

Smooth-transition regression models for non-stationary extremes

J. Hambuckers^{1,†} and T. Kneib²

Abstract

We introduce a smooth-transition generalized Pareto (GP) regression model to study the time-varying dependence structure between extreme losses and a set of economic factors. In this model, the distribution of the loss size is approximated by a GP distribution, and its parameters are related to explanatory variables through regression functions, which themselves depend on a time-varying predictor of structural changes. We use this approach to study the dynamics in the monthly severity distribution of operational losses at a major European bank. Using the VIX as a transition variable, our analysis reveals that when the uncertainty is high, a high number of losses in a recent past is indicative of less extreme losses in the future, consistent with a self-inhibition hypothesis. On the contrary, in times of low uncertainty, only the growth rate of the economy seems to be a relevant predictor of the likelihood of extreme losses.

Keywords: Extreme value theory, generalized Pareto distribution, operational risk, VIX.

JEL: C24, C46, C58, G21.

¹ University of Liège, HEC Liège, Finance Department, 14 rue Louvrex, 4000 Liège, Belgium.

² Georg-August-Universität Göttingen, Chair of Statistics, Humboldtallee 3, 37073 Göttingen, Germany.

† Corresponding author: jhambuckers@uliege.be.

1 Introduction

In the financial sector, operational risk is defined as the likelihood of suffering losses resulting from inadequate or failed internal processes, people and systems, or from external events [Basel Committee on Banking Supervision (BCBS), 2004]. Typical examples include losses resulting from unauthorized trading, embezzlement, or legal settlements, and the tremendous amounts involved regularly make the headlines of the media ¹. Early on, this risk was acknowledged as a major concern by the banking industry: a survey conducted by the BCBS in 2003 revealed that around 15% of the total capital of banks was held to cover against operational losses [De Fontenouvelle et al., 2006]. This concern persists to this day: an updated monitoring report of the BCBS concluded that 17% of the Tier-I equity of the largest banks in the world, amounting to \$411bn, is dedicated to protection against operational events [Sands et al., 2018]. The proper modeling and understanding of operational losses is thus a crucial economic challenge for banks. This paper contributes to both aspects and provides a set of recommendations useful for risk managers.

Recently, the research community has turned its attention to the link between the severity distribution of operational losses (i.e. the distribution of the loss sizes) and various economic indicators. This interest stems from the observation that the economic context impacts the incentives and the allocation of resources inside banks, which in turn impact the frequency and the size of operational losses [Cope et al., 2012]. Therefore, studying the nature and the drivers of this dependence can help towards designing risk models in which economic variables are used as indicators of changes in risk levels.

Owing to the heavy-tailed nature of operational loss data, generalized Pareto (GP) regression models are appropriate tools to investigate this dependence [Chavez-Demoulin et al., 2016, Hambuckers et al., 2018a]. In these models, one relies on extreme value theory (EVT) and the peaks-over-threshold (POT) procedure to select extreme events defined as losses taking values that exceed a high threshold. The interest of the POT procedure lies in the fact that the limiting distribution for the exceedances has been shown to be a GP distribution [Balkema and de Haan, 1974, Pickands, 1975, Davison and Smith, 1990], giving us a parametric basis to build a distributional model. Relying on this result, the parameters of the GP distribution are further assumed to be functions of covariates through a set of structural equations, thus specifying an implicit dependence structure between risk factors and the severity distribution. For instance, Chernobai et al. [2011] and Wang and Hsu [2013] report that operational losses are caused by a breakdown of internal controls, suggesting that the number of losses is indicative of the quality of internal controls. Therefore, including the observed number of recent operational losses in a GP regression model would connect the severity distribution of extreme losses with a firm-specific measure of the quality of internal controls.

However, a major issue with GP regression models is the implied assumption of a constant dependence structure over time, whereas one often observes fundamental events such as financial crises or managerial and regulatory decisions that impact how a financial institution responds to economic

¹See our discussion in Section 4 for prominent examples.

incentives. In particular, Adrian and Shin [2010] found that changes in financial uncertainty modify the risk aversion of financial institutions, a major driver in the intensity of the control of business processes after loss events. This suggests that variations in the uncertainty level may impact how internal controls respond to previous losses, thereby influencing the severity distribution of future operational events.

A solution to this problem lies in using regime-switching models. This approach has been successfully applied to the study of market volatility and interest rates [Guidolin, 2011] as well as to the modeling of insurance and operational losses with compound processes [Guillou et al., 2013, Hambuckers et al., 2018b]. However, the regime-switching framework suffers from two important limitations. First, the switching between regimes relies on a Markovian hypothesis. Not only is this hardly true in practice, but this hypothesis also does not tell us anything about the reasons of a structural change, making the interpretation of this model hard in practice. Second, the whole framework implies abrupt changes of regression models from one period to another, ignoring possible transitory states. From a risk management perspective, it has the undesirable effect of generating volatile risk indicators.

In light of these considerations, the primary purposes of this paper are to provide an alternative econometric approach that solves the issues mentioned above and investigate the time-varying dependence structure in the monthly operational loss data of a typical bank.

To begin with, we introduce a novel framework for the study of extremes, which relaxes the assumptions made on the transitions between regression models. To do so, we rely on the concept of *smooth transition* introduced in Teräsvirta [1994]: in our model, the regression equations of the distribution's parameters are defined by the combination of two limiting regression models, with mixing weights depending on a time-varying transition variable. The main advantages of our specification are drastic gains in flexibility while keeping the number of parameters low as well as including the two-state regime as a special case. In addition, the conditional likelihood function is readily available for maximum likelihood estimation. An important difference with regime-switching models, though, is the need to define explicitly the factors driving the transitions. However, although it exposes us to a potential mis-specification of the model, it enhances the interpretation of the results. Whereas state decoding of regime-switching models must be interpreted *ex post*, smooth-transition models enable the formulation of *ex ante* hypotheses about the switching process, which can be used to interpret the results. In particular, we are able to disentangle the loss formation process between a baseline process, i.e. a structural component linking economic determinants to the loss severity distribution, and an amplification mechanism that exacerbates or attenuates the effects of changes in the covariates.

The concept of smooth transition has previously been applied, among others, to the study of economic growth rate, exchange rate, and unemployment rate (see, e.g., van Dijk et al. [2007] for a review) but never to the study of extremes. Since this is the first time that this model is presented and that smooth-transition models have been found particularly hard to estimate [Chan and MacAleer, 2002], we detail several procedures to reduce numerical instabilities. In a simulation study, we demonstrate that our approach leads to good and stable estimations of the proposed smooth-transition model.

Our approach also relates in several ways to dynamic tail risk models recently proposed in the literature, such as the POT-dynamic conditional score (POT-DCS) model of Massacci [2017] or the realized-POT model of Bee et al. [2019]. As in the POT-DCS and realized-POT models, our goal here is to estimate the time-varying parameters of a tail distribution and to measure the influence of explanatory variables on its dynamics. However, whereas Massacci [2017] specifies a time-series model for the law of motion of the parameters and explains the estimated parameters with the covariates through classical regression models in a second step, we take another route, similar to the one taken in Bee et al. [2019]: we start from an initial predictive GP-regression model that explicitly accounts for the effect of covariates on the tail dynamics and not directly for the effect of time. Then, relaxing the linear structure used in Bee et al. [2019], we specify a dynamic law of motion for the parameters *with respect to the covariates*, i.e. allowing for an effect of a variation in the covariates that changes over time. In that sense, our approach should be viewed as complementary to that of Massacci [2017] and Bee et al. [2019], since we focus on non-stationarities arising from the covariates rather than on those arising in the time dimension.

Secondly, we conduct an empirical study of the dependence structure in the severity distribution of extreme operational losses for a large financial institution. To do so, we use a sample of monthly extreme operational losses from the Italian bank UniCredit, recorded over the period 2005M1 - 2014M6. While in earlier studies yearly or quarterly periods of time were used to match the horizon of the regulators, we focus here on a monthly horizon to detect distributional changes relevant at the level of a financial institution. As transition variable, we employ the (log) VIX, motivated by earlier findings demonstrating that it conveys information related to future risk perception. As risk factors, we use the number of operational losses recorded the previous month, relaxing the usual assumption of full independence between severity and frequency processes. The past number of losses is seen here as a proxy for the efficiency of internal controls. With this approach, we investigate the existence of a *self-inhibition mechanism*, i.e. situations in which breakdowns of internal controls are followed by corrective actions reducing the severity of future operational losses. In addition, we use a set of macro and financial variables to capture the effect of the economic context.

We find major changes in the relationship between the severity distribution and the considered variables along changes in VIX values: during periods of high VIX values, the severity distribution is characterized by a significant self-inhibition effect, suggesting that months with a high number of losses are predictive of future months with losses of small sizes. This effect may be explained by a tightening of control procedures following operational events, in anticipation of financial stress for the firm. On the contrary, periods of low VIX values exhibit no self-inhibition effects, and only the macro-financial context appears to influence the severity distribution. In particular, the likelihood of extreme losses appears to be connected with economic growth, potentially through the channel of transaction sizes. Finally, we demonstrate that the proposed specification significantly improves the fit of the severity distribution over time, with respect to simpler GP regression models, and provides value-at-risk (VaR) estimates of the total loss distribution with better coverage probabilities far in the

tail.

The structure of the paper is as follows: In Section 2, we provide a description of the various models and estimation procedures used to analyze the severity distributions. In Section 3, we study the finite sample properties of the proposed methodology in a realistic Monte Carlo simulation. In Section 4, we present the data and then discuss the results of our empirical study in Section 5, including the economic interpretation and the policy implications for risk management. We conclude in Section 6.

2 Models, estimation methods and testing procedures

In this section, we detail the set of econometric procedures used to study the time-varying dependence structure of extremes. We first define the model, before discussing estimation, inference, testing, and model selection issues.

2.1 Severity model in a non-stationary context

The focus here is on defining a regression model for the severity distribution of extreme losses, i.e. losses $Z_{t,i}$ larger than a threshold τ ². The index i goes from 1 to n_t , n_t being the observed number of losses larger than τ over the period t . We denote by $\mathbf{n} = \{n_t\}_{t=1,\dots,T}$ the time series of observed counts up to T . We also define the exceedances as $Y_{t,i} = Z_{t,i} - \tau$. Relying on EVT and the POT approach, the natural distribution of $Y_{t,i}$ is the GPD. Its cumulative distribution function (cdf) is given by

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

with $y \geq 0$. $\gamma \in \mathbb{R}$ and $\sigma > 0$ are the shape and scale parameters, respectively. For $\gamma < 0$, we have $0 < y < -\sigma/\gamma$ and y is bounded. In the case of $\gamma = 0$, the GPD has an exponential decay. For $\gamma \in (0, 1)$, it can be shown that Y has a finite first moment [Embrechts et al., 1997]. This is often a desirable property from a methodological perspective (e.g. for moment-based inference) and for applications involving the computation of a conditional expectation. We restrict our attention to the case where $\gamma > 0$, i.e. the heavy-tail case.

The 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]. In particular, if the cdf of Z belongs to the max-domain of attraction of a generalized extreme value distribution, then the limiting distribution of exceedances over a high threshold is a GPD [see, e.g. Embrechts et al., 1997, Davison and Smith, 1990, for theoretical details]. In practice, we assume that the severity distribution of the exceedances above a high threshold is effectively GPD and the regression analysis is performed on $Y_{t,i}$ [Chavez-Demoulin et al., 2016, Hambuckers et al., 2018b,a].

²The index t refers to the time, whereas the index i refers to the i^{th} loss larger than τ during a given period.

Here, we apply the POT method in a non-stationary context: we assume that the severity distribution of the exceedances, conditional on the chosen threshold, is GPD but with γ and σ determined by a *smooth-transition regression model* following the idea of Teräsvirta [1994]. This model implies that both distribution parameters are weighted combinations of two limiting regression models. The weights are derived from a logistic function depending on a *transition variable*, denoted by s_t . Mathematically, for the i^{th} exceedance taking place over the time period $]t - 1, t]$, we assume that

$$Y_{t,i} \sim F(y; \gamma(\mathbf{x}_{t,i}^\gamma, s_t), \sigma(\mathbf{x}_{t,i}^\sigma, s_t)), \quad (2)$$

with $Y_{t,i} \geq 0$, $\gamma(\mathbf{x}_{t,i}^\gamma, s_t) > 0$, $\sigma(\mathbf{x}_{t,i}^\sigma, s_t) > 0$, and where $\mathbf{x}_{t,i}^\gamma$, $\mathbf{x}_{t,i}^\sigma$ and s_t are the vector of economic covariates for the shape parameter, the scale parameter, and the transition variable, respectively. The transition variable is assumed here to be some exogenous variable. The parameters $\gamma(\mathbf{x}_{t,i}^\gamma, s_t)$ and $\sigma(\mathbf{x}_{t,i}^\sigma, s_t)$ are the shape and scale parameters for the severity distribution of the i^{th} exceedance in period t . They are given by:

$$\gamma(\mathbf{x}_{t,i}^\gamma, s_t) = \phi(s_t)h\left((\mathbf{x}_{t,i}^\gamma)^T \boldsymbol{\beta}_1^\gamma\right) + (1 - \phi(s_t))h\left((\mathbf{x}_{t,i}^\gamma)^T \boldsymbol{\beta}_2^\gamma\right), \quad (3)$$

$$\sigma(\mathbf{x}_{t,i}^\sigma, s_t) = \phi(s_t)h\left((\mathbf{x}_{t,i}^\sigma)^T \boldsymbol{\beta}_1^\sigma\right) + (1 - \phi(s_t))h\left((\mathbf{x}_{t,i}^\sigma)^T \boldsymbol{\beta}_2^\sigma\right), \quad (4)$$

with

$$\phi(s_t) = \frac{1}{1 + \exp\{-g(s_t - c)\}}, \quad (5)$$

where $\phi(\cdot) \in (0, 1)$ is the logistic weight function, $g > 0$ is the dispersion parameter, c the location parameter, whereas $\boldsymbol{\beta}_r^\theta$, with $\theta \in \{\gamma, \sigma\}$ and $r = 1, 2$ are the vectors of regression parameters in the different limiting states. $h(\cdot)$ is a suitable response function ensuring the strict positivity of the distribution parameters (e.g. the exponential function). Equations (2)–(4) describe our *smooth-transition generalized Pareto* (ST-GPD) regression model.

This model can be viewed as combining two limiting regression models via the weight component $\phi(s_t)$. It defines particular regression models at time $t = 1, \dots, T$, possibly “in between” the two limiting models. Indeed, owing to ϕ taking a continuum of values over $(0, 1)$ for s_t being continuous, this structure implies that we do not necessarily shift abruptly from one limiting model to the other, as in a Markov switching model. Instead, depending on the speed of the transition function (5) (regulated by g) and the process driving s_t , we have $0 < \phi(s_t) < 1$, and the effect of a change in covariates is determined by both $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ (which explains why the transition is referred to as “smooth”).

Of course, the “smoothness” of the model strongly depends on the values taken by g and s_t . For $g \rightarrow +\infty$, (5) becomes a step function, and the model is similar to threshold regression, leading to abrupt changes between the two limiting models. Similarly, a large variance in s_t can also lead to abrupt shifts, owing to ϕ taking values close to either 0 (for $s_t \ll c$) or 1 (for $s_t \gg c$). In our empirical application, we find a transition function that is rather steep. Nevertheless, the merit of the present approach consists in enabling a high degree of flexibility and letting the data determine whether the transitions are more or less smooth without *a-priori* choices by the analyst. It also reflects our belief

that the regression structure is not stable over time and that more than two (separated) regression models are needed to depict the process adequately.

To ease the notation, the explicit references to $\mathbf{x}_{t,i}^\gamma$, $\mathbf{x}_{t,i}^\sigma$ and s_t are omitted in the following, so that $\gamma(\mathbf{x}_{t,i}^\gamma, s_t)$ and $\sigma(\mathbf{x}_{t,i}^\sigma, s_t)$ are replaced by $\gamma_{t,i}$ and $\sigma_{t,i}$ in the rest of the paper. An important specification of the present model is the assumption of a unique transition function, driving both γ and σ . This has the advantage that c and g appear in the two structural equations, which potentially leads to an easier estimation of these parameters. Notice also that we apply the transition function to $h(\mathbf{x}^T\boldsymbol{\beta})$ instead of $\mathbf{x}^T\boldsymbol{\beta}$. This structure facilitates the interpretation of the model, defining the terms $h(\mathbf{x}^T\boldsymbol{\beta})$ as the value of the distribution parameter in the limiting regression models.

