	σ_3^+
IFRAUD	8.38	1.032	89,863	2.27	0.91	80,910	6.12	0.85	93,133	10.14	0.88	98,022
EFRAUD	42.06	0.538	16,744	12.80	0.37	15,151	30.41	0.43	17,440	45.76	0.53	18,355
EPWS	15.08	0.72	32,941	4.86	0.62	26,289	11.54	0.58	30,354	17.34	0.60	32,095
CPBP	106.16	0.778	31,583	22.62	0.67	20,336	109.91	0.63	36,274	220.31	0.65	44,770
BDSF	4.42	0.828	12,167	1.34	0.74	10,902	3.19	0.69	12,549	4.80	0.71	13,208
EDPM	86.89	1.07	15,450	35.38	0.93	13,966	70.93	0.87	16,076	99.89	0.90	16,920
DPA	5.88	1.15	3,331	2.03	1.01	2,975	4.82	0.94	3,424	7.22	0.98	3,604

Appendix

Table 15: Results of the KS tests. The rejection regions have been computed with $B = 2,000$. The second row indicate the proportion of significant differences, among all comparisons.

% rejected.	IFRAUD	EFRAUD	EPWS	CPBP	BDSF	EDPM	DPA	All
τ_0	1.99 %	8.68%	7.82%	17.78%	0.14%	4.41%	0%	30.4%
$\tau_{.5}$	-	-	-	-	-	-	-	3.13%
$\tau_{.75}$	-	-	-	-	-	-	-	1.14%
$\tau_{.85}$	-	-	-	-	-	-	-	0.43%

Table 16: Likelihood ratio test statistics (LRT). The bootstrap bound of the rejection region at the 95% confidence level (denoted $Q_{1-\alpha}^*$) and obtained with $B = 10,000$ is 52.58. * indicates a rejection of H_0 at the 5% test level. When we don't have any extreme during one of the time periods, the period is excluded from the test.

τ	IFRAUD	EFRAUD	EPWS	CPBP	BDSF	EDPM	DPA	All
τ_0	137.33*	726.87*	269.4*	1,778.5*	318.6*	931.39*	108.12*	2,713.9*
$\tau_{.5}$	150.22*	569.83*	238.98*	1,074.99*	184.01*	610.73*	97.36*	2,092.5*
$\tau_{.75}$	138.65*	292.86*	220.46*	579.11*	93.69*	358.88*	67.37*	1,197.7*
$\tau_{.85}$	104.14*	195.86*	173.58*	427.4*	55.64*	284.35*	47.67	910.6*

Table 17: Results of the unpenalized estimation for Model 1 and 2.

