Operational risk, uncertainty, and the economy: a smooth transition extreme value approach.

J. Hambuckers[1,2,†] and T. Kneib[2]

Abstract

We study the link between the distribution of extreme operational losses and the economic context, a fundamental task to compute adequate risk measures over time. In particular, we allow for time-varying dependencies due to structural changes, thanks to a newly-introduced smooth transition Generalized Pareto (GP) regression model. In this model, the parameters of the GP distribution are related to explanatory variables through regression functions, which depend themselves on a predictor of structural changes. Relying on this model, we study the dependence of the monthly loss severity distribution at UniCredit, over the period 2005-2014. As indicator of structural changes, we use the VIX, accounting for the general uncertainty on financial markets. We show that both the goodness-of-fit far in the tail and Value-at-Risk estimates of the total loss distribution obtained from such models are superior to a set of alternatives. We also show that in periods of high uncertainty, conditions favorable to a lax monetary policy are synonym of an increased likelihood of extreme losses. Finally, we discover evidence of a self-inhibition mechanism, where a high number of losses in a recent past are indicative of less extreme losses in the future, probably due to improved monitoring.

Keywords: Extreme value theory; generalized Pareto distribution; operational loss; smooth transition; VIX.

JEL: C24, C46, C58, G21.

[1] University of Liege, HEC Liege, 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

Early on, operational risk has been acknowledged as a major concern: a survey in 2003 by the Basel Committee for Banking Supervision (BCBS) revealed that a large share - around 15% - of the total capital of banks was held against operational risks [De Fontenouvelle et al., 2006]. Since then, this importance remained: 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]. Hence, because of its high consumption of capital, the proper modeling and understanding of the operational loss phenomenon is a crucial economic challenge for banks. This paper contributes to these two aspects, and provides a set of recommendations useful for the decision making process of risk managers.

Recently, the operational risk literature[1] gained interest in the dependence between the severity distribution and economic factors [Cope et al., 2012, Chavez-Demoulin et al., 2016, Hambuckers et al., 2018b,a]. Indeed, the economic context is expected to impact the incentives and the allocation of resources inside banks, which in turn impact the loss formation process. Therefore, studying the nature and the drivers of this dependence structure is fundamental to design proper risk models. In particular, it allows to control for time-inhomogeneity of the severity distribution and to improve distribution forecasts. In addition, it proves helpful to meet two important requirements of the regulators: accounting for dependencies across event types (ET), as well as between severity and frequency distributions. The former requirement is particularly relevant since banks are expected to compute an aggregate risk measure over all ET [Chapelle et al., 2008, Shi et al., 2015]. On the other hand, the latter requirement proves useful to improve total loss distribution estimations based on compound processes, as it is traditionally the case for operational losses [Garrido et al., 2016].

A way to study this dependence consists in using extreme value analysis (EVA) and its extension to generalized Pareto (GP) regression as in Chavez-Demoulin et al. [2016] and Hambuckers et al. [2018a]. The extensive use of EVA in this area is due to two major facts: on the one hand, the empirical distribution of these losses is particularly heavy-tailed. On the other hand, practitioners are mostly interested in the probability of extreme losses that can bankrupt the bank. Consequently, an extensive literature studies the operational loss phenomenon relying on the Peak-over-Threshold (POT) approach (see e.g. Aue and Kalkbrener [2006], Chapelle et al. [2008], Soprano et al. [2009], Chavez-Demoulin et al. [2016]). In this approach, an extreme event is defined as a value that exceeds some high threshold, and only the exceedances over this threshold are used for statistical inference [Rootzen et al., 2018]. Thanks to a series of theoretical results, it can be shown that the natural distribution for these exceedances is a GP distribution [Balkema and de Haan, 1974, Pickands, 1975, Davison and Smith, 1990]. In the context of operational losses, this

[1] Operational risk relates to operational losses, defined as *"direct or indirect losses resulting from inadequate or failed internal processes, people and systems or from external events"* [Basel Committee on Banking Supervision (BCBS), 2004].

distribution is called the *severity distribution* whereas the number of exceedances follow a *frequency distribution*. In the natural extension to the GP regression case, we assume additionally that the distribution parameters depend on covariates through a structural equation, in the spirit of the generalized additive model for location, scale and shape (GAMLSS) of Rigby and Stasinopoulos [2005]. Hence, by the introduction of common risk factors, we can easily create an implicit dependence structure across ET. In addition, by using the number of claims as a risk factor itself, we allow for an indirect dependence structure between the severity and the frequency distributions, meeting the second requirement of the regulators.

Nevertheless, a major drawback of this technique is the implied assumption of a constant dependence structure over time. In practice, one often observes fundamental events, like financial crises, important political, managerial and regulatory decisions that alter how the loss formation process is impacted by changing economic conditions. For instance, Hambuckers et al. [2018b] found that the relationship between the severity distribution and economic indicators is different before and after the financial crisis. Similarly, Power [2005] notices that managers can actively improve controls of the business process after loss events, thereby influencing the severity of future operational events.

To deal with this issue, Ang and Timmermann [2012] suggest using regime-switching models. This approach has been successfully applied to model market volatility [Klaassen, 2002], interest rates [Pesaran et al., 2006] and the size of insurance claims [Guillou et al., 2013]. In the specific context of operational losses, Hambuckers et al. [2018b] consider a semiparametric regime-switching GP regression model. However, the regime-switching framework suffers from three important limitations. First, the switching between regimes relies on a Markovian hypothesis. Aside from being hardly true in practice, this hypothesis does not tell us anything about the reasons of a structural change. Hence, both the analysis and the communication of such a model are difficult. Secondly, it is difficult to consider more than a few regimes, due to cumbersome estimation of these models. Third, the whole framework implies abrupt changes of regression models from one period to another, ignoring possible transitory states. From a risk management perspective, such models have the undesirable effect to provide extremely volatile risk indicators.

In light of these considerations, the primary purpose of this paper is to provide an adequate econometric set-up that solves the mentioned issues, and to unravel some of the time-varying dependence structures underlying monthly operational loss data in a typical bank.

In a first step, we introduce a novel framework for the study of extremes, that relaxes the assumptions made on the transitions between regression models. To do so, we rely on the concepts of smooth transition first introduced in Teräsvirta [1994]: the regression equation of the distribution parameters is defined by a mixture of two limiting regression models, with mixing weights depending on a time-varying covariate. The main advantage of our specification is that, contrary to the regime-switching model, we allow for an infinity of regimes while keeping a low number of parameters. It drastically enhances the flexibility

of the estimation, whereas the usual two-state regime is included as a special case. In addition, the conditional likelihood function is readily available, allowing us to avoid the use of expectation-maximization algorithm or forward recursion, procedures that suffer from computational constraints [Langrock et al., 2017]. An important difference with switching regime models, though, is the need for explicitly defining the factors driving the transitions. However, whereas it exposes us to a potential misspecification 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 allow to formulate *ex ante* an hypothesis about the switching process, which can be used for interpretation. In particular, we can disentangle the loss formation process between a baseline process i.e. a structural component linking determinants to the loss severity distribution, and an amplification mechanism that exacerbates or smoothes the effect of changes in the determinants.