2.2 Estimation procedure

For our severity model, the log-likelihood function is given by

$$\mathcal{L}(\mathbf{y}; \Theta, \mathbf{x}, \mathbf{s}, \boldsymbol{\tau}, \mathbf{n}) = \sum_{t=1}^T \sum_{i=1}^{n_t} \log(f(y_{t,i}; \gamma_{t,i}, \sigma_{t,i})), \quad (6)$$

where $f(\cdot)$ denotes the probability density function (pdf) of the GPD, Θ is the set of all parameters of the model, \mathbf{y} is the vector of all losses, \mathbf{x} a design matrix of all observed covariates, \mathbf{s} the vector of transition variables, $\boldsymbol{\tau}$ the threshold parameter, and \mathbf{n} the observed frequency process. Notice that (6) relies on the *contemporaneous* conditional independence between frequency and severity, implying that n_t and $y_{t,i}$ are independent given the covariates. With this assumption, we can split the full likelihood into a severity part and a frequency part and then study the two processes separately [Chavez-Demoulin et al., 2016]. However, it does not prevent n_{t-1} and $y_{t,i}$ from being dependent. Therefore, in \mathbf{x} , we can include n_{t-1} ³. An estimator $\hat{\Theta}$ of the regression parameters is obtained by maximizing Eq. (6) with respect to Θ :

$$\hat{\Theta} = \arg \max_{\Theta} \{\mathcal{L}(\mathbf{y}; \Theta, \mathbf{x}, \mathbf{s}, \boldsymbol{\tau}, \mathbf{n})\}. \quad (7)$$

This maximization is both notoriously difficult to conduct and sensitive to starting values; see the discussions in Chan and MacAleer [2002], Chan and MacAleer [2003] and Chan and Theoharakis [2011]. Several papers (e.g. Dungey et al. [2015]) use grid-search or two-step maximizations. Here, we resort to two different procedures to improve numerical stability. Firstly, following Chan and Theoharakis [2011], we re-parametrize the transition function in terms of (\bar{g}, c) , where $\bar{g} = 1/\sqrt{g}$, $g > 0$. Secondly, following Silvennoinen and Teräsvirta [2009], we set boundary conditions on the location parameters of the transition function. More details are given in Appendix A. In Section 3, we demonstrate with realistic simulations that we reliably estimate the different coefficients with this approach and do not need two-step procedures.

Regarding inference on the regression coefficients, we assume that asymptotic normality of (7) holds, and we use the inverse of the Fisher information matrix of the log-likelihood as asymptotic covariance matrix (see Appendix A for more details on the computation of these quantities). We

³To limit the complexity of the final model, we do not consider variables beyond a lag of one period.

can therefore conduct Wald-type tests to test restrictions on the regression coefficients of the type $H_0 : \beta_r^\theta = 0$, with $\theta \in \{\gamma, \sigma\}$ and $r = 1, 2$. Coverage properties of this approach are examined in the simulation study. Overall, Wald-type tests exhibit good sizes at the 5% test level and satisfactory power for the considered data-generating processes (DGPs, see Section 3). Notice that the asymptotic normality is likely to hold only if the model is well identified, i.e. if $\beta_1 = (\beta_1^\gamma, \beta_1^\sigma) \neq (\beta_2^\gamma, \beta_2^\sigma) = \beta_2$. Thus, a first step would be to test for a correct specification of the time-varying structure. This testing issue is discussed in Section 2.3.

2.3 Testing for a time-varying structure

The use of smooth-transition models inevitably raises questions concerning the correct specification of an additional time structure. Early on, Lukkonen et al. [1988] discussed testing for linearity against alternative ST models. This task is drastically complicated by the presence of nuisance parameters not identified under the null hypothesis of no ST specification. Indeed, for our model, notice that equations (3) and (4) can be rewritten as

$$\gamma(\mathbf{x}_{t,i}^\gamma, s_t) = h\left((\mathbf{x}_{t,i}^\gamma)^T \beta_2^\gamma\right) + \phi(s_t) \left(h\left((\mathbf{x}_{t,i}^\gamma)^T \beta_1^\gamma\right) - h\left((\mathbf{x}_{t,i}^\gamma)^T \beta_2^\gamma\right) \right), \quad (8)$$

$$\sigma(\mathbf{x}_{t,i}^\sigma, s_t) = h\left((\mathbf{x}_{t,i}^\sigma)^T \beta_2^\sigma\right) + \phi(s_t) \left(h\left((\mathbf{x}_{t,i}^\sigma)^T \beta_1^\sigma\right) - h\left((\mathbf{x}_{t,i}^\sigma)^T \beta_2^\sigma\right) \right). \quad (9)$$

Under the null hypothesis H_0 of no ST effect, we have $\beta_1^\gamma = \beta_2^\gamma$ and $\beta_1^\sigma = \beta_2^\sigma$, making the second terms of the right-hand sides equal to zero. Consequently, parameters of ϕ (i.e. c and \bar{g}) are not identified under H_0 , and the distribution of a test statistic like the likelihood ratio (LR) is non-standard. One way of circumventing this issue consists in testing an auxiliary regression model, based on a Taylor expansion of ϕ that avoids identification problems [van Dijk et al., 2007]. However, the performance of this test heavily depends on the quality of that approximation, which might be very bad for a steep transition function. In addition, whereas simplifications arise for restricted models being linearly additive, it is unclear as to how the residual terms in the Taylor series affect the random part of our model. Thus, we prefer using simulation-based methods to derive the distribution of the test statistic under the null hypothesis. We consider the following test statistic:

$$LR(\hat{\Theta}^{H_0}, \hat{\Theta}) = -2(\mathcal{L}(\mathbf{y}; \hat{\Theta}^{H_0}, \mathbf{x}, \boldsymbol{\tau}, \mathbf{n}) - \mathcal{L}(\mathbf{y}; \hat{\Theta}, \mathbf{x}, \mathbf{s}, \boldsymbol{\tau}, \mathbf{n})), \quad (10)$$

with $\mathcal{L}(\mathbf{y}; \hat{\Theta}^{H_0}, \mathbf{x}, \boldsymbol{\tau}, \mathbf{n})$ being the log-likelihood function of the null GP-regression model with maximum likelihood estimate $\hat{\Theta}^{H_0}$. It is obtained by fitting a GP-regression model on our sample of extremes, where

$$\gamma(\mathbf{x}_{t,i}^\gamma) = h\left((\mathbf{x}_{t,i}^\gamma)^T \beta_{H_0}^\gamma\right), \quad (11)$$

$$\sigma(\mathbf{x}_{t,i}^\sigma) = h\left((\mathbf{x}_{t,i}^\sigma)^T \beta_{H_0}^\sigma\right). \quad (12)$$

The distribution of (10) under H_0 is obtained by means of the fixed-regressors bootstrap procedure of Hansen [1996]. In the present case, the steps of this procedure are:

- I. Using $\hat{\Theta}^{H_0}$, fixed \mathbf{x}, \mathbf{n} (i.e. the number of extremes at each point in time) and the GP-regression model, simulate a time series of exceedances $\mathbf{y}_{(b)}^*$ under H_0 (i.e. following the restricted model). With this approach, we enforce the null hypothesis to be true, given the observed values of the regressors and the selected thresholds.
- II. Estimate the vectors of parameters $\hat{\Theta}_{(b)}^{*,H_0}$ and $\hat{\Theta}_{(b)}^*$ from the null and the ST models, respectively.
- III. Compute $LR(\hat{\Theta}_{(b)}^{*,H_0}, \hat{\Theta}_{(b)}^*)$, a bootstrap realization of the likelihood ratio.
- IV. Repeat for $b = 1, \dots, B$, e.g. 1,000 times.

The rejection region at level α is defined by $[q_{1-\alpha}^*, +\infty)$, where $q_{1-\alpha}^*$ is the empirical $1 - \alpha$ quantile of $LR(\hat{\Theta}_{(b)}^{*,H_0}, \hat{\Theta}_{(b)}^*)$. Notice that, as for the estimation procedure in Section 2.2, we neglect the uncertainty stemming from the threshold selection procedure and work only with the extremes. This is necessary, since the time series of the number of extremes enters the set of regressors. In Hansen [1996], the author concentrates on additive non-linear models and ordinary least squares estimators, not on non-separable models as ours. However, his discussion on page 1, as well as footnote 1, suggest that this procedure can be extended to models requiring non-linear maximum likelihood estimation procedures. We follow this suggestion, although a formal proof is beyond the scope of the present paper. We demonstrate in our simulation study in Section 3 that the test exhibits adequate size, hinting to its validity in the present context.

Finally, note that with the definition of our model, we easily test for the absence of ST structure in one of the two distribution parameters. Indeed, since (11) and (12) are driven by a unique transition function, \bar{g} and c remain identified for this less restrictive null hypothesis. In this situation, classical likelihood ratio theory applies.

2.4 Model selection and specification

We now briefly discuss the selection and specification criteria used in Section 5. First, we rely on the censored likelihood score (CLS) and its cross-validated equivalent, denoted CLS_α^{CV} . Introduced by Diks et al. [2011], CLS is useful to assess the goodness-of-fit for extreme observations, avoiding both selection bias and outcome conditioning (see, e.g. the discussions in Gneiting and Ranjan [2011]). Fitting extreme observations well is what matters most here, since it determines how good our risk measures will be. The CLS for a given censoring threshold α is given by

$$CLS_\alpha = - \sum_{t=1}^T \sum_{i=1}^{n_t} \left\{ \mathbb{1}(y_{t,i} > \kappa(\alpha)) \log(\text{gpd}(y_{t,i}; \hat{\gamma}_{t,i}, \hat{\sigma}_{t,i})) \right. \\ \left. + \mathbb{1}(y_{t,i} \leq \kappa(\alpha)) \log(\text{GPD}(\kappa(\alpha); \hat{\gamma}_{t,i}, \hat{\sigma}_{t,i})) \right\}, \quad (13)$$

where $\hat{\gamma}_{t,i}$ and $\hat{\sigma}_{t,i}$ are estimated parameters for observation $y_{t,i}$. The smaller the CLS, the better the fit, particularly in the tail of the distribution. We choose $\kappa(\alpha)$ as the empirical α -quantile of all losses pooled together for various values of α . This gives special weight to extremely large observations from a historical perspective, as in Hambuckers et al. [2018a]. Increasing α gives increasing weight to extreme

observations. To control for potential in-sample overfitting, we also use the cross-validated version of (13), CLS_{α}^{CV} . This quantity is simply obtained by replacing $\hat{\gamma}_{t,i}$ and $\hat{\sigma}_{t,i}$ by their corresponding cross-validated estimators (see also Hambuckers et al. [2018a])⁴.

Second, we look at the quality of VaR estimates of the total loss distribution⁵ provided by different models. The idea here is to test whether a given model is correctly specified far in the tail of the distribution. To test this claim under experimental conditions independent from a frequency model, we compute quantile estimates conditional on knowing the exact number of losses taking place over each time period. That is, for each model under consideration and time period, we simulate 1 million samples of losses, using fixed covariates and fixed count data. Then, we compute the simulated total loss and retrieve the empirical quantile at various levels (95%, 97.5% and 99%)⁶. The correct specification of each model is assessed via the multivariate unconditional coverage test proposed in Colletaz et al. [2013]. In this approach, we test jointly for the correctness of the number of exceptions and *super exceptions* of the VaR estimates. An exception is defined as a time period in which the total loss is larger than the estimated VaR. This procedure accounts not only for the number of VaR exceedances for a level of interest α_1 , but also for their magnitudes, by computing simultaneously a VaR at level α_2 , with $\alpha_2 > \alpha_1$. More formally, we test the following null hypothesis:

$$H_0 : \mathbb{E}(I_t(\alpha_1)) = \alpha \text{ and } \mathbb{E}(I_t(\alpha_2)) = \alpha_2, \quad (14)$$

where $I_t(\alpha_1)$ is an indicator function taking value 1 if the total loss at time t is larger than the VaR at level α_1 ($I_t(\alpha_2)$ is defined similarly). Defining $N(\alpha) = \sum_{t=1}^T I_t(\alpha)$ as the number of exceptions for level α , we test (14) relying on the likelihood ratio test statistic given by

$$\begin{aligned} LR_{MUC}(\alpha_1, \alpha_2) = & -2 \log \left[\alpha_1^{T-N(\alpha_1)} (\alpha_2 - \alpha_1)^{N(\alpha_1)-N(\alpha_2)} (1 - \alpha_2)^{N(\alpha_2)} \right] \\ & + 2 \log \left[\left(\frac{N_0}{T} \right)^{N_0} \left(\frac{N_1}{T} \right)^{N_1} \left(\frac{N(\alpha_2)}{T} \right)^{N(\alpha_2)} \right], \end{aligned} \quad (15)$$

with $N_0 = T - N(\alpha_1)$ and $N_1 = N(\alpha_1) - N(\alpha_2)$. The finite sample distribution of (15) is obtained via simulations, using the procedure proposed in Section 2.3 of Colletaz et al. [2013]⁷. Rejecting (14) leads to the conclusion that the model under study is mis-specified. Of course, the result of the test depends on the value chosen for α_2 . To neutralize this effect, in Section 5, we consider several values of α_2 between α_1 and 1.

⁴We use 10-fold cross-validation: we randomly split the sample into 10 subgroups. Then, we remove one subgroup at a time, fit the models on the remaining 9 groups, and compute the CLS statistic on the omitted group. The final cross-validated CLS is the sum of the CLS on the 10 groups.

⁵The VaR (or quantile) of Y for a level α , denoted Q_{α} , is defined by $\mathbb{P}(Y \leq Q_{\alpha}) = \alpha$.

⁶Notice that some results exist to retrieve the distribution of the sum of heavy-tail random variables analytically, but only in the i.i.d. case, see e.g. Kratz [2014].

⁷In this procedure, we simulate 1 million time series of length T from a $U(0,1)$ distribution and compute how many times the random variables are larger than α_1 and α_2 . Then, the distribution of (15) is obtained from the empirical distribution.

3 Simulation study

In this section, we investigate the finite-sample properties of the proposed procedures. We generate data from equations (2) to (4). To study these properties in a realistic context, we simulate covariates from processes exhibiting features similar to those of our empirical analysis. First, we assume that the p covariates are the same in the equations of σ and γ , as well as for losses taking place during the same period:

$$\mathbf{x}_{ti}^\sigma = \mathbf{x}_{ti}^\gamma = \mathbf{x}_{ti} = \mathbf{x}_t. \quad (16)$$

This has the effect of introducing correlation between γ and σ , as in our empirical study. Secondly, since we use the past number of losses as a predictor, we simulate one covariate from an inhomogeneous Poisson process with a self-exciting structure [Porter and White, 2012], defined by

$$x_{t1} \sim \text{Pois}(\lambda_t), \quad (17)$$

with

$$\lambda_t = \exp(1.3 + .2x_t^\lambda) + .5 \sum_{i:t_i < t} b(t - t_i; 10), \quad (18)$$

and

$$x_t^\lambda = .1 + .65x_{t-1}^\lambda + e_t^\lambda, \quad e_t^\lambda \stackrel{iid}{\sim} N(0, 1), \quad (19)$$

where $b(\cdot; 10)$ is the standardized geometric distribution with decay rate $r = 10$. These assumptions reflect that the frequency process of operational losses was found to follow a Poisson process depending on covariates (see Chernobai and Yildirim [2008] and Wang and Hsu [2013] among others), but also that occurrences of financial extremes exhibit self-exciting features [e.g Chavez-Demoulin et al., 2005, 2006, Grothe et al., 2014]. Equation (18) incorporates both assumptions. As a consequence, the loss distribution is also (indirectly) affected by the self-exciting structure, leading to a clustering of severe events, a feature similar to Chavez-Demoulin et al. [2005]. We choose x_t^λ to follow a simple AR(1) process with a high persistence parameter, as it is common for macro or financial variables. The values of the parameters in Eq. (18) are chosen, such that (17) mimics the historical time series of counts. For the $p - 1$ other covariates, we simulate from AR(1) processes with correlated error terms:

$$x_{tj} = .1 + .65x_{t-1j} + e_{tj}, \quad j = 2, \dots, p \quad (20)$$

$$\begin{pmatrix} e_{t2} \\ \vdots \\ e_{tp} \end{pmatrix} \stackrel{iid}{\sim} \text{MN}(\mathbf{0}, \mathbf{\Sigma}), \quad (21)$$

with $\mathbf{\Sigma}$ being the variance-covariance matrix with 1 on the diagonal elements and .5 otherwise. This has the effect of introducing a high degree of correlation between covariates and strong time dependencies. These choices are consistent with several studies highlighting these features in macro-finance time series (see the discussion in Phillips and Lee [2013]). Finally, we consider a similar AR(1) process for the state variable:

$$s_t = \alpha_s s_{t-1} + e_t^s, \quad e_t^s \stackrel{iid}{\sim} N(0, 1).$$

We choose $\alpha_s \in \{.7, .8, .9\}$, so that the transition variable exhibits highly persistent time-dependencies like the VIX. We consider $n_t \in \{10, 20, 30, 50\}$ fixed for a given simulation, reflecting the observed number of losses per time period in our data, and $T \in \{50, 100, 150, 200\}$. Furthermore, we set $p = 3$, $c = 0$ and $\bar{g} \in \{.4, .5, .7\}$ to assess the effect of the transition speed on the estimation. The true values of the regression coefficients β_1^γ , β_2^γ , β_1^σ and β_2^σ for the different data generating processes are displayed in Table 1 (second column). For each set of parameters, we generate $N = 200$ samples.