Covariate	Model 1				Model 2			
	$\hat{\alpha}^\sigma$	(p-value)	$\hat{\alpha}^\gamma$	(p-value)	$\hat{\alpha}^\sigma$	(p-value)	$\hat{\alpha}^\gamma$	(p-value)
(Intercept)	10.000	(0.00)	-0.19	(0.00)	10.010	(0.00)	-0.223	(0.00)
IFRAUD	0.573	(0.00)	-0.02	(0.26)	0.569	(0.00)	-0.018	(0.29)
EFRAUD	0.589	(0.00)	-0.28	(0.00)	0.586	(0.00)	-0.276	(0.00)
EPWS	0.529	(0.00)	-0.11	(0.00)	0.526	(0.00)	-0.111	(0.00)
CPBP	1.103	(0.00)	-0.19	(0.00)	1.103	(0.00)	-0.200	(0.00)
BDSF	0.165	(0.00)	-0.04	(0.02)	0.162	(0.00)	-0.044	(0.02)
EDPM	0.721	(0.00)	-0.03	(0.05)	0.725	(0.00)	-0.041	(0.02)
Unemp. EU	-	-	-	-	0.576	(0.00)	0.126	(0.00)
Unemp. IT	-	-	-	-	-0.266	(0.00)	-0.119	(0.00)
GDP EU	-	-	-	-	0.100	(0.00)	-0.104	(0.00)
GDP IT	-	-	-	-	-0.065	(0.00)	0.075	(0.00)
HPI	-	-	-	-	-0.018	(0.25)	0.069	(0.00)
M1	-	-	-	-	-0.122	(0.00)	0.026	(0.15)
LOR EU	-	-	-	-	0.129	(0.00)	0.141	(0.00)
LOR IT	-	-	-	-	0.136	(0.00)	-0.222	(0.00)
LT rates	-	-	-	-	-0.019	(0.25)	0.026	(0.16)
ST rates	-	-	-	-	0.169	(0.00)	0.029	(0.12)
Stock returns	-	-	-	-	-0.047	(0.00)	-0.065	(0.00)
TRSI	-	-	-	-	-0.047	(0.00)	0.172	(0.00)
S&P500	-	-	-	-	0.057	(0.00)	-0.119	(0.00)
VIX	-	-	-	-	-0.021	(0.19)	0.162	(0.00)
VFTSE	-	-	-	-	-0.054	(0.00)	-0.019	(0.28)
MIB	-	-	-	-	0.024	(0.14)	0.084	(0.00)
PRF	-	-	-	-	0.048	(0.00)	0.082	(0.00)
DGR	-	-	-	-	-0.016	(0.30)	0.032	(0.08)
TCR	-	-	-	-	0.020	(0.19)	-0.058	(0.00)
LR	-	-	-	-	-0.059	(0.00)	-0.062	(0.00)

Table 18: Results of the unpenalized estimation for Model 3, for σ

Model 3, unpenalized, σ										
	(Intercept)	IFRAUD	EFRAUD	EPWS	CPBP	BDSF	EDPM			
(Intercept)	10.03	0.48	-0.03	2.70	2.57	-0.38	-0.61	(0.00)	(0.00)	(0.00)
Unemp. EU	0.8084	0.48	-0.251	-1.390	-0.744	2.676	0.272	(0.00)	(0.00)	(0.00)
Unemp. IT	-0.2895	-0.14	-0.021	0.248	0.104	-0.948	-0.130	(0.00)	(0.00)	(0.00)
GDP EU	-0.7879	0.26	0.373	0.058	0.688	0.126	0.533	(0.00)	(0.00)	(0.00)
GDP IT	0.557	-0.16	-0.264	-0.094	-0.422	-0.093	-0.389	(0.00)	(0.00)	(0.00)
HPI	0.2915	0.48	-0.060	-0.108	-0.449	-0.202	-0.007	(0.00)	(0.00)	(0.68)
MI	0.2804	-0.36	-0.218	0.043	-0.525	-0.106	-0.701	(0.00)	(0.00)	(0.00)
LOR EU	-0.0674	4.22	1.599	-0.611	-0.787	-0.087	1.705	(0.00)	(0.00)	(0.00)
LOR IT	-0.6956	-1.35	0.519	0.699	2.663	0.810	1.809	(0.00)	(0.00)	(0.00)
LT rates	0.1653	-0.24	-0.212	-0.395	-0.776	-0.737	-0.324	(0.00)	(0.00)	(0.00)
ST rates	0.9674	-1.01	-0.447	0.086	-0.618	0.317	-1.002	(0.00)	(0.00)	(0.00)
Stock returns	-0.5457	0.12	0.144	0.164	0.298	-0.219	0.419	(0.00)	(0.00)	(0.00)
TRSI	0.0463	-0.08	0.090	-0.083	-0.145	0.188	-0.055	(0.00)	(0.00)	(0.00)
S&P500	-0.5214	0.17	0.153	0.120	0.406	-0.114	0.424	(0.00)	(0.00)	(0.00)
VIX	1.0354	-0.66	-0.902	-0.924	-1.191	-0.875	-0.951	(0.00)	(0.00)	(0.00)
VFTSE	-0.7853	0.62	0.708	0.809	0.799	0.550	0.554	(0.00)	(0.00)	(0.00)
MIB	1.2765	-0.27	-0.471	-0.236	-0.760	-0.059	-0.922	(0.00)	(0.00)	(0.00)
PRF	-0.0451	-0.43	0.477	-0.269	0.500	-0.790	0.654	(0.00)	(0.00)	(0.00)
DGR	0.1838	-0.09	-0.065	-0.060	-0.123	0.003	-0.123	(0.00)	(0.81)	(0.00)
TCR	0.0662	-0.81	0.187	0.204	-0.415	0.083	-0.115	(0.00)	(0.00)	(0.00)
LR	0.2425	-0.77	-0.990	-0.569	-0.688	-0.263	-0.834	(0.00)	(0.00)	(0.00)