In the following, we refer to this new severity distribution model as ST-GPD (for smooth transition GP distribution). The concept of smooth transition has been previously applied, among others, to the study of gross domestic product (GDP) growth rate [Florackis et al., 2014], stocks comovements [Chelley-Steeley et al., 2013] and exchange rate market [Kilian and Taylor, 2003], but never for 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, 2003, Chan and Theoharakis, 2011], we detail a number of useful tools to reduce numerical instabilities. In addition, we propose a likelihood-ratio test to test the correct specification of ST-GPD against simpler alternatives, tackling identification issues under the null hypothesis. In a simulation study, we show that the combination of these tools allows for a good and stable estimation of the considered smooth transition models.

In a second step, we conduct an empirical study of the monthly time-varying dependence structure between economic factors and the severity distribution of extreme operational losses. To do so, we study a sample of extreme operational losses from the Italian bank UniCredit[2], recorded over the period 2005M1 - 2014M6 and across three event types (execution, delivery and process management; client, product and business practices; external frauds). Whereas previous literature studied yearly or quarterly periods of time to match regulators' horizon, we focus here on a monthly horizon to detect structural change relevant at the level of a financial institution. As explanatory variables, we use a set of macro, financial and firm-specific variables. In particular, we introduce a dependence with the past number of operational losses, relaxing the usual assumption of full independence between severity and frequency processes. Whereas other variables can be seen more as control variables, the past number of losses is seen here as the core of the baseline process, and as a proxy for the efficiency of internal controls. As an indicator of structural changes in the transition function, we use the (log) Volatility Index (VIX), drawing on previous

[2]Some of these data have been used, e.g. in Hambuckers et al. [2018b] and Hambuckers et al. [2018a], but considering only quarterly dependencies.

findings that highlight the forward-looking ability of this index. This specification allows us to specify the uncertainty on the financial market as the channel through which changes in internal controls affect differently the severity over time. Thus, our analysis focuses on the three following aspects: relevance of the proposed smooth transition structure with respect to simpler alternatives, magnitude and direction of the effect of past counts (as a proxy of internal controls) on the severity, and history of this dependence over time.

Using a battery of econometric tests, we show that the ST-GPD specification significantly improves the fit of the severity distribution over time with respect to simple GP regression models. Especially, it consistently provides better density estimates in the tail of the distribution. Lastly, the ST-GPD is shown to provide risk indicators for the total loss distribution with better properties, especially in terms of coverage probabilities far in the tail.

We discuss the economic significance of our results and show major changes in the relationship between the severity distribution and the economic context. We show that during time periods of high financial uncertainty, the severity distribution is characterized by a self-inhibition effect, suggesting that months with a high number of losses are predictive of future months with low severity losses. Such an effect can be explained by a tightening of control procedures following operational events in a stressed financial context for the firm. At the contrary, periods of low uncertainty exhibit no self-inhibition effects, and only the macro-financial context seems to influence the severity. In addition, our findings suggest that in a high uncertainty context, situations associated with laxer monetary policies leads to a higher risk of large losses. On the other side, in low uncertainty situations, the likelihood of extreme losses will positively depend on financial instability i.e. liquidity effects and transaction sizes implied by economic growth.

Our contribution can be summarized by the following points:

- We provide a detailed econometric approach to study time-varying dependencies in the context of extreme regression models.
- We show that ST models provide better fits to our data than non-ST models. In particular, our approach provides better estimates of risk indicators like quantiles far in the tail for the total loss distribution.
- We find that in times of low uncertainty, transaction sizes and financial instability plays an important role whereas in the high uncertainty regime, situations characterized by laxer monetary policies and low financing costs are synonym of a higher likelihood of large losses.

The structure of the paper is as follows: In Section 2, we provide the various models and estimation procedures used for analysing the severity distributions. In Section 3, we study the finite sample properties of the proposed methodology in a realistic Monte Carlo simulation. Then, in Section 4, we present the data and discuss in Section 5 the results of our empirical study, discussing 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 define first the model, before discussing estimation, inference and testing 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 [3]. The index i goes from 1 to n_t, n_t being the observed number of losses in period t larger than τ. We denote by $\mathbf{n} = \{n_t\}_{t=1,...,T}$ the time series of observed counts up to T. We also define the exceedances as $Y_{t,i} = Z_{t,i} - \tau$. Relying on extreme value theory (EVT) and the POT approach, the natural distribution of $Y_{t,i}$ is the GPD. Its cumulative distribution function (cdf) is given by

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

with $y \geq 0$ (we follow the same notation as in Hambuckers et al. [2018a]). $\gamma \in \mathbb{R}$ and $\sigma > 0$ are the shape and scale parameters, respectively. For $\gamma < 0$, then $0 < y < -\sigma/\gamma$ and y is bounded. In the case of $\gamma = 0$, we observe an exponential decay. For $\gamma \in (0, 1)$, it can be shown that Y has a finite first moment [Chavez-Demoulin et al., 2016]. This is often a desirable property from a methodological perspective (e.g. for moment-based inference) as well as for applications involving the computation of a conditional expectation. We restrict our attention to the case where $\gamma > 0$ (i.e. the heavy-tail case).

This GPD approximation stems from fundamental results in extreme value analysis, known under the names Gnedenko and Pickands-Balkema-De Haan theorems [Gnedenko, 1943, Balkema and de Haan, 1974, Pickands, 1975]. The GPD approximation holds when the cdf of Z, $F(z)$, is of the type

$$F(z) = (z)^{-1/\gamma}G(z), \tag{2}$$

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

$$\lim_{z\to\infty} \frac{G(\omega z)}{G(z)} = 1, \tag{3}$$

for $\omega > 0$. This condition, known as *the regular variation* property, implies that the tail of the loss distribution decays at a power rate of Z and that the GPD is the limiting distribution of the exceedances [see, e.g. Embrechts et al., 1997, Davison and Smith, 1990, for theoretical details]. In practice, it is assumed 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].

[3] The 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 with γ and σ following a smooth transition regression model. This model implies that both distribution parameters are a combination of two limiting regression models. Thus, for the i^{th} exceedance taking place over the time period $]t-1,t]$, we assume that

$$Y_{t,i} \sim GPD(Y_{t,i}; \gamma(\mathbf{x}_{t,i}^{\gamma}, s_t), \sigma(\mathbf{x}_{t,i}^{\sigma}, s_t)), \tag{4}$$

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 $\mathbf{x}_{t,i}^{\sigma}$, $\mathbf{x}_{t,i}^{\sigma}$ and s_t are the vector of economic covariates for the shape parameter, the scale parameter and the transition variable, respectively. $\gamma(\mathbf{x}_{t,i}^{\gamma}, s_t)$ 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. Following the idea of Teräsvirta [1994] and van Dijk et al. [2007], we characterize the shape and scale parameters by the following smooth transition structure:

$$\begin{aligned}
\phi(s_t) &= \frac{1}{1+\exp\{-g(s_t-c)\}}, &(5)\\
\gamma(\mathbf{x}_{t,i}^{\gamma}, s_t) &= \phi(s_t) h\left((\mathbf{x}_{t,i}^{\gamma})^T \boldsymbol{\beta}_1^{\gamma}\right) + (1-\phi(s_t)) h\left((\mathbf{x}_{t,i}^{\gamma})^T \boldsymbol{\beta}_2^{\gamma}\right), &(6)\\
\sigma(\mathbf{x}_{t,i}^{\sigma}, s_t) &= \phi(s_t) h\left((\mathbf{x}_{t,i}^{\sigma})^T \boldsymbol{\beta}_1^{\sigma}\right) + (1-\phi(s_t)) h\left((\mathbf{x}_{t,i}^{\sigma})^T \boldsymbol{\beta}_2^{\sigma}\right), &(7)
\end{aligned}$$