Table 1 provides mean estimates of the regression coefficients with $\bar{g} = .4$ (for clarity, only a subset of the simulation setups are presented, please refer to the supplementary material for the other simulations). We focus on simulations with $T = 100$ and $T = 150$, time-series lengths in line with our application. We see that all parameters are well-estimated, even when n_t is small. When n_t or T increases, we see that variability of the estimation decreases. With respect to the transition parameters known to be difficult to estimate, we observe very good estimations, close to the true parameters. Focusing on the case $\alpha_s = .8$ and $\bar{g} = .5$ for clarity, Figure 1 displays the root mean squared error (RMSE) for \bar{g} and c across all samples, respectively. For both parameters, we observe a decrease in RMSE with increasing T and n_t . A similar effect is observed when the degree of persistence for s_t decreases (i.e. when α_s decreases). Regarding the Wald-type test proposed in Section 2, we observe an excellent size of the test (Figure 2, left panel). For the 5% test level, on average across all parameters and the 200 simulations, the type-I error is always close to the test level⁸. This result indicates that, beyond good point estimates, we obtain good density estimates of the regression coefficients. Regarding the power of the test, we observe satisfactory rejection levels of the null of a zero coefficient (Figure 2, right panel). The average proportions of rejection vary between 25% and 75% for the considered DGP and samples sizes in line with our application. The results for the other combinations of α_s and \bar{g} are similar (see the supplementary material).

Lastly, we briefly investigate the size of the likelihood ratio test proposed in Section 2.3. Since it is time-consuming to study the bootstrap method in a simulation setup, we use the warp-speed technique proposed by Giacomini et al. [2013] to shorten the computation time. For our approach, this simply comprises resampling only $B' = 1$ sample of bootstrap data for each Monte Carlo sample and computing $q_{1-\alpha}^*$ from $N = B$ samples. We can then easily investigate cases with $p = 4$ and $p = 5$. We set $N = 1,000$. We simulate samples following a GP-regression model without smooth-transition structure, and with its parameters given by β_1^γ and β_1^σ . We use the same assumptions regarding the predictors. For s_t , we set $\alpha_s = .8$. We focus on the case $n_t = 20$ and $T = 100$. We compare the size of this test with the sizes of tests based on χ^2 distributions. Under the null, the true model has $2 \times p$ parameters, whereas the ST model has $2 \times 2 \times p + 2$ parameters. Therefore, under H_0 and a correct identification of both models, the test statistic (10) would follow a $\chi_{2 \times p + 2}^2$ distribution. Table 2 depicts the obtained sizes. The classical LR test is strongly oversized, leading us to reject the null wrongly too frequently. On the contrary, our bootstrap test exhibits rejection frequencies

⁸As pointed out by Teräsvirta [1994], we might sometimes obtain negative estimates of the asymptotic standard deviations preventing us from computing a test statistic, owing to numerical approximations. Here, on average across simulations, it concerns only 0.17% of the samples.

closer to the expected nominal levels (5% and 10%). As highlighted in Self and Liang [1987], the true distribution is likely a mixture of χ^2 distributions, leading to a null distribution with heavier tails than a $\chi^2_{2 \times p+2}$ distribution. Figure 3 illustrates these results, displaying various distributions against the true empirical distribution of (10), for $p = 5$. We see that the bootstrap distribution matches the true distribution pretty well.

Thus, our simulation study highlights the correctness of our procedure in terms of estimation, uncertainty quantification, and testing. Notice that for short time series ($T < 150$), we observe a sharp reduction in estimation variability when the number of observations per time period n_t is larger than 10. We would therefore advise future users to consider sufficiently dense data, with $n_t > 10$ and $T \geq 100$. Finally, please note that additional simulations relaxing the conditional normality and homoscedasticity in Eq. (20) and (21) can be found in the supplementary material. The results are qualitatively similar to those presented here.

Table 1: Average estimated regression coefficients, across the 200 simulations and the different values of α_s . In parentheses, average standard deviation of the estimates, across the specifications of the simulations.

Param.	Truth	n_t	T=100				n_t	T=150			
			10	20	30	50		10	20	30	50
\bar{g}	.4		.399 (0.03)	.398 (0.02)	.4 (0.02)	.4 (0.01)		.398 (0.03)	.4 (0.02)	.399 (0.02)	.399 (0.01)
c	0		.001 (0.06)	.001 (0.04)	.001 (0.03)	.001 (0.03)		.003 (0.04)	.00 (0.03)	-.001 (0.03)	0.001 (0.02)
β_{10}^γ	-1.5		-1.58 (0.73)	-1.68 (1.72)	-1.54 (0.34)	-1.54 (0.52)		-1.58 (0.49)	-1.55 (0.34)	-1.52 (0.29)	-1.52 (0.2)
β_{11}^γ	.04		.02 (0.13)	.046 (0.12)	.038 (0.04)	.041 (0.05)		.036 (0.06)	.04 (0.04)	.038 (0.03)	.04 (0.02)
β_{12}^γ	.2		.207 (0.3)	.174 (0.42)	.196 (0.12)	.205 (0.14)		.207 (0.16)	.196 (0.12)	.202 (0.1)	.197 (0.07)
β_{13}^γ	-1		-.095 (0.3)	-.069 (0.56)	-.102 (0.13)	-.114 (0.18)		-.099 (0.16)	-.097 (0.11)	-.099 (0.09)	-.098 (0.07)
β_{20}^γ	-1.3		-3.82 (22.07)	-1.48 (1.49)	-1.41 (0.51)	-1.33 (0.35)		-1.449 (0.83)	-1.412 (0.46)	-1.34 (0.39)	-1.335 (0.28)
β_{21}^γ	-.04		-.093 (1.42)	-.048 (0.12)	-.0437 (0.08)	-.044 (0.05)		-.054 (0.14)	-.041 (0.06)	-.043 (0.05)	-.04 (0.04)
β_{22}^γ	-1		-5.63 (4.86)	-1.05 (0.48)	-1.08 (0.19)	-1.05 (0.14)		-0.11 (0.28)	-.098 (0.17)	-.099 (0.14)	-.098 (0.1)
β_{23}^γ	.1		.814 (6.92)	.108 (0.26)	.12 (0.26)	.092 (0.14)		.079 (0.25)	.102 (0.18)	.1 (0.13)	.098 (0.1)
β_{10}^σ	4		4.001 (0.21)	4.004 (0.14)	4.006 (0.1)	4.003 (0.1)		4.007 (0.15)	4.0050 (0.11)	4.005 (0.09)	4.001 (0.07)
β_{11}^σ	.02		.021 (0.02)	.021 (0.02)	.02 (0.01)	.02 (0.01)		.021 (0.02)	.02 (0.01)	.02 (0.01)	.02 (0.01)
β_{12}^σ	.16		.164 (0.07)	.159 (0.05)	.161 (0.04)	.161 (0.03)		.164 (0.05)	.162 (0.04)	.161 (0.03)	.16 (0.02)
β_{13}^σ	-.05		-.054 (0.07)	-.048 (0.05)	-.048 (0.04)	-.05 (0.03)		-.054 (0.05)	-.049 (0.03)	-.051 (0.03)	-.051 (0.02)
β_{20}^σ	1.5		1.509 (0.25)	1.507 (0.18)	1.509 (0.15)	1.506 (0.12)		1.514 (0.19)	1.521 (0.13)	1.498 (0.11)	1.502 (0.09)
β_{21}^σ	-.035		-.036 (0.03)	-.035 (0.03)	-.036 (0.02)	-.035 (0.01)		-.036 (0.02)	-.037 (0.02)	-.035 (0.01)	-.035 (0.01)
β_{22}^σ	-.11		-.113 (0.09)	-.111 (0.07)	-.109 (0.06)	-.11 (0.04)		-.114 (0.07)	-.113 (0.05)	-.113 (0.04)	-.112 (0.03)
β_{23}^σ	.1		.107 (0.09)	.102 (0.06)	.096 (0.06)	.1 (0.05)		.107 (0.07)	.101 (0.05)	.1 (0.04)	.101 (0.03)

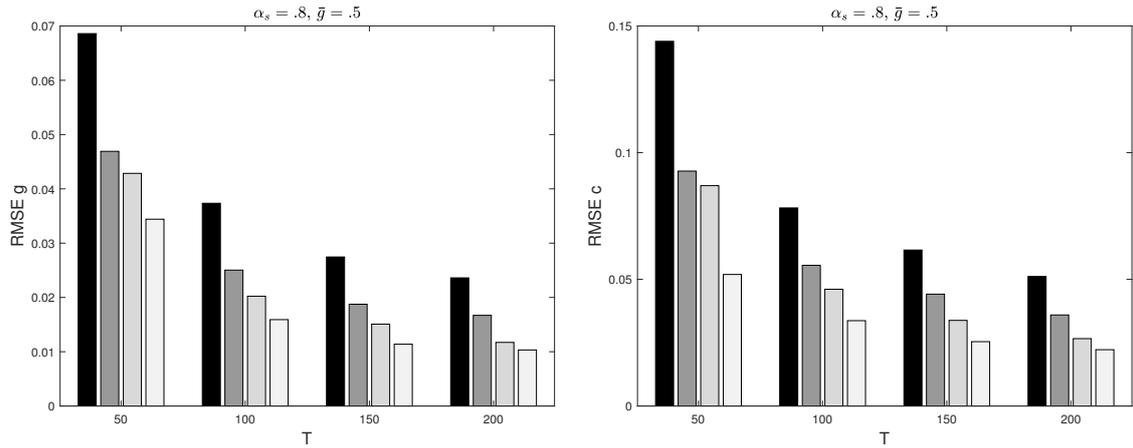


Figure 1: Simulation results for the DGP with $\bar{g} = .5$ and $\alpha_s = .8$. X-axis: length T of the time series. Shade of grey: number of observations n_t per time period (from 10 (black) to 50 (white)). Left: RMSE for parameter \bar{g} . Right: RMSE for parameter c .

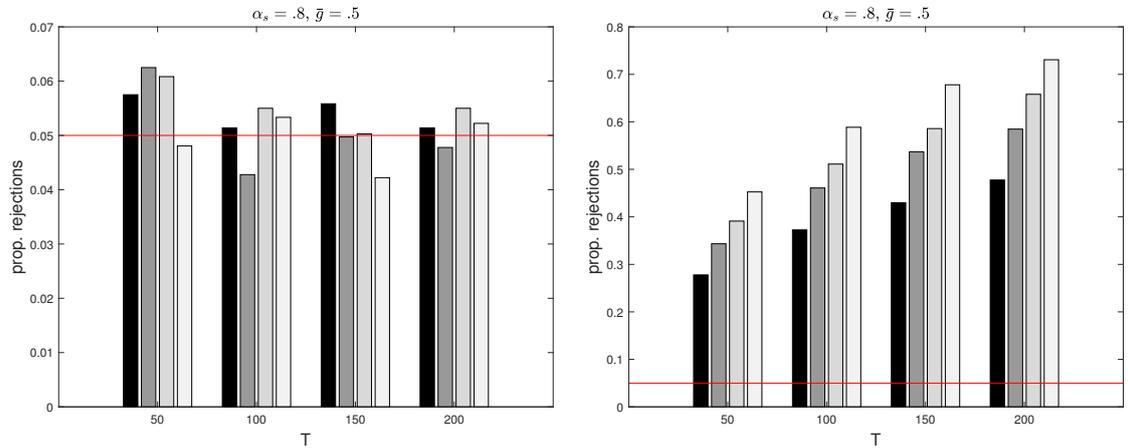


Figure 2: Simulation results for the DGP with $\bar{g} = .5$ and $\alpha = .8_s$. Average power of 5% Wald tests, across the 18 parameters under the null of the regression coefficients taking their true values (left panel) and the null of the regression coefficients being equal to 0 (right panel). Red solid: test level. Shade of grey: number of observations n_t per time period (from 10 (black) to 50 (white)).

Table 2: Sizes of the LR test given by Eq. (10), with various methods to compute the rejection region. BS refers to the bootstrap procedures, assessed through the warp-speed approach. Sizes are obtained from $N = 1,000$ Monte Carlo samples.

Test level	5%		10%	
# cov.	BS	$\chi^2_{2 \times p+2}$	BS	$\chi^2_{2 \times p+2}$
$p = 3$	4.7%	8.1%	8.8%	16.7%
$p = 4$	5.5%	9.4%	10.9%	17.6%
$p = 5$	6.5%	9.9%	10.7%	16.2%

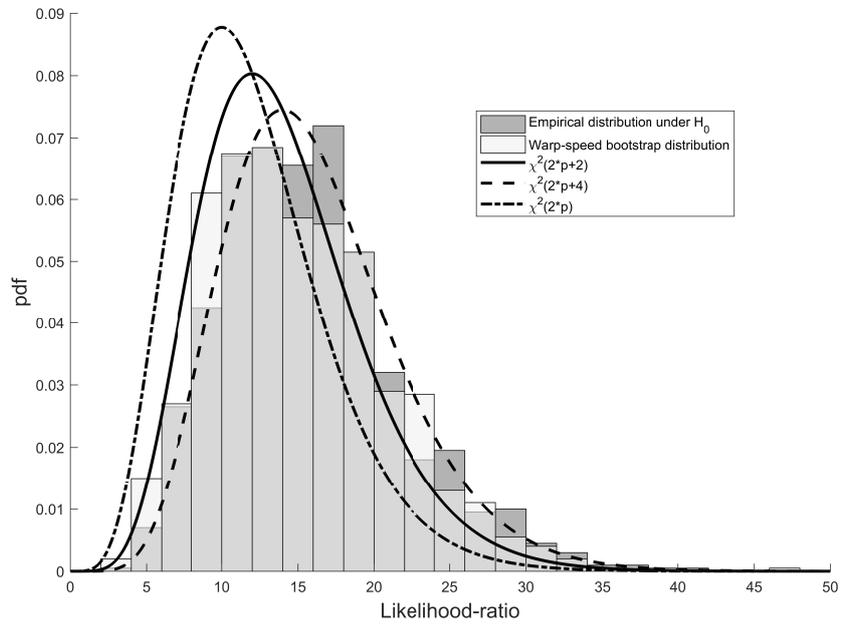


Figure 3: Comparison between the empirical distribution of the likelihood ratio (dark grey), the warp-speed bootstrap distribution (light grey), $\chi^2_{2 \times p}$ (dashed dotted), $\chi^2_{2 \times p+2}$ (solid) and $\chi^2_{2 \times p+4}$ (dashed) distributions.