Table 19: Results of the unpenalized estimation for Model 3, for γ

Model 3, unpenalized, γ														
	(Intercept)	IFRAUD	EFRAUD	EPWS	CPBP	BDSF	EDPM							
(Intercept)	-0.458	(0.00)	-1.642	(0.00)	4.324	(0.00)	2.818	(0.00)	-1.369	(0.00)	2.185	(0.00)	0.887	(0.00)
Unemp. EU	0.064	(0.00)	-0.941	(0.00)	-0.559	(0.00)	0.807	(0.00)	-0.523	(0.00)	6.216	(0.00)	0.026	(0.14)
Unemp. IT	-0.567	(0.00)	0.239	(0.00)	0.749	(0.00)	-0.939	(0.00)	1.377	(0.00)	-1.248	(0.00)	0.647	(0.00)
GDP EU	0.722	(0.00)	-0.043	(0.01)	-0.203	(0.00)	0.097	(0.00)	-0.675	(0.00)	0.089	(0.00)	-0.399	(0.00)
GDP IT	-1.086	(0.00)	0.273	(0.00)	0.398	(0.00)	0.037	(0.17)	0.834	(0.00)	0.465	(0.00)	0.646	(0.00)
HPI	0.089	(0.00)	-0.717	(0.00)	0.044	(0.05)	-0.192	(0.00)	0.339	(0.00)	-0.581	(0.00)	-0.137	(0.00)
MI	-0.235	(0.00)	-0.692	(0.00)	0.091	(0.00)	-0.236	(0.00)	0.390	(0.00)	-0.944	(0.00)	0.287	(0.00)
LOR EU	-0.005	(0.78)	-3.367	(0.00)	0.978	(0.00)	-1.152	(0.00)	1.828	(0.00)	2.294	(0.00)	0.206	(0.00)
LOR IT	0.986	(0.00)	2.658	(0.00)	-2.156	(0.00)	-0.938	(0.00)	-3.511	(0.00)	-1.207	(0.00)	-2.336	(0.00)
LT rates	0.399	(0.00)	0.720	(0.00)	-1.389	(0.00)	-2.002	(0.00)	-0.903	(0.00)	1.052	(0.00)	-1.279	(0.00)
ST rates	-0.317	(0.00)	-0.042	(0.02)	-0.225	(0.00)	-0.719	(0.00)	0.069	(0.00)	-1.895	(0.00)	0.385	(0.00)
Stock returns	0.816	(0.00)	0.104	(0.00)	-0.360	(0.00)	-0.495	(0.00)	-0.610	(0.00)	-0.040	(0.10)	-0.571	(0.00)
TRSI	-0.562	(0.00)	0.235	(0.00)	0.073	(0.00)	0.168	(0.00)	0.665	(0.00)	-0.400	(0.00)	0.415	(0.00)
S&P500	1.062	(0.00)	-0.403	(0.00)	-0.404	(0.00)	-0.327	(0.00)	-0.822	(0.00)	-1.053	(0.00)	-0.764	(0.00)
VIX	0.316	(0.00)	1.260	(0.00)	0.330	(0.00)	-0.469	(0.00)	-0.273	(0.00)	1.717	(0.00)	-0.417	(0.00)
VFTSE	-0.671	(0.00)	-1.531	(0.00)	-0.369	(0.00)	0.647	(0.00)	1.382	(0.00)	-2.743	(0.00)	0.881	(0.00)
MIB	-1.034	(0.00)	0.165	(0.00)	0.357	(0.00)	0.514	(0.00)	0.739	(0.00)	0.331	(0.00)	0.738	(0.00)
PRF	0.136	(0.00)	1.386	(0.00)	-2.015	(0.00)	0.817	(0.00)	-0.225	(0.00)	3.938	(0.00)	-0.407	(0.00)
DGR	-0.282	(0.00)	0.294	(0.00)	0.177	(0.00)	0.163	(0.00)	0.195	(0.00)	0.416	(0.00)	0.182	(0.00)
TCR	0.105	(0.00)	0.544	(0.00)	-1.688	(0.00)	-0.922	(0.00)	0.109	(0.00)	-8.968	(0.00)	0.022	(0.22)
LR	-0.656	(0.00)	1.663	(0.00)	1.477	(0.00)	1.851	(0.00)	1.680	(0.00)	-1.466	(0.00)	1.396	(0.00)