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). For s_t continuous, we have an infinity of regimes. To ease the notation, the explicit references to $\mathbf{x}_{t,i}^{\sigma}$, $\mathbf{x}_{t,i}^{\sigma}$ and s_t are omitted in the following such 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, potentially leading to an easier estimation of these parameters. Lastly, notice that we apply the transition function on $h(x^T\boldsymbol{\beta})$ and not on $x^T\boldsymbol{\beta}$. This structure facilitates the interpretation of the model, allowing us to see the terms $h(x^T\boldsymbol{\beta})$ as the value of the distribution parameter in the limiting regimes.

2.2 Estimation procedure

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

$$\mathcal{L}(\mathbf{y}; \Theta, \mathbf{x}, \mathbf{s}, \boldsymbol{\tau}, \mathbf{n}) = \sum_{t=1}^{T} \sum_{i=1}^{n_t} \log(gpd(y_{t,i}; \gamma_{t,i}, \sigma_{t,i})), \tag{8}$$

where *gpd* denotes the probability density function (pdf) of the GPD, Θ is the set of all parameters of the model; $\mathbf{y}$ the vector of all losses, $\mathbf{x}$ a design matrix of all observed covariates, $\mathbf{s}$ the vector of transition variables, τ the threshold parameter and $\mathbf{n}$ the observed

frequency process. Notice that (8) relies on the *contemporaneous* conditional independence between frequency and severity, implying that $n_t \perp y_{t,i}$ given the covariates. This assumption allows splitting the full likelihood into a severity part and a frequency part, and to study the two processes separately [Chavez-Demoulin et al., 2016]. However, it does not prevent $n_{t-1} \not\perp y_{t,i}$. Therefore, in $\mathbf{x}$, we can include n_{t-1}. An estimator $\hat{\Theta}$ of the regression parameters is obtained by maximizing equation (8) with respect to Θ:

$$\hat{\Theta} = \arg\max_{\Theta} \{\mathcal{L}(\mathbf{y}; \Theta, \mathbf{x}, \mathbf{s}, \tau, \mathbf{n})\}. \tag{9}$$

This maximization is 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 $\bar{g} = 1/\sqrt{g}$, $g > 0$. This alternative specification has the advantage to make the likelihood function less flat in the dimensions 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

$$\phi(s_t) = \frac{1}{1 + \exp\{-\frac{1}{(\bar{g})^2}(s_t - c)\}}. \tag{10}$$

Secondly, following Silvennoinen and Teräsvirta [2009], 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. Since this is the first time that a ST-GP regression model is proposed, in Section 3 we perform a simulation study to assess the reliability of the estimation method. We show that, with our procedure, we estimate well the different coefficients and there is no need to resort to two-step procedures[4].

Regarding inference on the regression coefficients, we assume that asymptotic normality of (9) holds, and we use the inverse of the variance-covariance matrix of the log-likelihood as asymptotic covariance matrix[5]. Due to the use of the `fmincon` option,

[4]Notice that here 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 choose here to neglect this additional uncertainty. This problematic is a long-standing question in EVT, that we let for further researches.

[5]See Chan and MacAleer [2002] for a formal proof in the case of STAR-GARCH models.

the returned Hessian is not the one 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. Coverage properties of this approach are examined in the simulation study. Overall, Wald-type tests show excellent sizes and satisfactory powers. Notice that the asymptotic normality is likely to hold only if the model is well identified, i.e. if $\boldsymbol{\beta}_1 = (\boldsymbol{\beta}_1^{\gamma}, \boldsymbol{\beta}_1^{\sigma}) \neq (\boldsymbol{\beta}_2^{\gamma}, \boldsymbol{\beta}_2^{\sigma}) = \boldsymbol{\beta}_2$. Thus, a first step would be to test for a correct specification of the time-varying structure. This testing issue is discussed in the Section 2.3.

2.3 Testing for a time-varying structure

The use of smooth transition models inevitably raises questions about the correct specification of an additional time structure. Early on, Lukkonen et al. [1988], Teräsvirta [1994] and van Dijk et al. [2007] discussed testing for linearity against alternative ST models. This task is drastically complicated by the presence of unidentified nuisance parameters under the null hypothesis of no ST specification. Indeed, for our model, notice that equations (6) and (7) can be rewritten as

$$\gamma(\mathbf{x}_{t,i}^{\gamma}, s_t) = h\left((\mathbf{x}_{t,i}^{\gamma})^T \boldsymbol{\beta}_2^{\gamma}\right) + \phi(s_t)\left(h\left((\mathbf{x}_{t,i}^{\gamma})^T \boldsymbol{\beta}_1^{\gamma}\right) - h\left((\mathbf{x}_{t,i}^{\gamma})^T \boldsymbol{\beta}_2^{\gamma}\right)\right), \quad (11)$$

$$\sigma(\mathbf{x}_{t,i}^{\sigma}, s_t) = h\left((\mathbf{x}_{t,i}^{\sigma})^T \boldsymbol{\beta}_2^{\sigma}\right) + \phi(s_t)\left(h\left((\mathbf{x}_{t,i}^{\sigma})^T \boldsymbol{\beta}_1^{\sigma}\right) - h\left((\mathbf{x}_{t,i}^{\sigma})^T \boldsymbol{\beta}_2^{\sigma}\right)\right). \quad (12)$$

Under the null of no ST, H_0: $\boldsymbol{\beta}_1^{\gamma} = \boldsymbol{\beta}_2^{\gamma}$ and $\boldsymbol{\beta}_1^{\sigma} = \boldsymbol{\beta}_2^{\sigma}$, making the second terms of the right hand sides equal to zero. Consequently, parameters of ϕ (i.e. c and $\bar{g}$) are unidentified under the null hypothesis, and distributions of test statistics like the likelihood ratio (LR) are likely non-standard. One way to circumvent this issue consists in testing an auxiliary regression model, based on a Taylor expansion of ϕ that avoids unidentification [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 how the Taylor rest affects the random part of our model. Finally, in case of semiparametric extensions, the distribution of the test statistic in the auxiliary procedure is unknown. Thus, we prefer using simulation-based methods, as advocated in Hansen [1999] and Becker et al. [2004]. To do so, we consider a LR-type test, where our test statistic is given by

$$LR(\hat{\Theta}^{H_0}, \hat{\Theta}) = -2(\mathcal{L}(\mathbf{y}; \hat{\Theta}^{H_0}, \mathbf{x}, \boldsymbol{\tau}, \mathbf{n}) - \mathcal{L}(\mathbf{y}; \hat{\Theta}, \mathbf{x}, \mathbf{s}, \boldsymbol{\tau}, \mathbf{n})), \quad (13)$$

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 where

$$\gamma(\mathbf{x}_{t,i}^{\gamma}) = h\left((\mathbf{x}_{t,i}^{\gamma})^T \boldsymbol{\beta}_{H0}^{\gamma}\right), \quad (14)$$

$$\sigma(\mathbf{x}_{t,i}^{\sigma}) = h\left((\mathbf{x}_{t,i}^{\sigma})^T \boldsymbol{\beta}_{H0}^{\sigma}\right). \quad (15)$$

The distribution of (13) under H_0 is obtained by means of a classical parametric bootstrap procedure, described by the following steps:

I. Using $\hat{\Theta}^{H_0}$, $\mathbf{x}$ and $\mathbf{n}$, simulate a time series of exceedances $\mathbf{y}^*_{(b)}$ under H_0.