4 Data

In this section, we present the characteristics of the operational loss data and of the explanatory variables used in our regression analysis.

4.1 Operational loss data at UniCredit

Operational loss data are provided by the risk department of UniCredit, one of the largest systemically important banks in Europe. UniCredit was also among the first to implement the advanced measurement approach to manage its operational risks. The data encompass losses classically defined as resulting from “inadequate or failed internal processes, people and systems, or from external events”. Further prominent examples from other institutions that made the headlines of newspapers include the \$ 6 billion of unauthorized trading losses at Société Générale, the \$ 2.4 billion paid in fines by Barclays following the LIBOR scandal; or the \$ 250 million paid by Merrill Lynch as a settlement in a gender discrimination lawsuit. We use here losses recorded over the period 2005M1 – 2014M6 and across three event types (execution, delivery, and process management – EDPM; client, product, and business practices – CPBP; external frauds – EFRAUD). Typical losses stem from transactions with counter-parties and suppliers, from the non-fulfilment of professional obligations towards clients, and from fraud, embezzlement, or violation of laws by subjects external to the bank. Initially, our database comprised 13,209, 16,138, and 6,391 losses, respectively. Loss amounts are adjusted for inflation using the Italian consumer price index and multiplied by an unknown factor for anonymity reasons. Note also that, on the time-series dimension, each loss is associated with a particular month, based on its registration date. This information is used later on to associate a set of time-varying predictors to individual losses. The monthly frequency is also our period of reference to study the dynamics in the severity distribution.

To select extreme losses, we set our threshold τ to the empirical third quartile for a given event type⁹. Although the third quartile might appear to be a low threshold from a statistical perspective, it is important to notice that the operational losses in our sample are all larger than UniCredit’s *reporting threshold*. Therefore, they are also extreme enough to be considered by UniCredit as *economically meaningful*. The use of the present threshold must be seen as a compromise between pure statistical criteria and the need to keep a representative sample of extremes. Our final samples are of size 3,288, 4,019, and 1,585, with a total sample of size 8,892. The selected extreme losses are displayed in Figure 4. In Section 5, we conduct a joint analysis accounting for potential event-type-specific dependencies.

The large dispersion of the losses makes it hard to take an intuitive look at these data. This issue highlights the need to account for these extremes to assess the dependence structure. Nevertheless, it is possible to distinguish several periods during which losses appear particularly large. For EFRAUD,

⁹Extensive discussion on threshold selection for this dataset can be found in Hambuckers et al. [2018a], who used event type-specific quantiles at the level of 75%. Owing to the complexity of the phenomenon under analysis, we believe that 75% offers a good balance between a large sample size and correct specification of extremes, as confirmed by ex-post goodness-of-fit measures.

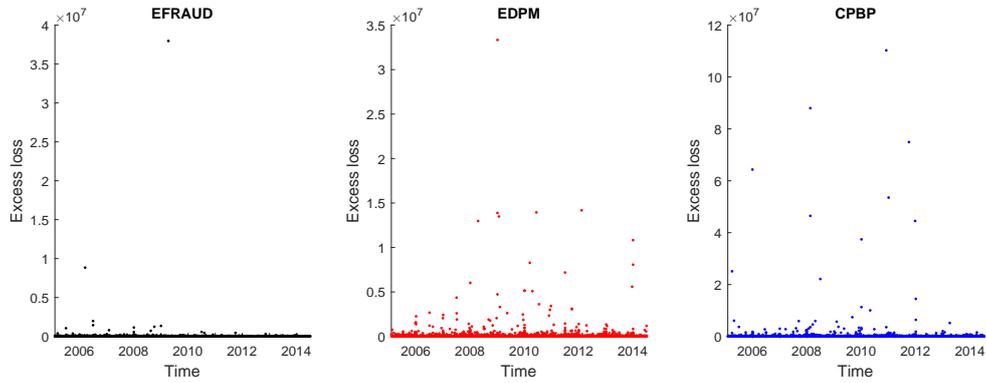


Figure 4: Excess losses, over time, for the three different event types. From left to right: EFRAUD, EDPM, and CPBP.

this is the case in 2007 and over the period 2009-2010. For EDPM, we observe an increase in the dispersion starting in 2007, peaking in 2009, and decreasing over the following years. Lastly, for CPBP, we observe several peaks: in 2007, 2009, and between mid-2010 and 2012. Overall, this suggests that some underlying process drives the surge of extremes during certain periods of time, e.g. via common risk factors or jumps.

The observed counts of the extremes are depicted in Figure 5. Overall, we observe a progressive decrease in the number of monthly losses over time as well as a clear seasonal pattern, probably a result of reporting. See Chernobai and Yildirim [2008] and Chernobai et al. [2011] as exemplary discussions of this phenomenon. The largest share belongs to the CPBP event type. We also observe a positive correlation between the number of losses for each event type, suggesting that the loss processes are dependent. This highlights the need for a joint modeling of the data.

4.2 Economic determinants and transition variable

To each extreme loss, we associate a set of risk factors registered over the previous month¹⁰. Our focus is on the number of past extreme losses, as a series of theoretical justifications have been presented to link the number of operational losses in a recent past with the severity distribution of future losses. In particular, Chernobai et al. [2011] report that most operational losses can be traced back to a *breakdown of internal controls*. Hence, if the number of losses is indicative of the quality of internal controls, it must also convey some information about the distribution of the loss sizes, given a particular quality level of internal controls. Similar reasoning is highlighted in Wang and Hsu [2013], who notice that a failed control environment is a major contributory factor to large operational losses. Therefore, we investigate this hypothesis by including this variable into the baseline part of our model. We distinguish between two situations of interest: either the existence of a *self-excitation* effect, where operational events are indicative of future larger losses; or a *self-inhibition* effect, indicating a lesser likelihood of suffering extreme losses. The former case is indicative of a persistent situation of

¹⁰Using lagged values of the predictors enables us to minimize reverse causality and endogeneity issues, as well as provide an obvious framework for prediction.

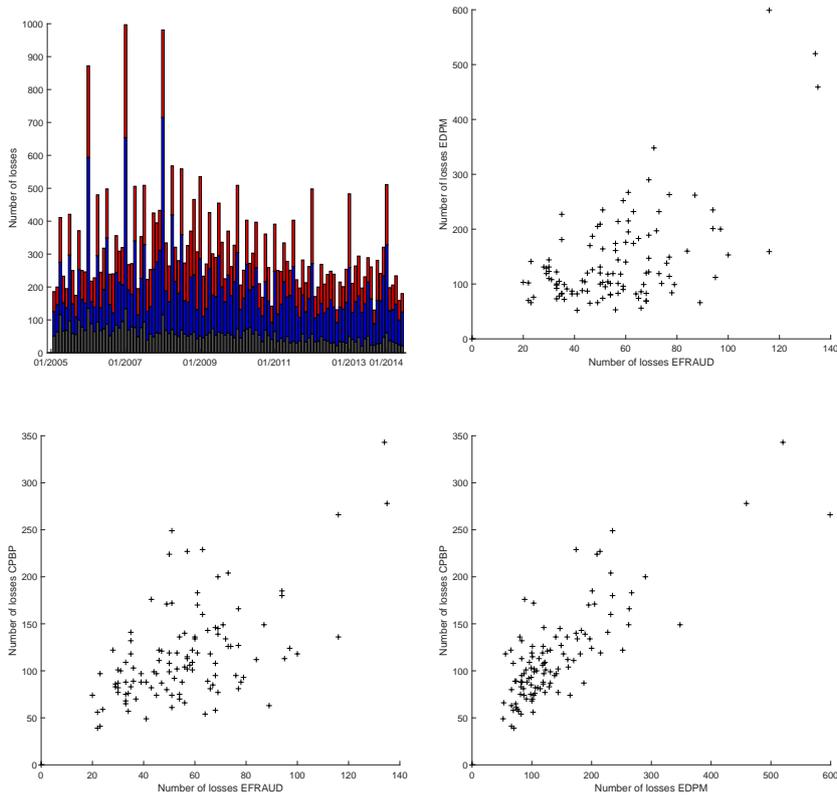


Figure 5: Number of losses in the different event types. Graph top left portrays the sum of the losses over time, on a monthly basis: grey for EFRAUD, blue for EDPM, and red for CPBP. Graphs top right to bottom right plot the number of the losses of one event type against those of another.

breakdowns in internal controls. On the contrary, the latter refers to situations in which the failure of internal controls is corrected and future loss sizes are mitigated. Whereas the relationship between severity and frequency processes has been extensively studied for insurance data [Frees and Valdez, 2008, Shi et al., 2015], it remains overlooked in the operational risk literature. This is surprising, as this feature is particularly important to establish the total operational risk capital with models based on compound processes [Brechmann et al., 2014], but also to provide risk managers with reliable indicators of future losses. To the best of our knowledge, the present study is the first to investigate this dependence in an extreme regression context. Notice that here we decompose past numbers of extreme losses into the contributions of the different event types: to a loss at time t that belongs to e.g. EFRAUD, we associate as predictor the number of EFRAUD losses registered over period $t - 1$.

For the transition variable, we use the lagged end-of-month level of the Chicago Board Options Exchange Market Volatility Index (CBOE VIX). The VIX can be seen as “the *risk-neutral* expected stock market variance for the US S&P500 contract and is computed from a panel of options prices” [Bekaert and Hoerova, 2014]. Among others, Bekaert et al. [2013] and Bekaert and Hoerova [2014] demonstrate that high values of the VIX are indicative of a high market volatility, high risk aversion, and future economic downturn. Similarly, in Adrian and Shin [2010], time periods characterized by high values of the VIX are associated with a decrease in the risk appetite of financial intermediaries owing to managerial constraints. Variation in the VIX also seems to be linked with future market liquidity, thus capturing changes in market structure [Chung and Chuwonganant, 2014]. Hence, the

VIX seems to be a good indicator of economic uncertainty and risk aversion of market participants, two factors that have an important impact on risk-taking behavior. Consequently, through the use of the VIX in our transition function as the driver of the time-varying dependence, we specify the uncertainty as the channel modulating the intensity of the link between our predictive variable (i.e. the number of losses) and the severity distribution. To limit numerical instabilities possibly caused by abrupt changes in the raw VIX data, we first apply a logarithmic transformation. This leads to distribution parameters being functions of the log-implied volatility, as in Christensen and Prabhala [1998]. In Appendix C, we investigate the sensitivity of our results to the choice of the transition variable, replacing the VIX by another measure of financial uncertainty. We discuss these results more extensively in Section 5.

Finally, relying on previous findings in Chernobai et al. [2011], Cope et al. [2012], and Hambuckers et al. [2018a], we use the following set of macro and financial control variables: European Union (EU) and Italian industrial production growth rates, the St. Louis Fed financial stress index (FSI), the economic policy uncertainty (EPU) indices of Baker et al. [2016] for the EU and Italy, and the Italian yield spread on 10-year government bond with respect to the German yield. The growth rate of industrial production is indicative of the size of transactions: in good economic times, transactions are larger, increasing the likelihood of large losses [Cope et al., 2012, Hambuckers et al., 2018a]. Booming economic conditions may also create incentives to commit larger frauds, and to increase the size of fine and compensation claims in court settlements [Povel et al., 2007]. Moreover, market sentiment, economic policy uncertainty, and credit conditions might convey information on risk aversion [Bekaert and Hoerova, 2016], future market volatility [Pastor and Veronesi, 2013], and lax monetary policy [Delis and Kouretas, 2011, Boubaker et al., 2017], three factors potentially affecting the loss formation process: a high degree of risk aversion leads to reinforced internal controls and less risky investment strategies. Market volatility generates unexpected price variations, increasing the likelihood of market crashes and making system failures or human mistakes quickly costly if they lead to market timing issues. Last but not least, a lax monetary policy drives down the banks' margins, which pushes them to soften their lending procedures and rely on more aggressive trading, with the consequence that costly operational mistakes become more likely.

We restrict our attention to this small set of predictors to limit the complexity of our model. In particular, we use aggregate indices to control indirectly for a large number of macroeconomic factors at a small cost in terms of parameters. For example, the FSI is constructed from the principal component of 18 weekly data series: seven interest rate series, six yield spreads, as well as five other macroeconomic indicators (one of which being the VIX), and measures the degree of financial stress in the markets along multiple dimensions. On the other hand, EPU is based on the newspaper coverage frequency of specific words¹¹, and proxies movements in policy-related economic uncertainty [Baker et al., 2016].

¹¹E.g. for the US, the index reflects the frequency of articles in 10 leading U.S. newspapers containing such terms as *economic*, *uncertain*, *deficit*, *Federal Reserve*, etc.

Time series of the explanatory variables are given in Figure 6. The considered time period covers both the Great Recession (characterized by high values of the VIX and a drop in industrial production growth) and the sovereign debt crisis (characterized by a high EPU and increasing yield spreads). EPU and growth rates for the EU and Italy are highly correlated. Thus, to assess the robustness of our results to issues related to the near-singularity of the predictors' matrix, we also consider models with either only EU or Italy-related variables.

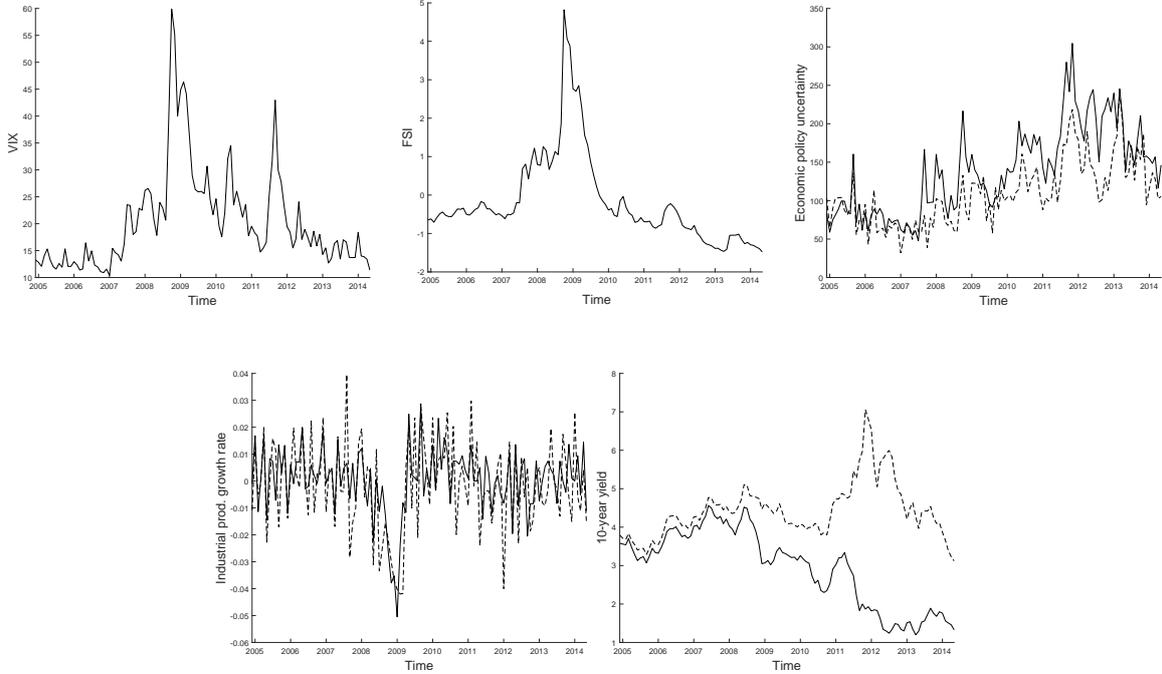


Figure 6: Measures of uncertainty on the financial markets and of the economic context. First row, from left to right: VIX, FSI, and EPU for EU (solid) and Italy (dashed). Second row: Growth rate of industrial production (monthly basis, solid: EU, dashed: Italy) and 10-year government bond yield for Germany (solid) and Italy (dashed). In the regression analysis, we use the difference between the time series.

5 Results

In this section, we estimate the proposed model on UniCredit operational losses, using the predictors described in the previous section. The predictors of the full model are given in Table 3. For γ , we restrict the predictors to the event types, whereas we include all economic variables in the structural equation of σ . The model is estimated jointly for the pooled losses from the different event types.