References

- F. Aue and M. Kalkbrener. LDA at work: Deutsche Bank's approach to quantifying operational risk. *Journal of Operational Risk*, 1(4):49–93, 2006.
- A. Balkema and L. de Haan. Residual life time at great age. *The Annals of Probability*, 2(5):792–804, 1974.
- Basel Committee on Banking Supervision (BCBS). Basel II: international convergence

- of capital measurement and capital standards. A revised framework. Technical report, Bank of International Settlements, Basel, Switzerland, 2004.
- Basel Committee on Banking Supervision (BCBS). Basel III: A global regulatory framework for more resilient banks and banking systems. Technical report, Bank of International Settlements, Basel, Switzerland, 2010.
- Basel Committee on Banking Supervision (BCBS). Standardised Measurement Approach for operational risk. Technical report, Bank of International Settlements, Basel, Switzerland, 2016.
- J. Beirlant and Y. Goegebeur. Local polynomial maximum likelihood estimation for the Pareto-type distributions. *Journal of Multivariate Analysis*, 89(1):97–118, 2004.
- J. Beirlant, Y. Goegebeur, J. Teugels, J. Segers, D. De Waal, and C. Ferro. *Statistics of extremes: Theory and applications*. Wiley, Chichester, 2005.
- G. Bekaert and M. Hoerova. The VIX, the variance premium and stock market volatility. *Journal of Econometrics*, 183(2):181–192, 2014.
- G. Bekaert, M. Hoerova, and Duca M.L. Risk, uncertainty and monetary policy. *Journal of Monetary Economics*, 60(7):771–788, 2013.
- S. Bhagat, B. Bolton, and J. Lu. Size, leverage, and risk-taking of financial institutions. *Journal of Banking & Finance*, 59:520–537, 2015.
- S. Boubaker, D. Gounopoulos, D.C. Nguyen, and N. Paltalidis. Assessing the effects of unconventional monetary policy and low interest rates on pension fund risk incentives. *Journal of Banking & Finance*, 77:35–52, 2017.
- E. Brechmann, C. Czado, and S. Paterlini. Flexible dependence modeling of operational risk losses and its impact on total capital requirements. *Journal of Banking & Finance*, 40:271–285, 2014.
- A. Chapelle, Y. Crama, G. Hübner, and J.-P. Peters. Practical methods for measuring and managing operational risk in the financial sector: a clinical study. *Journal of Banking & Finance*, 32(6):1049–1061, 2008.
- V. Chavez-Demoulin, P. Embrechts, and S. Sardy. Extreme-quantile tracking for financial time series. *Journal of Econometrics*, 181(1):44–52, 2014.
- V. Chavez-Demoulin, P. Embrechts, and M. Hofert. An extreme value approach for modeling Operational Risk losses depending on covariates. *Journal of Risk and Insurance*, 83(3):735–776, 2016.
- A. Chernobai, P. Jorion, and F. Yu. The derminants of Operational Risk in U.S. financial institutions. *Journal of Financial and Quantitative Analysis*, 46(8):1683–1725, 2011.