II. Estimate the vectors of parameters $\hat{\Theta}^{*,H_0}_{(b)}$ and $\hat{\Theta}^*_{(b)}$ from the null and the ST models, respectively.

III. Compute $LR(\hat{\Theta}^{*,H_0}_{(b)}, \hat{\Theta}^*_{(b)})$, a bootstrap realization of the likelihood-ratio.

IV. Repeat for $b = 1, \ldots, B$, e.g. 500 times.

The rejection region at level α is defined by $\left[q^*_{1-\alpha}, +\inf\right)$, where $q^*_{1-\alpha}$ is the empirical $1-\alpha$ quantile of $LR(\hat{\Theta}^{*,H_0}_{(b)}, \hat{\Theta}^*_{(b)})$.

Among others, Andrews and Plöberger [1994], Becker et al. [2004] and Lee et al. [2011] advocate that, instead of (13), one should consider a statistic of the type $\sup_{g_i, c_i} LR(g'_i, c_i)$, using grids of values for g'_i and c_i. However, this approach is computationally costly in a bootstrap procedure. In our simulation study in Section 3, we show that this is not necessary since our test exhibits an adequate size. We also show that a χ^2 approximation with corrected degrees of freedom is actually pretty good, and can be a viable alternative to more time-consuming procedures.

Eventually, notice that our model's definition allows easily to test for the absence of ST structure in one of the two distribution parameters. Indeed, since we assume that $\bar{g}$ and c take the same values in (14) and (15), these parameters stay identified under the null for a less restrictive null hypothesis. In this situation, classical likelihood ratio theory applies.

3 Simulation study

Since ST-GPD models have not been considered previously in the literature, we investigate the finite-sample properties of the proposed procedures. We generate data following the model defined by equations (4) to (7). To study the properties of estimator (9) in a realistic context, we pay attention to specify predictors with distributions similar to those encountered in our empirical analysis, i.e. count data, time series and collinearity. Thus, we choose the following data generating processes for the covariates:

$$\mathbf{x}^{\sigma}_{ti} = \mathbf{x}^{\gamma}_{ti} = \mathbf{x}_{ti} = \mathbf{x}_t,$$

$$x_{t1} \sim \text{Poiss}(\lambda_t),$$

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

$$x_{tj} = .1 + .65x_{t-1\,j} + e_{tj},\ j = 2, \ldots, p$$

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

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

λ_t is driven by a self-exciting process, as in Porter and White [2012]. $b(\cdot; 10)$ is the standardized geometric distribution with decay rate $r = 10$. $\mathbf{\Sigma}$ is a variance-covariance matrix with one on the diagonal elements and .5 otherwise. We consider $n_t \in \{10, 20, 30, 50\}$ fixed for a given simulation, and $T \in \{50, 100, 150, 200\}$. We choose $\alpha \in \{.7, .8, .9\}$, such that the transition variable exhibits highly persistent time-dependencies like the VIX. 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 different values for the regression coefficients $\boldsymbol{\beta}_1^\gamma$, $\boldsymbol{\beta}_2^\gamma$, $\boldsymbol{\beta}_1^\sigma$ and $\boldsymbol{\beta}_2^\sigma$ 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 (for clarity, only a subset of the simulation set-ups are presented, see supplementary material for more details). 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, though, we see a decrease in the variability of the estimation. Regarding the transition parameters known to be difficult to estimate, we observe very good estimations, close to the true parameters. 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 an increase in T and n_t. A similar effect is observed when the degree of persistence for s_t decreases (i.e. when α decreases). Regarding the Wald-type test proposed in Section 2, we observe an excellent size of the test (Figure 2, panel (a)). Focusing on the 5% test level, we see that on average across all parameters and the 200 simulations, the type-I error is always close to the test level[6]. 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, panel (b)). The average proportions of rejections vary between 25% and 75%.

Lastly, we shortly investigate the size of the likelihood-ratio test proposed in Section 2.3. The bootstrap method is particularly time-consuming to study in a simulation set-up, therefore we use the warp-speed method proposed by Giacomini et al. [2013] to shorten the computing time. For our approach, it consists in resampling $B' = 1$ sample of bootstrap data for each Monte Carlo sample, and to compute $q^*_{1-\alpha}$ from $N = B$ samples. It allows us to easily investigate cases with $p = 4$ and $p = 5$. We set $N = 1,000$. We simulate samples following a classical GP-regression model, with its parameters given by $\boldsymbol{\beta}_1^\gamma$ and $\boldsymbol{\beta}_1^\sigma$. We

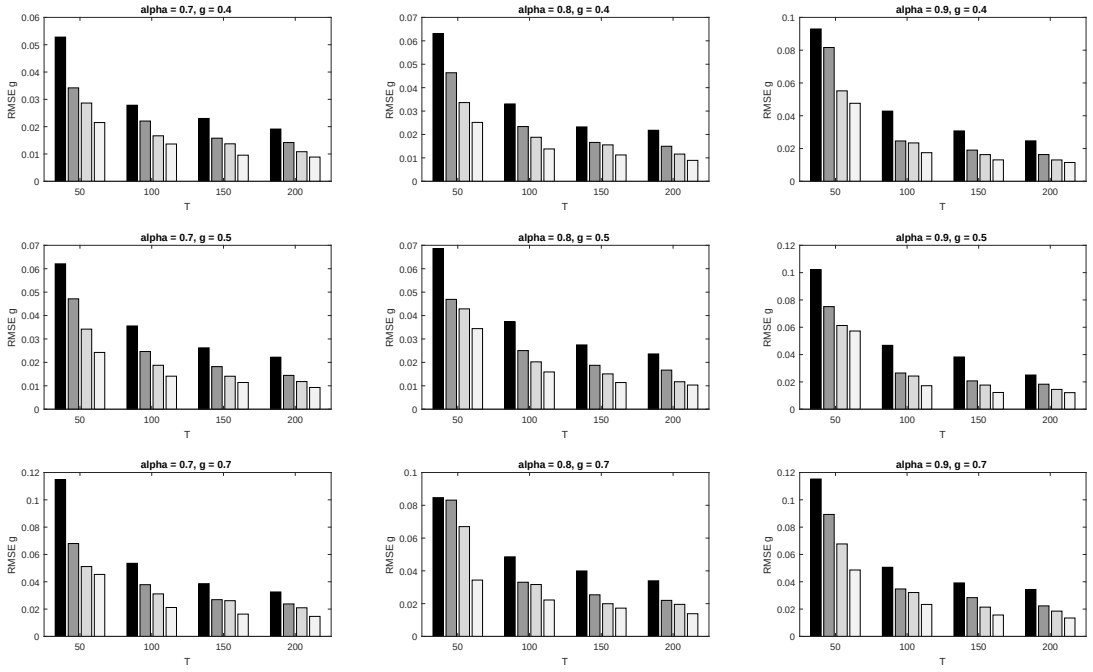
[6] As pointed out by Teräsvirta [1994], we might sometimes obtain negative estimates of the asymptotic standard deviations preventing us from computing a test statistic, due to numerical approximations. Here, on average across simulations, it concerns only 0.17% of the samples.