5.1 Dynamics in operational loss severity

We begin our analysis by discussing the estimated regression coefficients of our model. Estimates and associated confidence intervals are displayed in Table 4. Remember that the distribution parameters are given by equations (8) and (9), implying that β_2^σ (resp. β_1^σ) and β_2^γ (resp. β_1^γ) summarize the dependence structure during periods of low (resp. high) uncertainty.

Table 3: Summary of the explanatory variables used in the full model. $\text{EFRAUD}_{t,i}$ and $\text{CPBP}_{t,i}$ are binary variables indicating if the i^{th} loss at time t belongs to EFRAUD (resp. CPBP) event type; n_{t-1}^{ET} is the number of losses taking place at time $t-1$ for the event type of the i^{th} loss at time t . The product between n_{t-1}^{ET} and the event type variables leads to event- type-specific effects of past counts.

Variable	Description	Notation
s_t	Logarithm of the VIX	$\log(\text{VIX}_{t-1})$
$x_{t,i}^\sigma$	Event types	$\text{EFRAUD}_{t,i}, \text{CPBP}_{t,i}$
	Past number of losses	$n_{t-1}^{\text{ET}}, n_{t-1}^{\text{ET}} \times \text{EFRAUD}_{t,i}, n_{t-1}^{\text{ET}} \times \text{CPBP}_{t,i}$
	Economic policy uncertainty index	$\text{EPU}_{t-1}^{\text{EU}}, \text{EPU}_{t-1}^{\text{IT}}$
	Growth rate in industrial production	$\Delta \text{indprod}_{t-1}^{\text{EU}}, \Delta \text{indprod}_{t-1}^{\text{IT}}$
	Financial stability index	FSI_{t-1}
	Italian spread on gov. bonds	$\Delta \text{yield}_{t-1}$
$x_{t,i}^\gamma$	Event types	$\text{EFRAUD}_{t,i}, \text{CPBP}_{t,i}$

We find coefficients significantly different from zero in both limiting models. We also test for differences in regression coefficients across limiting models, and find significant differences (Table 4 in Appendix B). This indicates a change of dependence structure according to the level of the transition variable and the shape of the transition function. As depicted in Figure 7 right panel, the transition function is rather steep: ϕ is around 0 for VIX values up to 20, and equal to 1 for VIX values around 30 and above. This steepness implies that we are quickly in situations in which one of the two limiting models fully dominates the dependence structure. From a historical perspective, Figure 7, left panel illustrates that low uncertainty periods roughly cover the time frames 2005M1-2007M7 and 2012M7-2014M6. The high uncertainty periods are roughly 2007M8-2008M4, 2008M10-2010M2 and 2011M8-2012M1. Nevertheless, there are a number of time periods (e.g. mid-2008, mid-2010) during which the dependence structure is a mixture of both limiting models.

During the low uncertainty period, σ is rather constant: we find few effects of the covariates, in particular of n_{t-1}^{ET} . Two notable exceptions are the industrial production growth rate and the FSI. Regarding the industrial production factor, the positive sign of the regression coefficient is consistent with previous studies [Cope et al., 2012, Hambuckers et al., 2018b], suggesting that an increase in economic activity, via a mechanical increase in the size of the transactions, leads to more extreme operational losses. Similarly, a high value of the financial stress index *conditional on low values of the VIX* can be related to an increase in liquidity issues, which in turn increases the likelihood of extreme losses. Indeed, the FSI is built from seven interest rate series, six yield spreads, as well as five other macroeconomic indicators (among which is the VIX). As suggested by the time series display on Figure 6, the VIX and FSI are positively correlated. It suggests therefore that, when the VIX takes a low value, a high FSI value is primarily driven by liquidity indicators such as interest rates and spreads. Looking now at differences across event types, CPBP has the highest event-type-specific

constant term, followed by EFRAUD then EDPM. Regarding γ , we observe discrepancies across event types as well, with EDPM having the heaviest tail, followed by CPBP then EFRAUD. CPBP is thus found to be a riskier event type than EFRAUD (all else equal), a result in line with Cope et al. [2012].

Now, looking at the high uncertainty regime (i.e. for high VIX values), we observe important changes: first, n_{t-1}^{ET} exhibits a significant and negative effect. We observe some discrepancies across event types as well, probably as a result of scaling differences in the number of losses. Overall, it suggests a self-inhibiting mechanism, in which recent loss events are associated with a lesser likelihood of future extreme losses. This effect might result from the reviews of internal processes, reinforced controls, and various adjustments in the monitoring procedures that usually follow operational events. Whereas regular monitoring of the processes and corrective actions of control systems are a reality in practice, they are rarely included in statistical analyses, owing to a lack of useful data. Our study suggests that such an effect might exist and should be accounted for. Of course, the direction of the effect here is probably specific to UniCredit, and it is likely that we would observe other mechanisms (self-excitation or self-inhibition) in another financial institution. However, this result highlights that one can use the proposed approach to detect and test for the resilience of internal controls following operational events. This result is also to be put in perspective with findings in Chernobai et al. [2011] and Wang and Hsu [2013] with respect to the impact of internal controls on the frequency process. Even if we indirectly test for the same effect on the severity distribution via the use of past counts, it indicates that the use of internal control weakness indicators can prove useful to model the severity distribution.

Second, the value of γ increases for every event type when the VIX is high, indicating a higher likelihood of extreme losses when the uncertainty level is high. This result is consistent with previous findings in Chernobai and Yildirim [2008], Cope et al. [2012] and Hambuckers et al. [2018b,a], who suggest that unexpected price variations (especially of derivatives) can impact losses stemming from delivery, pricing, selling issues, or IT system crashes: in case of an operational mistake during a highly volatile time period, abrupt changes in price can easily prevent a financial institution from correcting its mistakes quickly, leading to more extreme losses.

Third, we observe negative and significant effects of $\Delta\text{indprod}_{t-1}^{EU}$, FSI_{t-1} and Δyield_{t-1} . For $\Delta\text{indprod}_{t-1}^{EU}$, we conjecture that a recession combined with a high level of uncertainty associated to higher extreme losses could be related to the counter-cyclicity of frauds, as found in Dionne and Wang [2013] for insurance claims. For the effect of FSI_{t-1} in a situation of high uncertainty, our results suggest that an increase leads to a greater likelihood of extremes. We can relate this effect to Bekaert and Hoerova [2016], who indicate that financial instability indices positively correlate with risk aversion, therefore indicating changes in risk preferences. As a consequence, a higher risk aversion combined with a high uncertainty pushes banks towards less risky investment behavior. The opposite sign of the FSI, compared to the one found in the low uncertainty regime, can be explained by the structure of the FSI: conditional on the VIX being high, high values of the FSI are mainly driven by the VIX. We can therefore postulate that risk aversion is the dominant mechanism here, and not

liquidity. This also illustrates the usefulness of a transition function based on the VIX. For Δyield_{t-1} , a decrease in spread can be associated with relatively cheap financing conditions from an Italian point of view. This phenomenon can be related to an incentivizing context for riskier investments and a "search for yields" effect, as described in Delis and Kouretas [2011]. Similar arguments are made in Bruno and Shin [2015], who highlight that a decrease in funding costs affects decisions by the banks as to how much exposure to take on, leading to greater risk taking. We can expect UniCredit, as a large European bank active in various countries, to be quite sensitive to this channel. Finally, EPU_{t-1}^{EU} seems positively related with the likelihood of large losses. This result can be analyzed in light of the findings in Pastor and Veronesi [2013]: first, they reveal that increasing values of EPU are associated with higher stock market volatility and correlation across assets. This effect would explain the positive effect on the likelihood of large losses, since high market volatility leads to unexpected price variations and more expensive operational mistakes. Second, Pastor and Veronesi [2013] also highlight that the link between economic policy uncertainty and market volatility is stronger in bad economic times. Since high VIX values (through its conditional variance component) are indicative of future recessions [Bekaert et al., 2013], the large positive effect of EPU found in the high uncertainty regime is consistent with their result.

Between these two limiting regression models, the transition component ϕ captures intermediate situations, in which uncertainty modulates the effect of the covariates. In particular, it gradually increases γ along an increase in uncertainty. From a theoretical perspective, it also combines the different channels (market volatility and transaction sizes versus laxer monetary policy and self-inhibition) to depict a more subtle situation than would have been inferred from non-ST models. Regarding the link with past losses, Figure 8 displays the time-varying effect of a change in n_{t-1}^{ET} on σ , as a function of the VIX¹². From these graphs, we can deduce regions in which σ is more sensitive to changes in n_{t-1}^{ET} i.e. when the VIX is high.

Figure 9 portrays the estimated γ and σ over time for the different ET. We also display the quantile at the level of 90% as a measure of the risk (bottom row). As highlighted previously, we observe a greater likelihood of extremes during the financial crisis (end of 2008) and the debt crises (2011 and 2012).

Are these results robust to the choice of the VIX as a transition variable? To investigate this question, we reproduce our main analysis using the lagged EPU for the US (denoted EPU_{t-1}^{US}) as our transition variable instead of the VIX. Both variables measure some forms of economic and financial stress and convey relatively similar information, as indicated by their high degree of correlation. The estimated coefficients and transition function are reported in Appendix C, Table 10, and Figure 13. We obtain qualitatively similar results in terms of signs, significance, and magnitudes of the estimated regression coefficients for the main variables. In addition, the transition function is also found to be quite steep, confirming the results of the main analysis. Overall, the estimated scale parameters are

¹²We display the marginal change, conditional on all the other covariates taking a value equal to their historical medians.

quite similar across specifications (Figure 13, right panel). Looking at the Akaike information criterion (AIC), the model based on the VIX is also found to be slightly better.

Table 4: Estimated regression coefficient obtained from the ST model 'Full'. In parentheses, estimated standard errors of the estimates. * (resp. ** and ***) denotes coefficients found significantly different from zero at the 10% (resp. 5% and 1%) test level.

Variable	β_1^σ	β_2^σ	β_1^γ	β_2^γ	\bar{g}	c
$\log(\text{VIX}_{t-1})$	-	-	-	-	0.17	3.12
					(.15)	(.04)
Constant	1.19***	.38**	0.13***	0.01	-	-
	(.34)	(.16)	(.06)	(.05)	-	-
$\text{EFRAUD}_{t,i}$	-.88*	.11	-0.50***	-0.75***	-	-
	(.47)	(.16)	(.14)	(.11)	-	-
$\text{CPBP}_{t,i}$.02	.57***	-0.20**	-0.42***	-	-
	(.40)	(.14)	(.09)	(.07)	-	-
n_{t-1}^{ET}	-2.18***	.01	-	-	-	-
	(.65)	(.19)	-	-	-	-
$n_{t-1}^{ET} \times \text{EFRAUD}$	1.63*	-.01	-	-	-	-
	(.87)	(.23)	-	-	-	-
$n_{t-1}^{ET} \times \text{CPBP}$	1.29 *	.28	-	-	-	-
	(.73)	(.21)	-	-	-	-
EPU_{t-1}^{EU}	.83**	.22	-	-	-	-
	(.42)	(.29)	-	-	-	-
EPU_{t-1}^{IT}	.81	-.02	-	-	-	-
	(.51)	(.23)	-	-	-	-
$\Delta \text{indprod}_{t-1}^{EU}$	-11.23***	4.91*	-	-	-	-
	(3.79)	(2.68)	-	-	-	-
$\Delta \text{indprod}_{t-1}^{IT}$	1.73	.26	-	-	-	-
	(2.51)	(1.96)	-	-	-	-
FSI_{t-1}	-.53***	.50*	-	-	-	-
	(.19)	(.27)	-	-	-	-
$\Delta \text{yield}_{t-1}$	-24.22***	3.66	-	-	-	-
	(6.99)	(4.15)	-	-	-	-

5.2 Model selection and specification

In this section, we investigate alternative specifications of the baseline model given in Table 3. We consider models without smooth-transition effects as well as models using only subsets of the covariates (EU variables only, Italian variables only, event-type variables only and no covariates). Our goal is to determine if the full ST model from Table 3 must be favored over more parsimonious specifications.

5.2.1 Do ST models provide a better fit to the data?

First, we apply the likelihood ratio test described in Section 2.3 to compare nested alternative specifications. For every pair of models, we reject the null of equal goodness-of-fit in favor of the ST model (right panel, Table 5). This is a first indication that ST models should be preferred over their non-ST

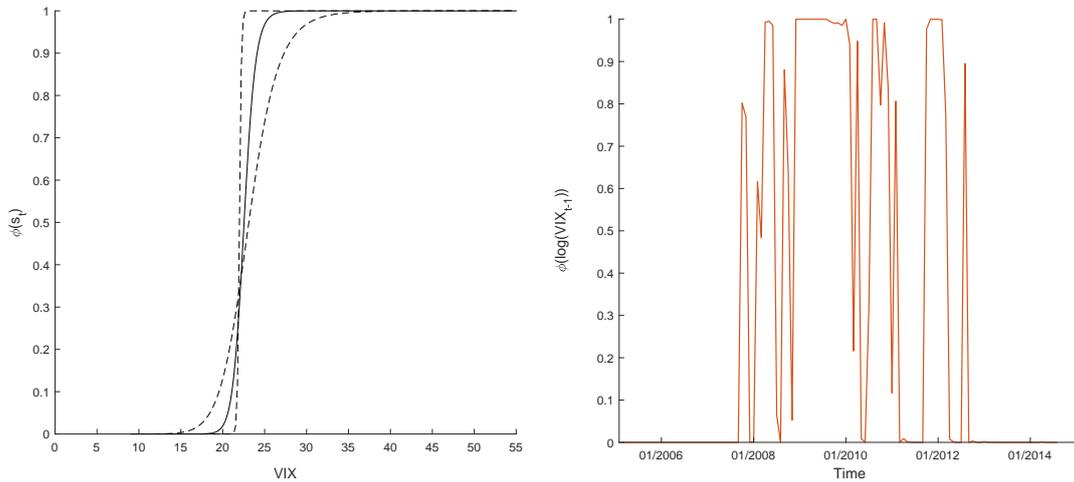


Figure 7: Value of the transition function ϕ as a function of the VIX (left) and over time (right). Dashed lines denote 50% asymptotic confidence intervals.

competitors. In addition, when comparing ST models between each other (left panel, same table), the full ST model is found superior to all alternatives except the ST model with EU covariates.

Second, we look at the censored likelihood score (CLS) described in Section 2.4: a smaller CLS is indicative of better goodness-of-fit, especially for tail data. Computing CLS_{α}^{CV} with α ranging between .5 and .99 (Figure 10), the conclusions are unequivocal: ST models have the smallest CLS for every censoring level and must be preferred over non-ST models when modeling extreme data. Figure 10 portrays the differences in CLS (as a percentage) of the alternative specifications with respect to the full ST model: a positive value indicates that the full ST model is better. For increasing values of α , the full ST model becomes even more advantageous. This result again highlights that the distribution of extremes is better modeled with the ST approach and that the smooth transition is an important component. As robustness checks, we repeat these operations using event-type-specific quantiles for the censoring thresholds (Figure 12 in Appendix B). We come to the same conclusion. We also test for a significant difference in CLS between the various models using the stepwise superior predictive ability (SPA) bootstrap test of Hsu et al. [2010] on the individual censored scores. This test is useful to control for data snooping issues. We find the full ST model, as well as the ST models with EU-only and Italy-only covariates to be significantly better than all non-ST models at the 5% test level. With this test, we also conclude that no model can beat the full ST model. Moreover, the full model and the one with EU-only covariates are both found to be significantly better than the ST model with Italy-only covariates. These results hold for various values of α between .5 and .95 (see Appendix B for the complete results).

We conclude from this exercise that the smooth-transition component is an important feature of the data. Regarding the various subsets of covariates considered, we see that the full ST model and the one with EU-only covariates provide significantly better fits in terms of CLS than all of the other specifications tested.