- E. Cope, M. Piche, and J. Walter. Macroenvironmental determinants of operational loss severity. *Journal of Banking & Finance*, 36(5):1362–1380, 2012.
- D. Creal, S.J. Koopman, and A. Lucas. Generalized Autoregressive Score Models with Applications. *Journal of Applied Econometrics*, 28(5):777–795, 2012.
- S. Das, O. Harel, D. Dey, J. Covault, and H. Kranzler. Analysis of extreme drinking in patients with alcohol dependence using Pareto regression. *Statistics in Medicine*, 29(11):1250–1258, 2010.
- A. Davison and R. Smith. Models for exceedances over high thresholds. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 52(3):393–442, 1990.
- M.D. Delis and G.P. Kouretas. Interest rates and bank risk-taking. *Journal of Banking & Finance*, 35(4):840–855, 2011.
- J. Dermine. Basel III leverage ratio requirement and the probability of bank runs. *Journal of Banking & Finance*, 53:266–277, 2015.
- C. Diks, V. Panchenko, and D. van Dijk. Likelihood-based scoring rules for comparing density forecasts in tails. *Journal of Econometrics*, 163(2):215–230, 2011.
- K. Dutta and J. Perry. A tale of tails: an empirical analysis of loss distribution models for estimating operational risk capital. *Research Review*, Jul-Dec(6):11–14, 2006.
- B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani. Least angle regression. *Annals of Statistics*, 32(2):407–499, 2004.
- P. Embrechts, C. Klupperlberg, and T. Mikosch. *Modelling extremal events for insurance and finance*. Springer - Verlag, Berlin, 1997.
- Y. Fan and C.Y. Tang. Tuning parameter selection in high dimensional penalized likelihood. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 75(3):531–552, 2012.
- C. Feng, H. Wang, and X.M. Tu. The asymptotic distribution of a likelihood ratio test statistic for the homogeneity of Poisson distribution. *Sankhya*, 74-A(Part 2):263–268, 2012.
- B. Gnedenko. Sur la distribution limite du terme maximum d’une série aléatoire. *Annals of Mathematics. Second Series*, 44(2):423–453, 1943.
- J.J. Goeman. L_1 penalized estimation in the Cox proportional hazards model. *Biometrical Journal*, 52(1):70–84, 2010.
- C.W. Granger and P. Newbold. Spurious regression in econometrics. *Journal of Econometrics*, 2(2):111–120, 1974.

- R. Gray. Flexible Methods for Analyzing Survival Data Using Splines, With Applications to Breast Cancer Prognosis. *Journal of the American Statistical Association*, 87(420): 942–951, 1992.
- A. Groll and G. Tutz. Variable selection for generalized linear mixed models by L_1 -penalized estimation. *Statistics and Computing*, 24(2):137–154, 2014.
- A. Guillaou, S. Loisel, and G. Stupfler. Estimating the parameters of a seasonal Markov-modulated Poisson process. *Statistical Methodology*, 26:103–123, 2015.
- A.C. Harvey. *Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series*. Cambridge University Press, Cambridge, 2013.
- V. Kharin, F. Zwiers, X. Zhang, and M. Wehner. Changes in temperature and precipitation extremes in the CMIP5 ensemble. *Climatic Change*, 119(2):345–357, 2013.
- C.T. Koch. Risky adjustments or adjustments to risks: Decomposing bank leverage. *Journal of Banking & Finance*, 45:242–254, 2014.
- L. Laeven and R. Levine. Is there a diversification discount in financial conglomerates? *Journal of Financial Economics*, 85(2):331–367, 2007.
- G. Löffler and A. Maurer. Incorporating the dynamics of leverage into default prediction. *Journal of Banking & Finance*, 35(12):3351–3361, 2011.
- D. Massacci. Tail Risk Dynamics in Stock Returns: Links to the Macroeconomy and Global Markets Connectedness. *Management Science*, 63(9):3072–3089, 2017.
- A. Mayr, N. Fenske, B. Hofner, and T. Kneib. 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, 2012.
- L. Meier, S. Van de Geer, and P. Bühlmann. The Group Lasso for Logistic Regression. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 70(1):53–71, 2008.
- M. Moscadelli. The modelling of operational risk: experience with the analysis collected by the Basel Committee. Technical report, Bank of Italy, 2004.
- R. Nishii. Asymptotic properties of criteria for selection of variables in multiple regression. *The Annals of Statistics*, 12(2):758–765, 1984.
- M.R. Oelker and G. Tutz. A uniform framework for the combination of penalties in generalized structured models. *Advances in Data Analysis and Classification*, 11(1): 97–120, 2017.
- M.Y. Park and T. Hastie. L_1 -regularization path algorithm for generalized linear models. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 19(4):659–677, 2007.