use the same assumptions regarding the predictors. For s_t, we set $\alpha = .8$. 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 correct identification of both models, the test statistics (13) would follow a $\chi^2_{2\times p+2}$ distribution under H_0. We also look at the size obtained from a χ^2 distribution with either $2 \times p$ or $2 \times p + 4$ degrees of freedom. Table 2 displays the obtained sizes. The classical LR test is strongly oversized, leading us to reject wrongly the null too frequently. At the contrary, our BS test exhibits rejection frequencies closer to the expected nominal levels (5% and 10%). Similarly, if we use a χ^2 distribution with corrected degrees of freedom, using $2 \times p + 4$ instead of $2 \times p + 2$, we obtain excellent sizes. Figure 3 illustrates these results, showing the various distributions against the true empirical distribution of (13), for $p = 5$. We see that both the bootstrap and the $\chi^2_{2\times p+4}$ match pretty well the true distribution.

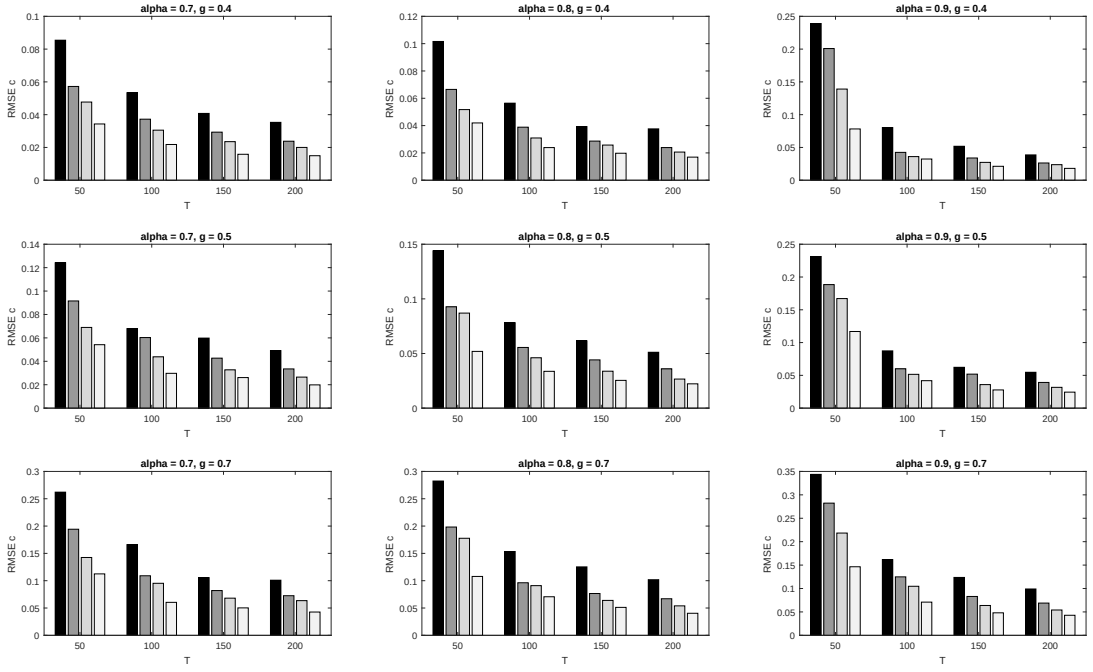
Thus, our simulation study highlights the correctness of our procedure in terms of estimation, uncertainty quantification and testing.

Table 1: Average estimated regression coefficients, across the 200 simulations and the differents specifications of α. In parentheses, average standard deviation of the estimates, accross simulations' specifications.

		T=100				T=150				
Param.	Truth	n_t	10	20	30	50	n_t	10	20	30	50
$\bar{g}$	.4		.399	.398	.4	.4		.398	.4	.399	.399
			(0.03)	(0.02)	(0.02)	(0.01)		(0.03)	(0.02)	(0.02)	(0.01)
c	0		.001	.001	.001	.001		.003	.00	-.001	0.001
			(0.06)	(0.04)	(0.03)	(0.03)		(0.04)	(0.03)	(0.03)	(0.02)
β^{γ}_{10}	-1.5		-1.58	-1.68	-1.54	-1.54		-1.58	-1.55	-1.52	-1.52
			(0.73)	(1.72)	(0.34)	(0.52)		(0.49)	(0.34)	(0.29)	(0.2)
β^{γ}_{11}	.04		.02	.046	.038	.041		.036	.04	.038	.04
			(0.13)	(0.12)	(0.04)	(0.05)		(0.06)	(0.04)	(0.03)	(0.02)
β^{γ}_{12}	.2		.207	.174	.196	.205		.207	.196	.202	.197
			(0.3)	(0.42)	(0.12)	(0.14)		(0.16)	(0.12)	(0.1)	(0.07)
β^{γ}_{13}	-.1		-.095	-.069	-.102	-.114		-.099	-.097	-.099	-.098
			(0.3)	(0.56)	(0.13)	(0.18)		(0.16)	(0.11)	(0.09)	(0.07)
β^{γ}_{20}	-1.3		-3.82	-1.48	-1.41	-1.33		-1.449	-1.412	-1.34	-1.335
			(22.07)	(1.49)	(0.51)	(0.35)		(0.83)	(0.46)	(0.39)	(0.28)
β^{γ}_{21}	-.04		-.093	-.048	-.0437	-.044		-.054	-.041	-.043	-.04
			(1.42)	(0.12)	(0.08)	(0.05)		(0.14)	(0.06)	(0.05)	(0.04)
β^{γ}_{22}	-.1		-.563	-.105	-.108	-.105		-0.11	-.098	-.099	-.098
			(4.86)	(0.48)	(0.19)	(0.14)		(0.28)	(0.17)	(0.14)	(0.1)
β^{γ}_{23}	.1		.814	.108	.12	.092		.079	.102	.1	.098
			(6.92)	(0.26)	(0.26)	(0.14)		(0.25)	(0.18)	(0.13)	(0.1)
β^{σ}_{10}	4		4.001	4.004	4.006	4.003		4.007	4.0050	4.005	4.001
			(0.21)	(0.14)	(0.1)	(0.1)		(0.15)	(0.11)	(0.09)	(0.07)
β^{σ}_{11}	.02		.021	.021	.02	.02		.021	.02	.02	.02
			(0.02)	(0.02)	(0.01)	(0.01)		(0.02)	(0.01)	(0.01)	(0.01)
β^{σ}_{12}	.16		.164	.159	.161	.161		.164	.162	.161	.16
			(0.07)	(0.05)	(0.04)	(0.03)		(0.05)	(0.04)	(0.03)	(0.02)
β^{σ}_{13}	-.05		-.054	-.048	-.048	-.05		-.054	-.049	-.051	-.051
			(0.07)	(0.05)	(0.04)	(0.03)		(0.05)	(0.03)	(0.03)	(0.02)
β^{σ}_{20}	1.5		1.509	1.507	1.509	1.506		1.514	1.521	1.498	1.502
			(0.25)	(0.18)	(0.15)	(0.12)		(0.19)	(0.13)	(0.11)	(0.09)
β^{σ}_{21}	-.035		-.036	-.035	-.036	-.035		-.036	-.037	-.035	-.035
			(0.03)	(0.03)	(0.02)	(0.01)		(0.02)	(0.02)	(0.01)	(0.01)
β^{σ}_{22}	-.11		-.113	-.111	-.109	-.11		-.114	-.113	-.113	-.112
			(0.09)	(0.07)	(0.06)	(0.04)		(0.07)	(0.05)	(0.04)	(0.03)
β^{σ}_{23}	.1		.107	.102	.096	.1		.107	.101	.1	.101
			(0.09)	(0.06)	(0.06)	(0.05)		(0.07)	(0.05)	(0.04)	(0.03)