Finally, one could question the suitability of the smooth-transition structure with respect to simpler threshold models, since the steep transition function might indicate over-parametrization. To

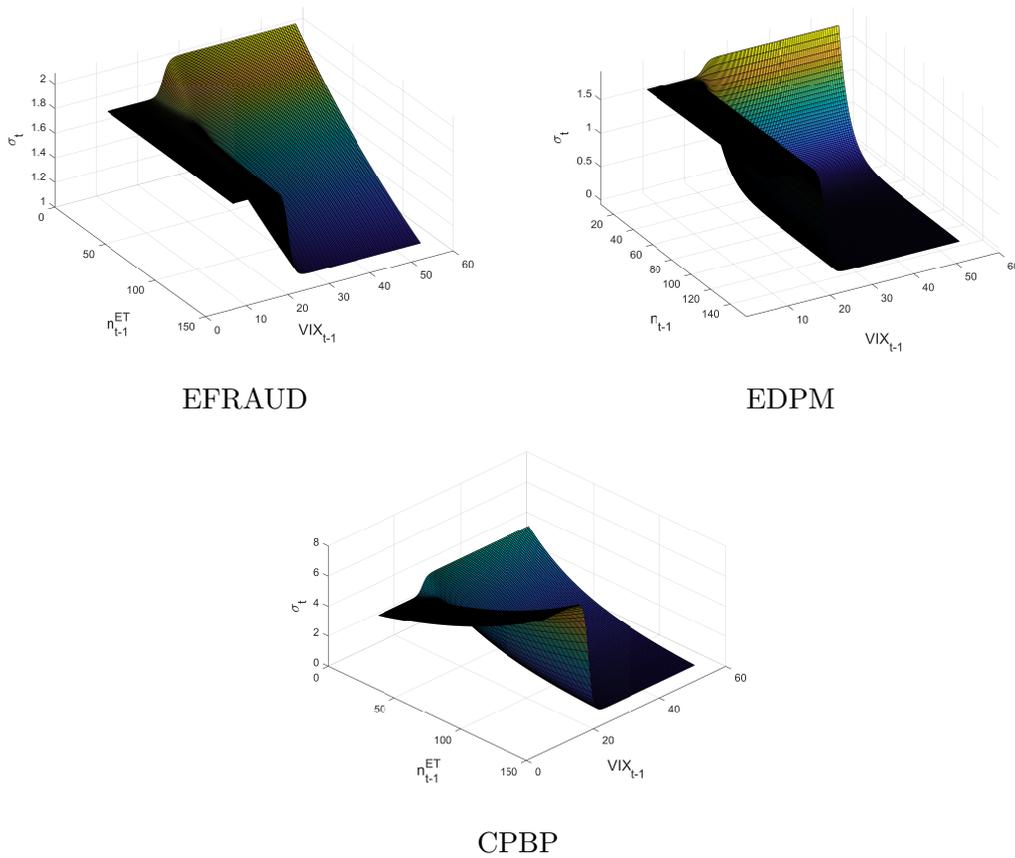


Figure 8: Response surface of parameter $\sigma_{t,i}$ as a function of past count (n_{t-1}^{ET}) and lagged value of the VIX, for EFRAUD, EDPM, and CPBP.

stress this hypothesis, we estimate three additional models, characterized by a threshold structure: a threshold-GP regression model, where the logistic function is replaced by a step function, and two GP regression models with additional predictors based on an indicator function of the log-VIX. Our conclusions stay alike: the ST-GPD model provides a better fit compared to these new alternatives, suggesting that the smooth-transition specification is more suitable than a step function. Detailed results of this analysis can be found in Appendix D. It is also interesting to note that over the time periods 2007M8-2008M1, 2008M5-2008M9, and 2009M12-2010M12 (overall 19 months out of 114, characterized by an important financial instability), the results of the ST-GPD model indicate that the dynamics are driven by various mixtures of the two limiting regression models (Figure 15, Appendix D). On the contrary, the results provided by the threshold approach imply that a same limiting regression model describes the severity distribution for the whole period. Thus, with the help of the ST-GPD model, we are able to grasp the different dynamics at play during these periods. In particular, our results suggest a certain heterogeneity across crisis periods, a feature not possible with the threshold model. Looking at a risk indicator such as the VaR at the level of 99.9% of the severity distribution, the threshold model also has the effect of generating almost systematically larger risk estimates (for example, for EFRAUD, see Figure 16 in Appendix D), whereas it might not be justified from a statistical point of view.

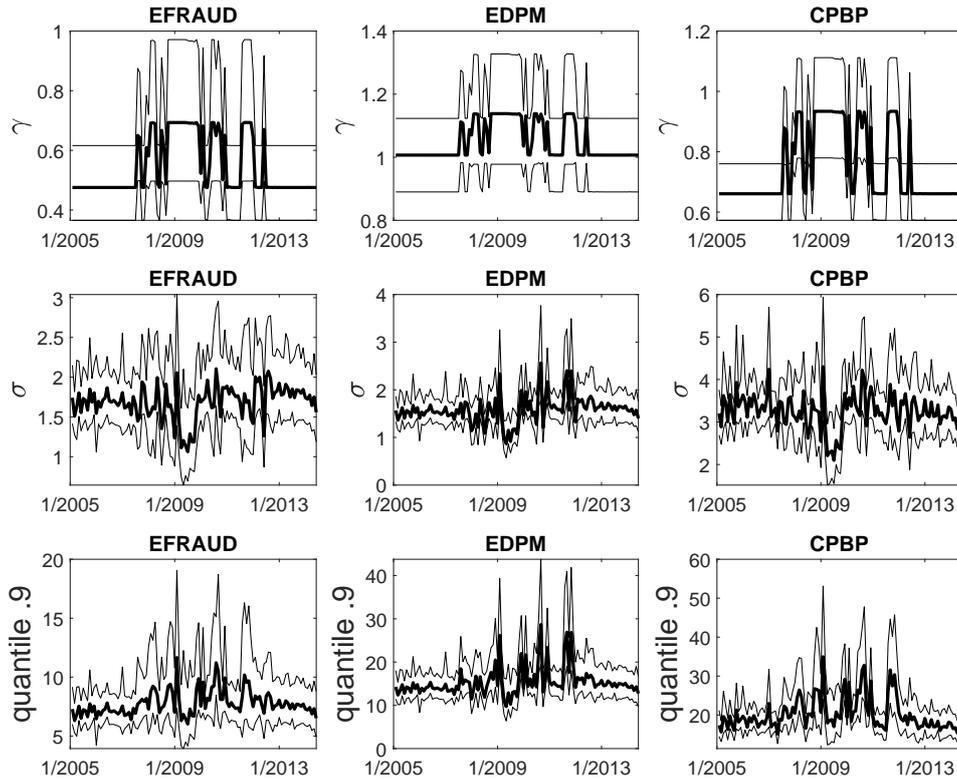


Figure 9: From top to bottom: estimates of γ , σ and the 90% quantile, respectively. Thick line: estimated parameters over time. Thin line: 99% confidence intervals obtained from the posterior simulation procedure of Wood [2006], for a fixed transition function.

5.2.2 Do ST models provide correct Value-at-Risk estimates?

Here, we test the correct specification of the various models, using VaR estimations and coverage tests for the total loss distribution (see Section 2.4). Table 6 displays the results of the LR_{MUC} test defined in (15), obtained from the observed number of exceptions and super exceptions for all models and various VaR levels (see Appendix B, Table 9 for a detailed presentation of the number of exceptions). Non-ST models lead to larger numbers of rejections, at all levels, compared to ST models. In particular, we reject the correct specification of all non-ST models at the level .99, whereas this is not the case for the ST models. Except for $\alpha_1 = .95$ and $\alpha_2 = .975$ (first row), the correct specifications of the full ST model and the ST model with EU-only covariates are never rejected at the 5% test level.

Hence, the full ST model seems appropriate for a bank to infer its operational risk capital reserves. Figure 11 displays the total loss and the VaR at the level of .975 for the full and ET-only ST model, as well as the full non-ST model. This illustrates that the alternative models imply either too large or too small VaR calculations, leading to a costly insurance mechanism or unwanted exceptions.

Table 5: p-values of the likelihood ratio tests. Left panel: the null model (on the horizontal line) is an ST model, and is tested against an alternative ST model (first column). Right panel: the null model is a GP regression model without ST structure, and is tested against an alternative ST model. For these tests, owing to the identification issue under H_0 discussed in Section 2.3, rejection regions are corrected accordingly.

	H0: ST models				H0: non-ST models				
H1/H0	EU	Italy	ET only	Cst	Full	EU	Italy	ET only	Cst
Full	.52	< .01	< .01	< .01	< .01	< .01	< .01	< .01	< .01
EU	-	-	< .01	< .01	-	< .01	-	< .01	< .01
Italy	-	-	.003	< .01	-	-	< .01	< .01	< .01
ET only	-	-	-	< .01	-	-	-	< .01	< .01

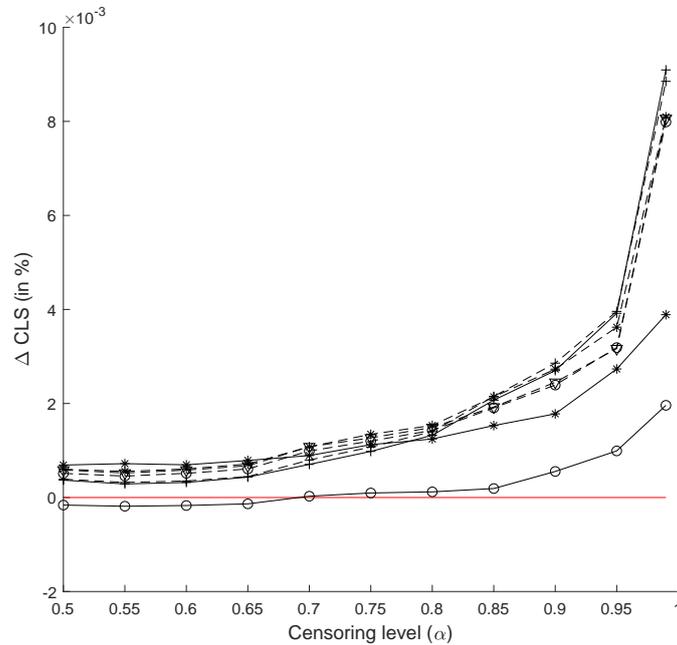


Figure 10: Differential in cross-validated CLS, as a percentage of CLS_{α}^{CV} for the ST model 'Full'. Positive differences imply a less favorable fit for the alternative model. Solid (resp. dashed): ST (resp. non-ST) models. +: 'ET only'; *: 'Italy', o: 'EU', ∇ : 'Full'. Red: ST model 'Full'. 'Cst' models have been omitted for legibility.

Table 6: Test statistic $LR_{MUC}(\alpha_1, \alpha_2)$ of the multiple LR test of Colletaz et al. [2013], using various values for α_1 and α_2 . In parentheses, p-values obtained from a simulation procedure with 10^6 samples of length $T = 113$.

Multivariate unconditional coverage test										
(α_1, α_2)	ST models					non-ST models				
	Full	EU	Italy	ET only	Cst	Full	EU	Italy	ET only	Cst
(.95, .975)	4.54 (.039)	4.54 (.039)	9.05 (.003)	1.9 (.314)	4.543 (.042)	6.10 (.016)	7.12 (.009)	6.10 (.016)	6.10 (.016)	7.12 (.009)
(.95, .99)	2.89 (.101)	2.89 (.101)	4.25 (.056)	2.29 (.156)	3.63 (.086)	5.94 (.019)	5.94 (.019)	5.94 (.019)	5.94 (.019)	5.94 (.019)
(.975, .99)	0.88 (.168)	0.88 (.168)	0.27 (.457)	0.84 (.233)	6.75 (.009)	2.21 (.116)	2.47 (.069)	2.21 (.116)	2.21 (.116)	6.57 (.011)
(.975, .995)	0.30 (.301)	0.30 (.301)	1.17 (.16)	0.45 (.301)	4.51 (.033)	1.40 (.123)	0.50 (.236)	2.45 (.08)	2.45 (.08)	6.67 (.007)
(.99, .995)	0.55 (.235)	0.55 (.235)	0.55 (.235)	0.06 (.533)	0.02 (.900)	2.49 (.024)	2.49 (.024)	2.49 (.024)	2.49 (.024)	2.49 (.024)

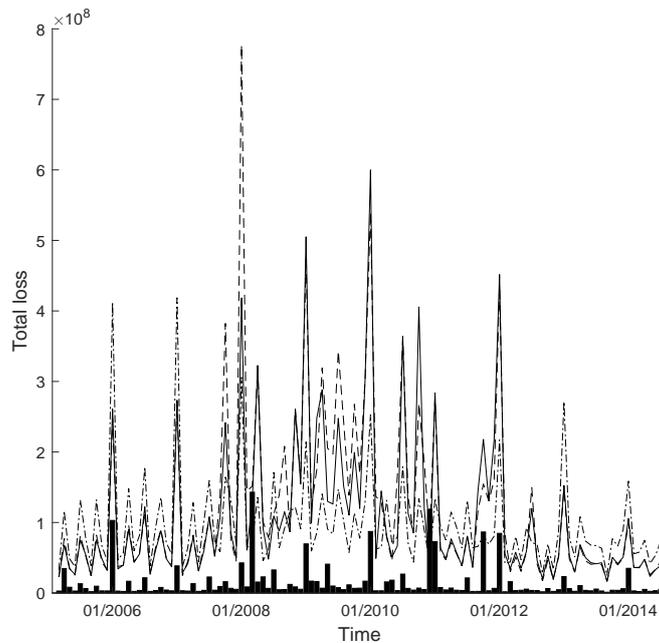


Figure 11: Comparison between various VaR estimates at the level .975 and the total loss distribution (black bar chart). Solid: ST model 'Full'. Dashed: ST model 'ET only'. Dashed-dotted: non-ST model 'Full'.

6 Conclusion

We introduce a smooth-transition generalized Pareto regression model, useful for handling the time-varying effect of risk factors on the severity distribution of financial losses. This model has the advantages of accounting explicitly for the high probability of extreme events and for a change in effects of the explanatory variables over time. In a simulation study, we highlight the good properties of the proposed estimation and testing procedures. Then, we use our model to conduct an empirical study of the dynamics driving operational losses severity at UniCredit. We focus on connecting the severity distribution of operational losses and the past number of losses (i.e. the frequency process), with the idea that past operational events proxy the quality of internal controls. As transition variable, we use the VIX, assuming that the uncertainty on the financial markets influences the link between the severity and frequency processes. We find that two different limiting mechanisms drive the severity distribution: in high uncertainty periods, we observe that a high number of operational events is followed by less extreme losses. This result suggests a self-inhibition effect, i.e. that the monitoring and supervision processes following operational events at UniCredit mitigate the likelihood of extremes in subsequent periods. In addition, during such periods of high uncertainty, an increase in the financial stability index (FSI), the Italian yield spread, and the industrial production growth rate are linked with a decrease in the likelihood of extreme losses. In light of these effects, we conjecture that these variables are proxies for an increase in risk aversion, a tight monetary policy, and the counter-cyclical nature of fraud losses, respectively. On the contrary, in periods of low uncertainty, only positive economic growth rate and FSI are significantly associated with more severe losses. Potential explanations are related to the effect of economic growth on transaction sizes and to the impact of low liquidity on timing issues. Several periods in our sample are driven by mixtures of the two limiting regimes, suggesting that a continuous transition function is necessary to model the data correctly. Finally, we demonstrate that the smooth-transition components improve the goodness-of-fit with respect to simpler alternatives.

Our findings have several implications from a risk-management perspective: first, measures of internal control stringency provide useful indications for the future likelihood of extreme operational events. However, their effectiveness can vary according to the uncertainty level. Therefore, including these indicators in an early-warning system as soon as the markets become turbulent could help for finer risk management. In particular, if a mechanism of self-inhibition exists, it provides counter-cyclical capital requirements that account for risks reductions after operational events. Second, operational risk models need to include a time-varying dependence structure like a smooth-transition feature to provide adequate risk estimates of the total loss distribution. Neglecting this effect would lead to too many VaR breaches or too much capital being reserved.