- G. Peters, P. Shevchenko, B. Hassani, and A. Chapelle. Should the advanced measurement approach be replaced with the standardized measurement approach for operational risk? *Journal of Operational Risk*, 11(2):1–49, 2016.
- G. Petrella and A. Resti. Supervisors as information producers: Do stress tests reduce bank opaqueness? *Journal of Banking & Finance*, 37(12):5406–5420, 2013.
- J. Pickands. Statistical inference using extreme order statistics. *The Annals of Statistics*, 3(1):119–131, 1975.
- P. Povel, R. Singh, and A. Winton. Booms, Busts, and Fraud. *Review of Financial Studies*, 20(4):1219–1254, 2007.
- R.A. Rigby and D.M. Stasinopoulos. Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society. Series C: Applied Statistics*, 54(3):507–554, 2005.
- S.M. Saadoghi-Alvandi and A. Malekzadeh. Simultaneous confidence intervals for ratios of means of several lognormal distributions: A parametric bootstrap approach. *Computational Statistics & Data Analysis*, 69:133–140, 2014.
- G. Schauburger and G. Tutz. Subject-specific modelling of paired comparison data - a lasso-type approach. *Statistical Modelling*, 17(3):223–243, 2017.
- M. Scholz and F. Stephens. K-Sample Anderson-Darling Tests. *Journal of the American Statistical Association*, 82(399):918–824, 1987.
- A. Soprano, B. Crielaard, F. Piacenza, and D. Ruspantini. *Measuring Operational and Reputational Risk: A Practitioner’s Approach*. Wiley, 2009.
- R. Tibshirani. Regression Selection and Shrinkage via the LASSO. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 58(1):267–288, 1996.
- F. Valencia. Bank capital and uncertainty. *Journal of Banking & Finance*, 69(S1):S1–S9, 2016.
- J.M. van Zyl. The generalized Pareto distribution fitted to research outputs of countries. *Scientometrics*, 94(3):345–357, 2013.
- J. Wadsworth and J. Tawn. Likelihood-based procedures for threshold diagnostics and uncertainty in extreme value modelling. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 74(3):543–567, 2012.
- T. Wang and C. Hsu. Board composition and operational risk events of financial institutions. *Journal of Banking & Finance*, 37(6):2042–2051, 2013.
- M.L. Weitzman. On modeling and interpreting the economics of catastrophic climate change. *The Review of Economics and Statistics*, 9(1):1–19, 2009.

- H. White. A reality check for data snooping. *Econometrica*, 68(5):1097–1126, 2000.
- B. Williams. The impact of non-interest income on bank risk in Australia. *Journal of Banking & Finance*, 73:16–37, 2016.
- H. Zou. The adaptive LASSO and its oracle properties. *Journal of the American Statistical Association*, 101(476):1418–1429, 2006.
- H. Zou and T. Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 67(2):301–320, 2005.
- H. Zou and R. Li. One-step sparse estimates in nonconcave penalized likelihood models. *Annals of Statistics*, 36(4):1509–1533, 2008.
- H. Zou, T. Hastie, and R. Tibshirani. On the “degrees of freedom” of the LASSO. *Annals of Statistics*, 35(5):2173–2192, 2007.