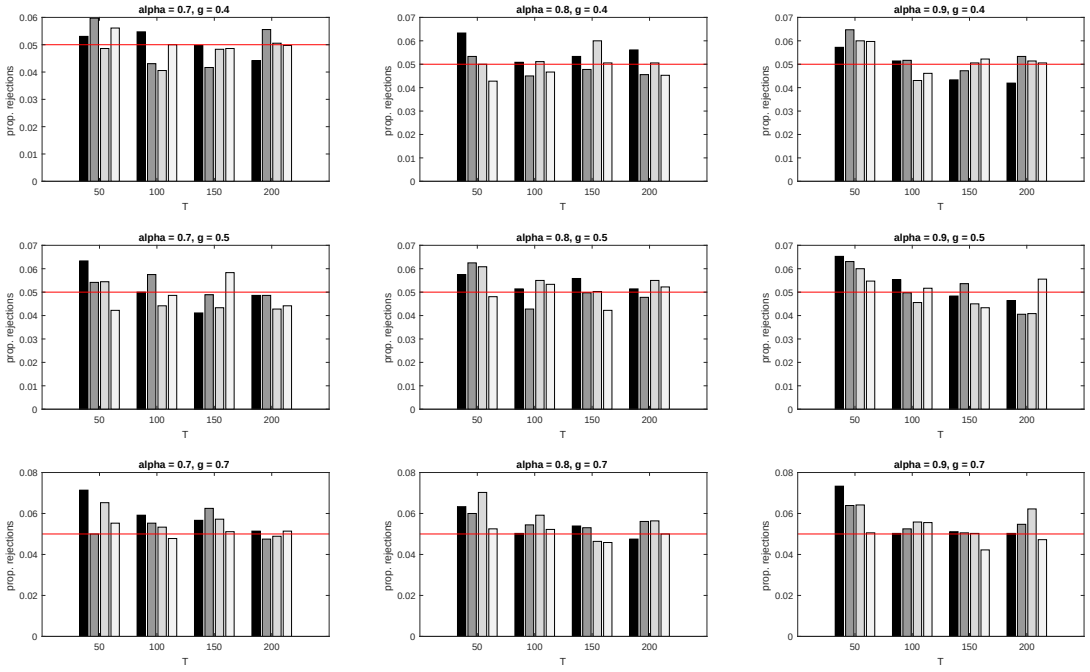


(a)

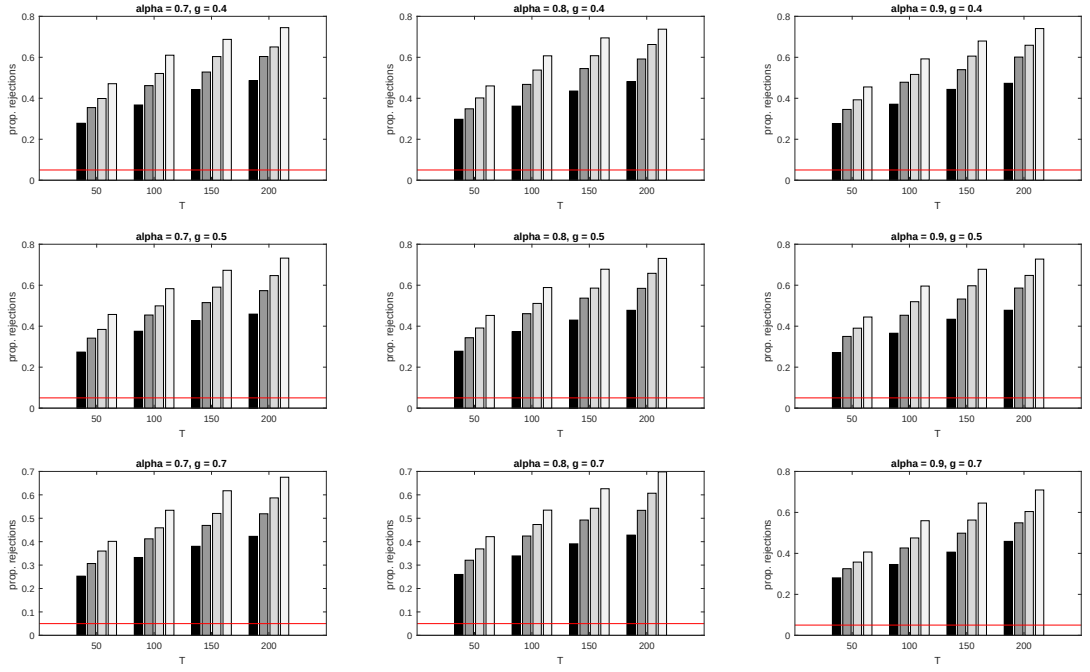


(b)

Figure 1: (a) RMSE for parameter $\bar{g}$ and (b) RMSE for parameter c, for different values of $\bar{g}$ and α. 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)).



(a)



(b)

Figure 2: Average power of 5% Wald tests, across the 18 parameters under (a) the null of the regression coefficients taking their true values and (b) the null of the regression coefficients being equal to 0. 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. (13), 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	$\chi^2_{2\times p}$	$\chi^2_{2\times p+2}$	$\chi^2_{2\times p+4}$	BS	$\chi^2_{2\times p}$	$\chi^2_{2\times p+2}$	$\chi^2_{2\times p+4}$
$p = 3$	4.7%	18.3%	8.1%	3.5%	8.8%	29.8%	16.7%	7.8%
$p = 4$	5.5%	18.4%	9.4%	4.7%	10.9%	29.3%	17.6%	9.4%
$p = 5$	6.5%	16.4%	9.9%	5.6%	10.7%	26.6%	16.2%	10.1%