Several extensions of the present paper may be considered in future work. First, we could extend the proposed smooth-transition structure to time-series tail models, such as the dynamic conditional score model proposed by Massacci [2017] and the realized-POT model of Bee et al. [2019]. By doing so, we would combined the best of two worlds (time-series and regression models for extremes), allowing

simultaneously for flexible structural changes over time and along covariates. This would solve a limitation of the present work, since we do not take the dependence on the time dimension explicitly into account. Second, instead of focusing on GP regression models, we could extend mutually self-exciting processes [Grothe et al., 2014] to the case of time-varying decay or impact functions depending on a transition variable. The merit of this approach over ST-GP regression models would be to model the severity-frequency dependence explicitly, particularly if covariates are included in the severity parameters. Third, we could also relax the linearity assumption on the limiting regression regimes and instead consider combinations of two non-parametric regression structures. This would decrease the likelihood of any mis-specification problems.

Finally note that, although the advanced measurement approach (AMA) has recently been removed from the regulatory framework of Basel III [Basel Committee on Banking Supervision (BCBS), 2016], we strongly believe that the present findings and statistical approaches are useful for a forward-looking approach to operational risks. As highlighted in Peters et al. [2016], the new standardized approach suffers from important drawbacks, notably through the assumption that all operational losses follow a single generating mechanism (Section 4, same reference). Therefore, it seems important not to base risk management solely on regulatory considerations, but also on sound statistical analyses.

Funding

Thomas Kneib acknowledges the financial support from the Deutsche Forschungsgemeinschaft (German Research Foundation, DFG) within research project KN 922/9-1. Julien Hambuckers acknowledges the financial support of Research Training Group 1644 funded by the DFG, and of the National Bank of Belgium. Parts of the research toward this paper were conducted while JH was postdoctoral researcher at the University of Göttingen (Germany), Chair of Statistics.

References

- T. Adrian and H.S. Shin. Liquidity and leverage. *Journal of Financial Intermediation*, 19(3):418–437, 2010.
- S.R. Baker, N. Bloom, and S.J. Davis. Measuring economic policy uncertainty. *The Quarterly Journal of Economics*, 131(4):1593–1636, 2016.
- 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). Standardised Measurement Approach for operational risk. Technical report, Bank of International Settlements, Basel, Switzerland, 2016.
- M. Bee, D. Dupuis, and L. Trapin. Realized Peaks over Threshold: A Time-Varying Extreme Value Approach with High-Frequency-Based Measures. *Journal of Financial Econometrics*, 17(2):254–283, 03 2019.
- G. Bekaert and M. Hoerova. The VIX, the variance premium and stock market volatility. *Journal of Econometrics*, 183(2):181–192, 2014.
- G. Bekaert and M. Hoerova. What do asset prices have to say about risk appetite and uncertainty? *Journal of Banking & Finance*, 67:103 – 118, 2016.
- G. Bekaert, M. Hoerova, and Duca M.L. Risk, uncertainty and monetary policy. *Journal of Monetary Economics*, 60(7):771–788, 2013.
- 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.
- V. Bruno and H.S. Shin. Capital flows and the risk-taking channel of monetary policy. *Journal of Monetary Economics*, 71:119–132, 2015.
- F. Chan and M. MacAleer. Maximum likelihood estimation of STAR and STAR-GARCH models: theory and Monte Carlo evidence. *Journal of Applied Econometrics*, 17(5):509–534, 2002.
- F. Chan and M. MacAleer. Estimating smooth transition autoregressive models with GARCH errors in the presence of extreme observations and outliers. *Applied Financial Economics*, 13(8):581–592, 2003.
- F. Chan and B. Theoharakis. Estimating m-regimes STAR-GARCH model using QMLE with parameter transformation. *Mathematics and Computers in Simulation*, 81(7):1385–1396, 2011.
- V. Chavez-Demoulin, A.J. McNeil, and A.C. Davison. Estimating value-at-risk: a point process approach. *Quantitative Finance*, 5(2):227–234, 2005.
- V. Chavez-Demoulin, P. Embrechts, and J. Neslehova. Quantitative models for operational risk: Extremes, dependence and aggregation. *Journal of Banking & Finance*, 30(10):2635 – 2658, 2006.
- 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 and Y. Yildirim. The dynamics of operational loss clustering. *Journal of Banking & Finance*, 32(12):2655–2666, 2008.

- A. Chernobai, P. Jorion, and F. Yu. The determinants of Operational Risk in U.S. financial institutions. *Journal of Financial and Quantitative Analysis*, 46(8):1683–1725, 2011.
- B.J Christensen and N.R. Prabhala. The relation between implied and realized volatility. *Journal of Financial Economics*, 50(2):125–150, 1998.
- K.H. Chung and C. Chuwongnant. Uncertainty, market structure, and liquidity. *Journal of Financial Economics*, 113(3):476–499, 2014.
- G. Colletaz, C. Hurlin, and C. Pérignon. The risk map: A new tool for validating risk models. *Journal of Banking & Finance*, 37(10):3843 – 3854, 2013.
- E. Cope, M. Piche, and J. Walter. Macroeconomic determinants of operational loss severity. *Journal of Banking & Finance*, 36(5):1362–1380, 2012.
- 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.
- P. De Fontenay, V. DeJesus-Rueff, J. Jordan, and E. Rosengren. Capital and risk: New evidence on implications of large operational losses. *Journal of Money, Credit and Banking*, 38(7):1819–1846, 2006.
- M.D. Delis and G.P. Kouretas. Interest rates and bank risk-taking. *Journal of Banking & Finance*, 35(4):840–855, 2011.
- 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.
- G. Dionne and K. Wang. Does insurance fraud in automobile theft insurance fluctuate with the business cycle? *Journal of Risk and Insurance*, 47(1):67–92, 2013.
- M. Dungey, G. Milunovich, S. Thorp, and M. Yang. Endogenous crisis dating and contagion using smooth transition structural GARCH. *Journal of Banking & Finance*, 58:71 – 79, 2015.
- P. Embrechts, C. Kluppelberg, and T. Mikosch. *Modelling extremal events for insurance and finance*. Springer - Verlag, Berlin, 1997.
- E.W. Frees and E.A. Valdez. Hierarchical insurance claims modeling. *Journal of the American Statistical Association*, 89(425):208–218, 2008.
- R. Giacomini, D.N. Politis, and H. White. A warp-speed method for conducting Monte Carlo experiments involving bootstrap estimators. *Econometric Theory*, 29(3):567–589, 2013.
- B. Gnedenko. Sur la distribution limite du terme maximum d’une série aléatoire. *Annals of Mathematics. Second Series*, 44(2):423–453, 1943.

- T. Gneiting and R. Ranjan. Comparing density forecasts using threshold-and quantile-weighted scoring rules. *Journal of Business & Economic Statistics*, 29(3):411–422, 2011.
- O. Grothe, V. Korniichuk, and H. Manner. Modeling multivariate extreme events using self-exciting point processes. *Journal of Econometrics*, 182(2):269 – 289, 2014.
- M. Guidolin. Markov Switching Models in Empirical Finance. In D.M. Drukker, editor, *Missing Data Methods: Time-Series Methods and Applications (Advances in Econometrics, Vol. 27 Part 2)*, pages 1–86. 2011.
- A. Guillou, S. Loisel, and G. Stupfler. Estimation of the parameters of a Markov-modulated loss process in insurance. *Insurance: Mathematics and Economics*, 53(2):388–404, 2013.
- J. Hambuckers, A. Groll, and T. Kneib. Understanding the Economic Determinants of the Severity of Operational Losses: A regularized Generalized Pareto Regression Approach. *Journal of Applied Econometrics*, 33(6):898–935, 2018a.
- J. Hambuckers, T. Kneib, R. Langrock, and A. Silbersdorff. A Markov-Switching Generalized Additive Model for Compound Poisson Processes, with Applications to Operational Losses Models. *Quantitative Finance*, 18(10):1679–1698, 2018b.
- B. Hansen. Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica*, 64(2):413–430, 1996.
- P.-H. Hsu, Y.-C. Hsu, and C.-M. Kuan. Testing the predictive ability of technical analysis using a new stepwise test without data snooping bias. *Journal of Empirical Finance*, 17:841–862, 2010.
- M. Kratz. Normex, a new method for evaluating the distribution of aggregated heavy tailed risks. *Extremes*, 17(4):661–691, 2014.
- R. Lukkonen, P. Saikkonen, and T. Teräsvirta. Testing Linearity Against Smooth Transition Autoregressive Models. *Biometrika*, 75(3):491–499, 1988.
- D. Massacci. Tail risk dynamics in stock returns: Links to the macroeconomy and global markets connectedness. *Management Science*, 63(9):3072–3089, 2017.
- L. Pastor and P. Veronesi. Political uncertainty and risk premia. *Journal of Financial Economics*, 110(3):520 – 545, 2013.
- G.W. Peters, P.V. 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(3):1–49, 2016.
- P.C.B. Phillips and J.H. Lee. Predictive regression under various degrees of persistence and robust long-horizon regression. *Journal of Econometrics*, 177(2):250 – 264, 2013.

- J. Pickands. Statistical inference using extreme order statistics. *The Annals of Statistics*, 3(1):119–131, 1975.
- M.D. Porter and G. White. Self-exciting hurdle models for terrorist activity. *The Annals of Applied Statistics*, 6(1):106–124, 2012.
- P. Povel, R. Singh, and A. Winton. Booms, Busts, and Fraud. *Review of Financial Studies*, 20(4):1219–1254, 2007.
- P. Sands, G. Liao, and Y. Ma. Rethinking operational risk capital requirements. *Journal of Financial Regulation*, 4(1):1–34, 2018.
- S. Self and K.-Y. Liang. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, 82(398):605–610, 1987.
- P. Shi, X. Feng, and A. Ivantsova. Dependent frequency-severity modeling of insurance claims. *Insurance: Mathematics and Economics*, 64:417–428, 2015.
- A. Silvennoinen and T. Teräsvirta. Modeling Multivariate Autoregressive Conditional Heteroskedasticity with the Double Smooth Transition Conditional Correlation GARCH Model. *Journal of Financial Econometrics*, 7(4):373–411, 2009.
- T. Teräsvirta. Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models. *Journal of the American Statistical Association*, 89(425):208–218, 1994.
- D. van Dijk, T. Teräsvirta, and P.H. Franses. Smooth transition autoregressive models - a survey of recent developments. *Econometric Reviews*, 21(1):1–47, 2007.
- Quang H. Vuong. Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses. *Econometrica*, 57(2):307–333, 1989.
- T. Wang and C. Hsu. Board composition and operational risk events of financial institutions. *Journal of Banking & Finance*, 37(6):2042–2051, 2013.
- S. N. Wood. *Generalized Additive Models: An Introduction with R*. Chapman & Hall, Boca Raton, Florida, 2006.

A Computational details

In this appendix, we provide additional details on the proposed maximization procedure. Firstly, as mentioned in Section 2.2, we re-parametrize the transition function in terms of (\bar{g}, c) , where $\bar{g} = 1/\sqrt{g}$, $g > 0$. This alternative specification has the advantage of making the likelihood function less flat in

the dimension of the transition parameters. This is particularly needed in the present context of GPD data, since the GPD is also known to exhibit a flat likelihood function. Equation (5) is now given by

$$\phi(s_t) = \frac{1}{1 + \exp\left\{-\frac{1}{(\bar{g})^2}(s_t - c)\right\}}. \quad (22)$$

Secondly, we set boundary conditions on the location parameters of the transition function. For the location parameter c , we impose that it cannot fall far from the range of values observed for s_t . Our additional constraint can be expressed as

$$c \in (1.025 \min(\mathbf{s}), .975 \max(\mathbf{s})).$$

Notice also that the positivity constraint on \bar{g} is explicitly taken into account as a bound constraint.

Solutions of the maximization problems are obtained via numerical procedures based on the `fmincon` function of MATLAB with the `trust-region` algorithm. Notice here that we work with the likelihood of the observed exceedances, given a threshold structure. Thus, to include the uncertainty related to threshold estimation, one would have to resort to more complex bootstrap procedures. We make the choice of neglecting this additional uncertainty. This problem is a long-standing issue in EVT, which we leave to further research.

With respect to inference on the regression coefficients, we use the inverse of the variance-covariance matrix of the log-likelihood as an approximate asymptotic covariance matrix. Owing to the use of the `fmincon` option, the returned Hessian is not that of the log-likelihood function, but of the Lagrangian. Therefore, to obtain an approximation of the Hessian of the log-likelihood function, we perform one additional step with the help of the `fminunc` function: we start from our constrained estimate and run one iteration of the `trust-region` algorithm. Then, the Hessian is obtained from finite-difference procedures.

B Additional results

We now provide additional results, highlighting the robustness of our findings to the correct specification of the different tests.

First, Table 4 displays the results of the test for significant differences across limiting regression models. Then, we provide the computed values for the CLS statistics (Figure 12) using an event-type-specific quantile as an alternative threshold. The ranking between models remains unchanged and we observe the same increase of the gap between non-ST and ST models when the censoring threshold increases. We also test for a significant difference in terms of CLS using the stepwise-SPA test of Hsu et al. [2010], for different values of α . Table 8 displays the p-values obtained through the tests. Starting with the non-ST model 'Full' as benchmark, we find all ST models superior to the non-ST models. Then, employing sequentially the worst ST models as benchmark, we conclude that the ST model 'Full' is significantly better than all models for at least some values of α .

Finally, in Table 9 we provide the number of VaR breaches suffered with the various models. The aim here is to investigate which model provides the most accurate risk measures. Once again, ST

models appear to provide better VaR breaches than their non-ST counterparts.

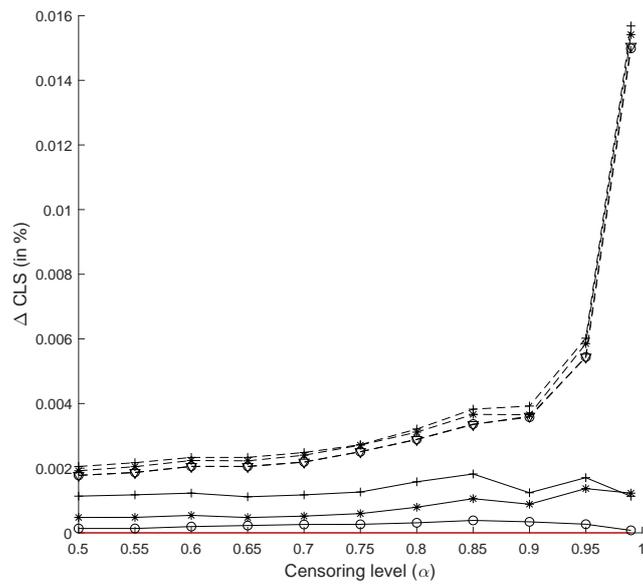


Figure 12: $\kappa(\alpha)$ is computed here as an event-type-specific quantile of order α . Differential in CLS (given by equation (13)), as a percentage of the CLS for the ST model 'Full'. Positive differences imply a less favorable fit for the alternative model. Solid (resp. dashed): ST (resp. non-ST) models. +: 'ET only'; *: 'Italy', o: 'EU', ∇ : 'Full'. Red: ST model 'Full'.

Table 7: Wald test of a significant difference between regression coefficients of the different limiting regimes. We report the observed difference and corresponding estimated standard errors. *, ** and *** indicate rejections of the null of no differences at the 10%, 5%, and 1% test levels, respectively.

Variable	$\Delta\beta^\sigma$	$\Delta\beta^\gamma$
Cst	0.81*** (0.39)	0.123 (0.07)
EFRAUD _{t,i}	-0.991** (.5)	0.254 (0.18)
CPBP _{t,i}	-0.554 (.43)	0.224* (.12)
n_{t-1}^{ET}	-2.188*** (.7)	-
$n_{t-1}^{ET} \times \text{EFRAUD}$	1.644* (.91)	-
$n_{t-1}^{ET} \times \text{CPBP}$	1.008 (.78)	-
EPU_{t-1}^{EU}	0.605 (.51)	-
EPU_{t-1}^{IT}	0.826 (.56)	-
$\Delta\text{indprod}_{t-1}^{EU}$	-16.145*** (4.71)	-
$\Delta\text{indprod}_{t-1}^{IT}$	1.472 (3.36)	-
FSI _{t-1}	-1.035*** (.32)	-
Δyield_{t-1}	-27.878*** (8.31)	-

Table 8: P-values of the stepwise SPA test of Hsu et al. [2010]. We test the null that no alternative model is better, in term of mean censored score, than a benchmark model. Only the p-values for the models with a better score are reported. We perform this test with various benchmarks: non-ST Full, ST Italy and ST EU. We use $B = 2,000$ resamples and the same MATLAB implementation as Hsu et al. [2010]

ST models					
Benchmark	α	Full	EU	Italy	ET only
non-ST Full	.5	.006	.008	.023	.210
	.75	.006	.010	.024	.100
	.90	.026	.036	.067	.097
	.95	.028	.036	.102	.127
	.99	.065	.068	.086	.088
ST Italy	.5	0.012	.097	-	-
	.75	0.012	.188	-	-
	.90	0.005	.099	-	-
	.95	< .001	.001	-	-
	.99	0.126	.152	-	-
ST EU	.5	0.101	-	-	-
	.75	0.004	-	-	-
	.90	0.01	-	-	-
	.95	0.131	-	-	-
	.99	0.812	-	-	-

Table 9: Number of exceptions, i.e. number of time periods during which the historical total loss is larger than the estimated quantile at level α .