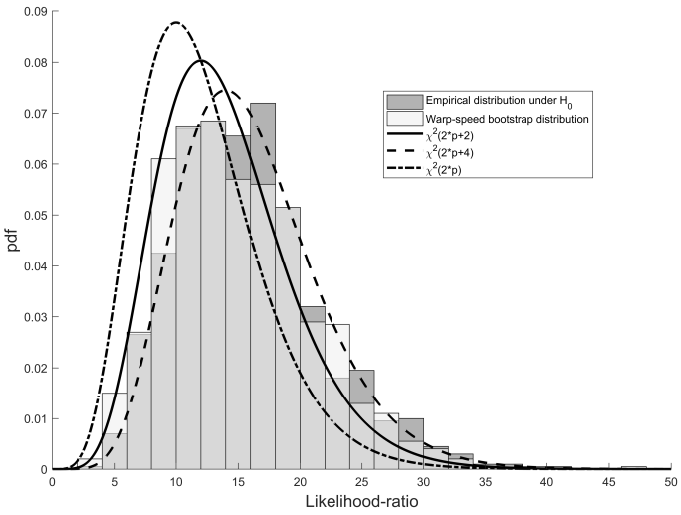


Figure 3: Comparison between the empirical distribution of the likelihood-ratio (dark grey), the warp-speed bootstrap distribution (light grey), $\chi^2_{2\times p}$ (dashed dotted), $\chi^2_{2\times p+}$ (solid) and $\chi^2_{2\times p+4}$ (dashed) distributions.

4 Data

In this section, we present the characteristics of the studied operational loss data, as well as the explanatory variables used for our regression analysis.

4.1 Operational loss data at UniCredit

Operational loss data have been provided by UniCredit risk department. Losses have been 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). Initially, our database consists in 13,209, 16,138 and 6,391 losses, respectively. Loss amounts have been adjusted for inflation using the Italian consumer price index. For anonymity reasons, losses have been multiplied by an unknown

factor.

To select extreme losses, we set our selection threshold τ as the empirical quantile at level 75% for a given ET[7]. Eventually, we conduct our analysis on the 25% largest losses, leading to final samples of size $3,288$, $4,019$, and $1,585$, and 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 ET-specific dependence structures.

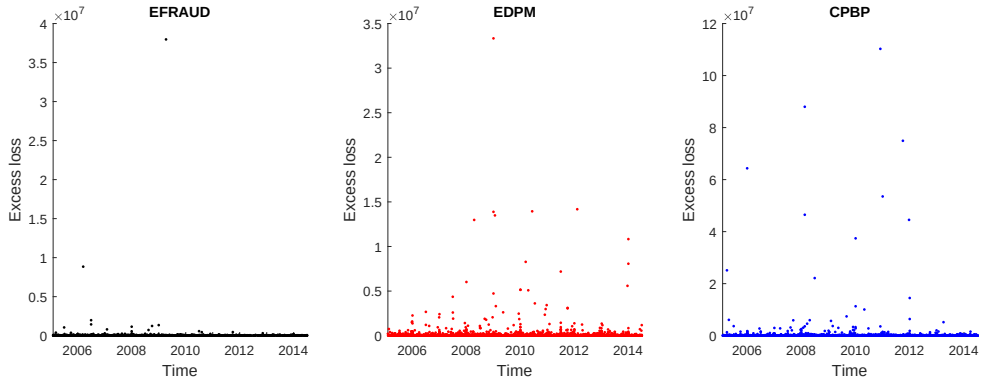


Figure 4: Excess losses, over time, for the three different event type. From left to right: EFRAUD, EDPM and CPBP.

The large dispersion of the losses makes it hard to have 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 where losses appear particularly large. For EFRAUD, 2007 and 2009-2010 seem more concerned. For EDPM, we observe an increase in the dispersion starting in 2007, with a peak in 2009, before a decrease the next years. Lastly, for CPBP, we observe several peaks: in 2007, 2009, as well as between mid-2010 and 2012. Overall, it seems clear that the losses have heterogeneous distributions over time.

The observed counts of the extremes are depicted in Figure 5, panel (a) to (d). Overall, we observe a progressive decrease of the number of monthly losses over time, as well as a clear seasonal pattern, probably due to reporting. See, e.g., Chernobai and Yildrim [2008] and Chernobai et al. [2011] for a discussion on this phenomenon. The largest share belongs to the CPBP event type. We also observe a clear positive correlation between the number of losses in each event type, suggesting that common factors are driving the loss process. It emphasizes the needs (and the usefulness) of a joint modeling of the data.

[7]Extensive discussion on threshold selection for this dataset can be found in Hambuckers et al. [2018a], who used ET-specific quantiles at level 75%. Due to the complexity of the phenomenon investigated, we believe that 75% offers a good balance between a large sample size and correct specification of extremes. Additional analyses with a 90% threshold are available in the supplementary material.