Number of exceptions										
α	ST models					non-ST models				
	Full	EU	Italy	ET only	Cst	Full	EU	Italy	ET only	Cst
.9	18	18	18	18	19	17	16	16	16	18
.95	10	10	11	9	10	12	12	12	12	12
.975	3	3	2	4	7	5	4	5	5	8
.99	2	2	2	1	1	3	3	3	3	3
.995	1	1	1	0	0	1	1	2	2	2
.999	0	0	0	0	0	0	0	0	0	0

C Alternative transition variable

In this appendix, we investigate the robustness of our results to the choice of the transition variable. We report the estimated model obtained when using the EPU for the USA (denoted EPU_{t-1}^{US}) instead of the VIX as a transition variable (Table 10). Both variables measure some forms of economic and financial stress, with a focus on the US economy. Correlation over the considered time period is .588, suggesting that both series partly convey the same information.

Signs, significance, and magnitudes of the coefficients for the main variables are comparable to those obtained with the VIX. As for the results presented in Section 5, the estimated transition function is quite steep (Figure 13, left panel). The estimated scale parameters are also alike (Figure 13, right panel, displays a scatter plot of these estimates obtained with the two models). Regarding the global fit, we achieve slightly better results (in terms of information criteria) when using the VIX: the model based on the VIX exhibits an AIC of 46,450 versus 46,464 for the model based on EPU_{t-1}^{US} .

Table 10: Estimated regression coefficient obtained from the ST model 'Full' but replacing the transition function by $\log(EPU_{t-1}^{US})$. In parentheses, estimated standard errors of the estimates. * (resp. ** and ***) denotes coefficients found significantly different from zero at the 10% (resp. 5% and 1%) test level.

Variable	β_1^σ	β_2^σ	β_1^γ	β_2^γ	\bar{g}	c
$\log(EPU_{t-1}^{US})$	-	-	-	-	.09	4.69
	-	-	-	-	(.05)	(.01)
Cst	.73***	.40**	0.09*	0.01	-	-
	(.21)	(.19)	(.05)	(.05)	-	-
EFRAUD $_{t,i}$	-.41	.26	-0.81***	-0.56***	-	-
	(.25)	(.22)	(.12)	(.11)	-	-
CPBP $_{t,i}$.15	.70***	-0.23***	-0.42***	-	-
	(.21)	(.19)	(.07)	(.08)	-	-
n_{t-1}^{ET}	-1.4***	.06	-	-	-	-
	(.35)	(.23)	-	-	-	-
$n_{t-1}^{ET} \times$ EFRAUD	1.03**	-.28	-	-	-	-
	(.50)	(.93)	-	-	-	-
$n_{t-1}^{ET} \times$ CPBP	1.05 **	.11	-	-	-	-
	(.42)	(.26)	-	-	-	-
EPU_{t-1}^{EU}	.95***	-.46	-	-	-	-
	(.31)	(.45)	-	-	-	-
EPU_{t-1}^{IT}	.23	.44	-	-	-	-
	(.26)	(.35)	-	-	-	-
$\Delta \text{indprod}_{t-1}^{EU}$	-2.30	6.39*	-	-	-	-
	(2.88)	(3.32)	-	-	-	-
$\Delta \text{indprod}_{t-1}^{IT}$	-.80	.96	-	-	-	-
	(1.96)	(2.02)	-	-	-	-
FSI $_{t-1}$	-.15	.75***	-	-	-	-
	(.13)	(.26)	-	-	-	-
$\Delta \text{yield}_{t-1}$	-13.53***	2.33	-	-	-	-
	(3.97)	(7.28)	-	-	-	-

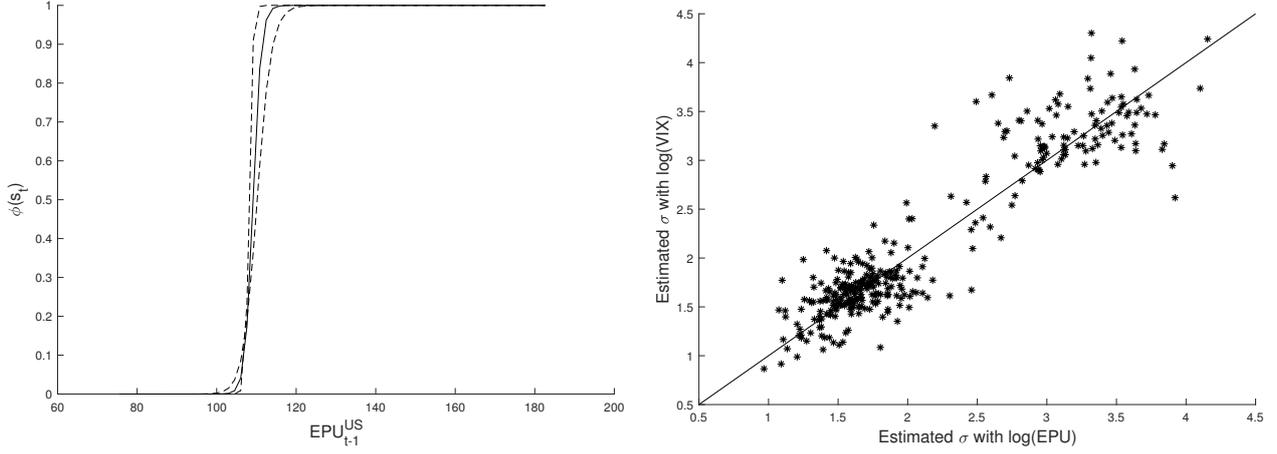


Figure 13: Value of the transition function ϕ as a function of the EPU_{t-1}^{US} . Dashed lines denote 50% asymptotic confidence intervals.

D Alternative transition mechanism

The results presented in Section 5 suggest a steep transition function when using the log of the VIX as a transition variable. Therefore, the proposed smooth-transition structure could be over-parametrized, since we perfectly discriminate between two limiting regression models beyond a certain threshold of the log-VIX. To stress this hypothesis, we consider two alternative models:

- a threshold-GP regression model, in which the logistic function of our ST-GPD model is replaced by an indicator function. In this model, the transition function is now given by

$$\phi(s_t) = \mathbb{1}(s_t > \tau^*),$$

where $\mathbb{1}(\cdot)$ is the usual indicator function taking value 1 if the condition in parentheses is met, 0 otherwise. τ^* is the threshold parameter, estimated by maximum likelihood procedures with the other regression parameters.

- a standard GP regression model, but with additional predictors based on an indicator function of the log-VIX. That is, we add a variable among the set of predictors that takes value 1 if the log-VIX is above a given threshold τ^* , and 0 otherwise. We also consider interactions of this variable with all the other predictors, allowing for two different regression models with a global intercept. This model is estimated by maximum likelihood procedures for fixed values of τ^* , covering a large grid of values.

Figure 14 (top left) displays the value of the log-likelihood function for the different models. It can be observed that the highest value of the log-likelihood is achieved with the ST-GPD model. Accounting for complexity, the smallest AIC is obtained with the ST-GPD model as well (same figure, top right)¹³. Then, approximately treating these models as strictly non-nested and non-overlapping¹⁴, we

¹³We favor AIC over BIC here because it is not clear in the present setting what sample size must be considered. Unreported results suggest that penalizing the degree of freedom by $\log(n)$ is overly conservative.

¹⁴This is not exactly the case for the threshold-GPD and ST-GPD models, since for $c = \tau^*$ and $g \rightarrow +\infty$, the

test for significant differences in goodness-of-fit using the test of Vuong [1989]. We reject the null of identical Kullback-Leibler distances (KLD) with respect to the true model, in favor of a significantly smaller KLD for the ST-GPD model. This result is obtained for almost all cases considered (Figure 14, bottom row). Figures 15 and 16 highlight differences in terms of transition functions and 99.9% quantiles between the threshold GP regression and the ST-GPD models. In particular, we see that the ST-GPD model gives a more nuanced picture during crisis periods.

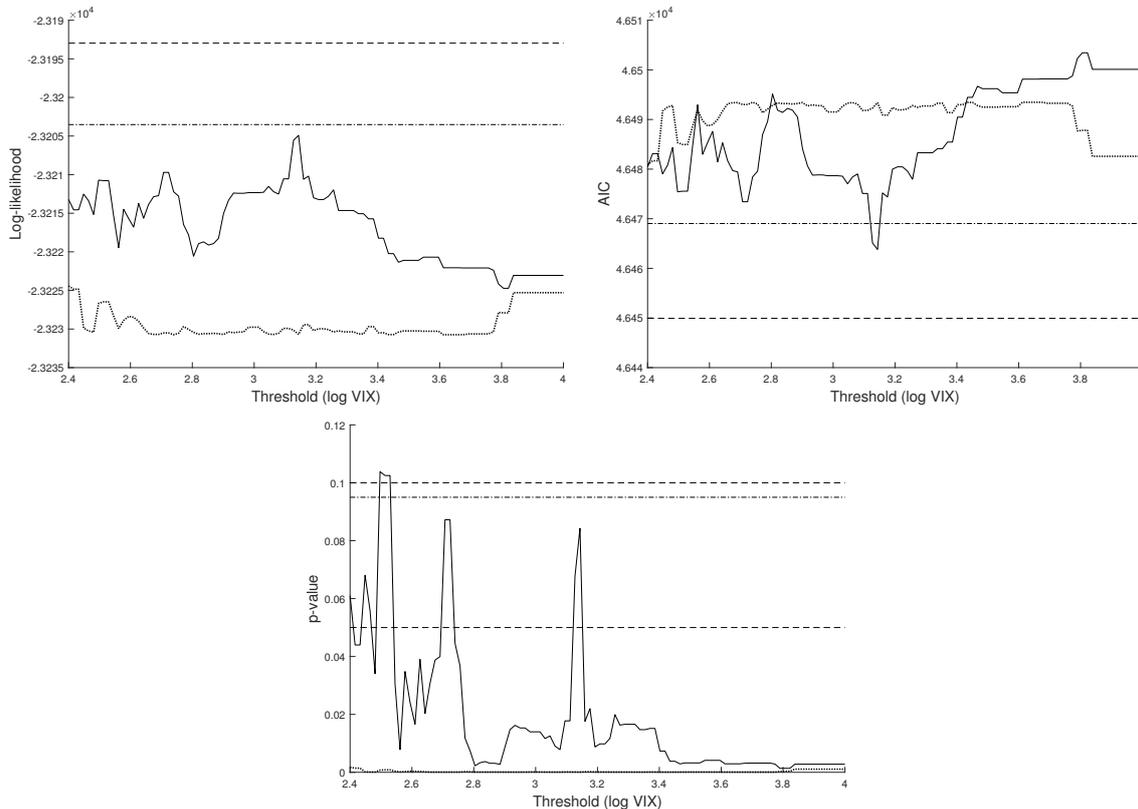


Figure 14: Dotted: Value of the log-likelihood function (top left) and the AIC (top right) of the baseline GP-regression model with a single indicator function of the log-VIX as additional predictor. Dashed: ST-GPD model. Solid: GP-regression model with interactions. Dashed-dotted: threshold-GP model. Bottom row: p-values of Vuong’s test comparing the ST-GPD models with the other models.

structural equations for the distribution parameters become identical. However, β_1^σ and β_1^γ would be unidentified under those constraints, making the problem non-standard. Therefore, if we assume g to be finite, the models can be treated as non-overlapping and Vuong’s results hold. Regarding the comparison with the GP regression models, it is clear from equations (8) to (12) that a change in the set of predictors prevents the models from being nested. However, if the regression parameter associated with the indicator function is constrained to zero and the smooth-transition structure does not exist, then the models overlap but the ST-GPD model is unidentified. We assume that such restrictions cannot apply. Hence, the results here must be seen as purely indicative.

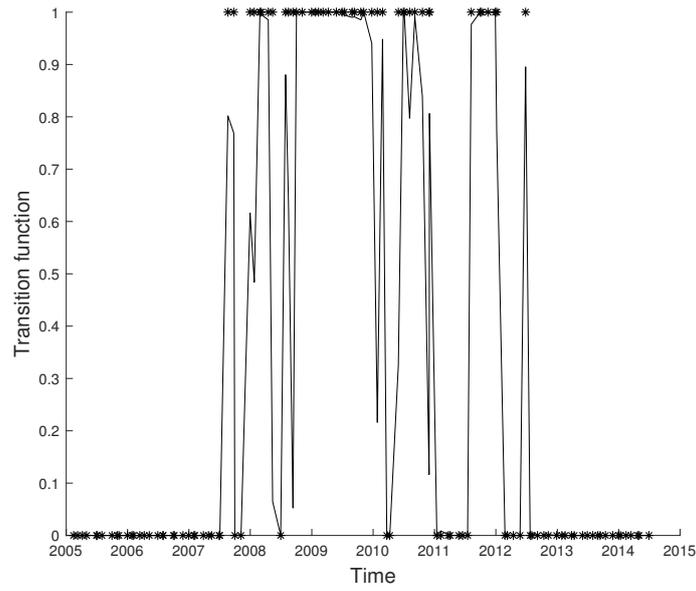


Figure 15: Solid line: transition function for the ST-GPD model. Stars: transition function for the threshold-GP regression model.

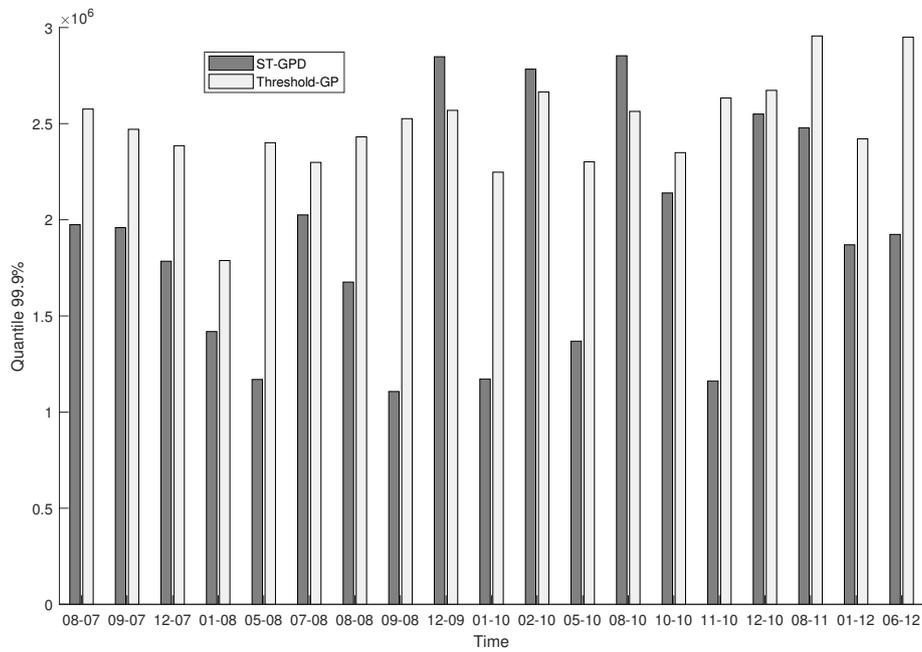


Figure 16: 99.9% quantile of the severity distribution for EFRAUD, at various dates where $0 < \phi(st) < 1$. Dark grey: ST-GPD model. Light grey: threshold GP regression model.