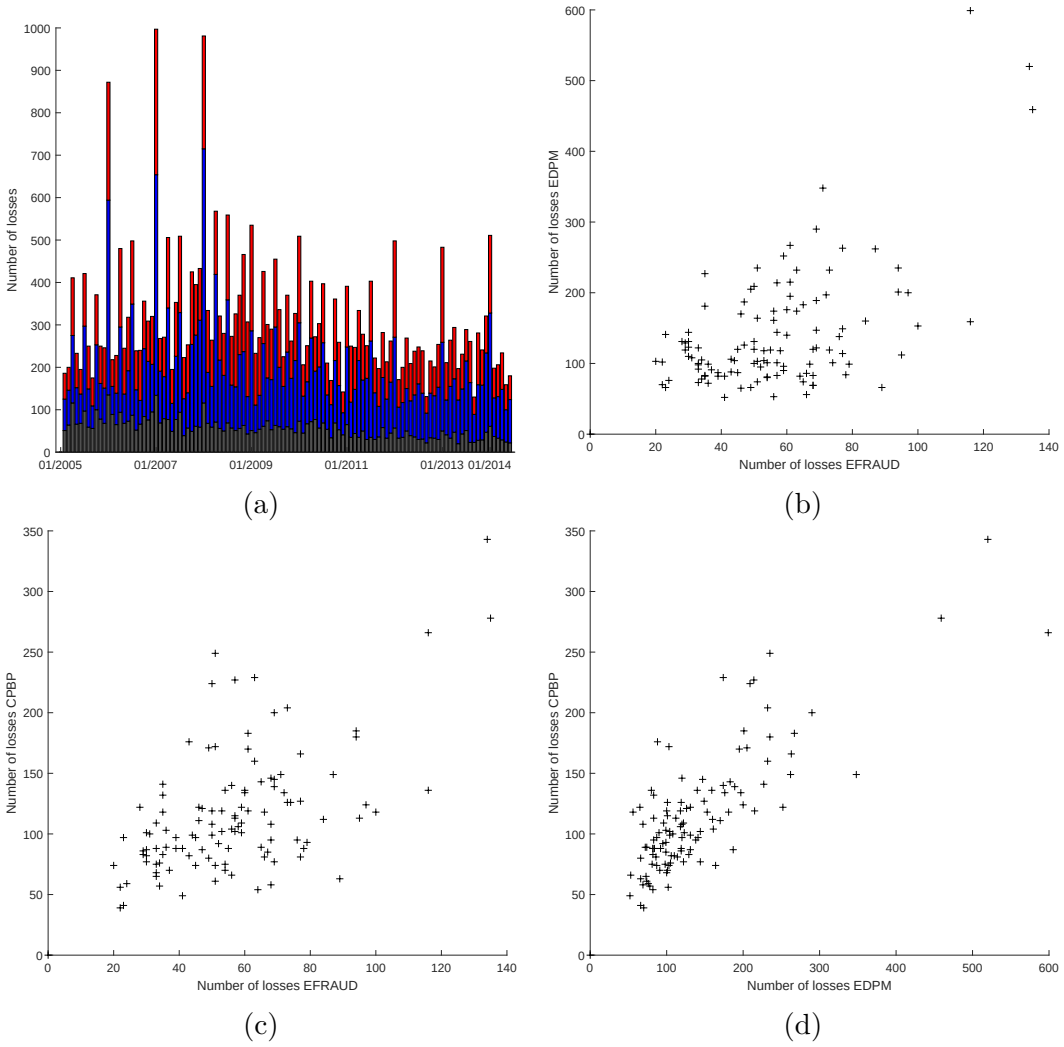


Figure 5: Number of losses in the different event types. Graph (a) gives the sum of the losses over time, on a monthly basis: grey for EFRAUD, blue for EDP and red for CPBP. Graphs (b) to (d) plot the number of the losses of one event type against those of another one.

4.2 Economic determinants and transition variable

To each extreme loss, we associate a set of economic determinants registered the previous month[8]. Relying on previous finding in Chernobai et al. [2011], Cope et al. [2012] and Hambuckers et al. [2018a], we use the following variables: the VIX, the St. Louis Fed financial stress index (FSI), the economic policy uncertainty indices (EPU) of Baker et al. [2016] for the European Union (EU) and Italy, EU and Italian industrial production growth rates, the Italian yield spread on 10-year government bond w.r.t. German yield, as well as the past number of extreme losses. Due to data availability and a desire to limit the complexity of our model, we restrict our attention to this small set of predictors. In particular, the use of aggregate indices allows us to indirectly pre-select and control for a large number of macroeconomic factors at a small cost in terms of parameters. Indeed, the

[8]Using lagged value of the predictors allows us to minimize reverse causality and endogeneity issues, as well as providing an obvious framework for prediction.

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 (like the VIX), and measures the degree of financial stress in the markets. On the other hand, EPU is based on newspaper coverage frequency of specific words[9], and proxies movements in policy-related economic uncertainty [Baker et al., 2016]. Here, the focus is on the dynamic of the dependence structure, especially with the past number of losses, rather than on the explicit identification of relevant predictors.

A series of theoretical justifications have been advanced to explain the apparent link between these predictors and loss severity. Regarding macroeconomic factors (i.e. industrial production growth), they are indicative of the size of transactions: in good economic times, transactions are larger, increasing the likelihood of large losses [Hambuckers et al., 2018a, Cope et al., 2012]. 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]. On the other end, financial stability and economic uncertainty (as measured by FSI, EPU and spreads) might impact the loss formation process, through its effect on the price of financial products: when volatility or instability of financial market are high, the likelihood of a market crash increases. Consequently, system's failures or people's mistakes causing timing issues are rapidly costly. In addition, high uncertainty and high risk aversion usually lead to a looser monetary policy, as measured by real interest rates [Bekaert et al., 2013]. This policy has been shown to incentivize banks to adopt more risky investment strategy, making costly operational mistakes more likely [Delis and Kouretas, 2011, Boubaker et al., 2017]. Therefore, we use the FSI and EPU to capture these effects. Italian spreads allow controlling for the general financial context in Italy.

Regarding the use of the past number of extreme losses, this feature is particularly important to establish the total operational risk capital with models based on compound processes [Chapelle et al., 2008, Brechmann et al., 2014]. Moreover, if a relationship exists between the number of past claims and future loss amounts, it would provide risk managers with readily available indicators, allowing them to take risk mitigation action. One possible reason for the existence of this relationship is suggested in Chernobai et al. [2011], who 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 might also convey some information about loss sizes implied by a given control level. This point is also highlighted in Wang and Hsu [2013], who noticed that a failed control environment is a major contributing factor to significant operational losses. The inclusion of this variable as an important component in the baseline part of our model allows to investigate this hypothesis. In particular, we face two possible situations: either a self-excitation effect, where operational events are indicative of future larger losses; or a self-inhibition effect, indicating a decrease in extreme losses. The former case is indicative

[9]E.g. for the US, the index reflects the frequency of articles in 10 leading U.S. newspapers that contain the following terms like *economic*, *uncertain*, *deficit*, *Federal Reserve*, etc.

of a persistent situation of breakdowns in internal controls. At the contrary, the latter refers to situations where the failure of internal controls are corrected, such that 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, Frees and Valdez, 2016], it is mostly overlooked in the operational risk literature. To the best of our knowledge, the present study is the first one to investigate this dependency in an extreme regression context. Notice that here we decompose past numbers of extreme losses into the contributions of the different ET: to a loss in t belonging to e.g. EFRAUD, we associate as predictor the number of EFRAUD losses registered over period $t-1$.

For the transition variable, we use lagged end-of-month level of the Chicago Board Options Exchange Market Volatility Index (CBOE VIX). 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]. In a number of studies, VIX has been shown to convey predictive power of future structural changes. Among others, Bekaert et al. [2013] and Bekaert and Hoerova [2014] show that high values of the VIX are indicative for future financial instability and laxer monetary policy. In Adrian and Shin [2010], the authors argue that, due to managerial constraints, periods of times characterized by high values of the VIX are associated with a decrease in risk appetite of financial intermediaries. Variation in the VIX seems also to have an effect on market liquidity, thus capturing changes in market structure [Chung and Chuwonganant, 2014]. Hence, VIX seems to be a good indicator of future changes in expectations and risk aversion of market participants. These two factors have necessarily an important impact on the loss formation process. Indeed, such changes in expectations imply a modification in the response of economic agents to economic changes; whereas a change in risk aversion implies a modified response to economic incentives. Consequently, the use of the VIX in our transition function, as the driver of the time-varying dependence, should allow to capture these structural changes in the sense that they modulate the strength between baseline factors (like internal controls) and the severity distribution. By doing so, we specify an explicit channel i.e. changes in risk aversion, uncertainty and expectation, that explain the time-varying dependence. Additionally, the use of the VIX in the transition function can be seen as introducing nonlinear interaction effects between VIX and other predictors, but with a limited number of parameters. To limit numerical instabilities that could be caused by abrupt changes in the raw VIX data, we apply first a logarithm transformation. It leads to distribution parameters being functions of the log-implied volatility, as in Christensen and Prabhala [1998].

An alternative transition variable could be the FSI, as it is also indicative of financial stress. However, as explained in Kliesen and Smith [2010], FSI "can be thought of as coincident indexes rather than as leading indexes - that is, they are designed to measure developments as they occur". Therefore, even though FSI can possess some forward-looking properties (after all, one of the series used in its construction is the VIX), we prefer using the VIX, since numerous empirical studies documented its predictive ability

(see references from the previous paragraph). From a practical perspective, unreported results obtained with the FSI were particularly unstable as well, leading us to focus our analysis on a transition model based on the VIX.

Time series of the explanatory variables are given in Figure 6. With these variables, we capture both the high uncertainty around the Great Recession (characterized by high values of the VIX and a drop in industrial growth), as well as 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 consider also models with either only EU or Italy-related variables.

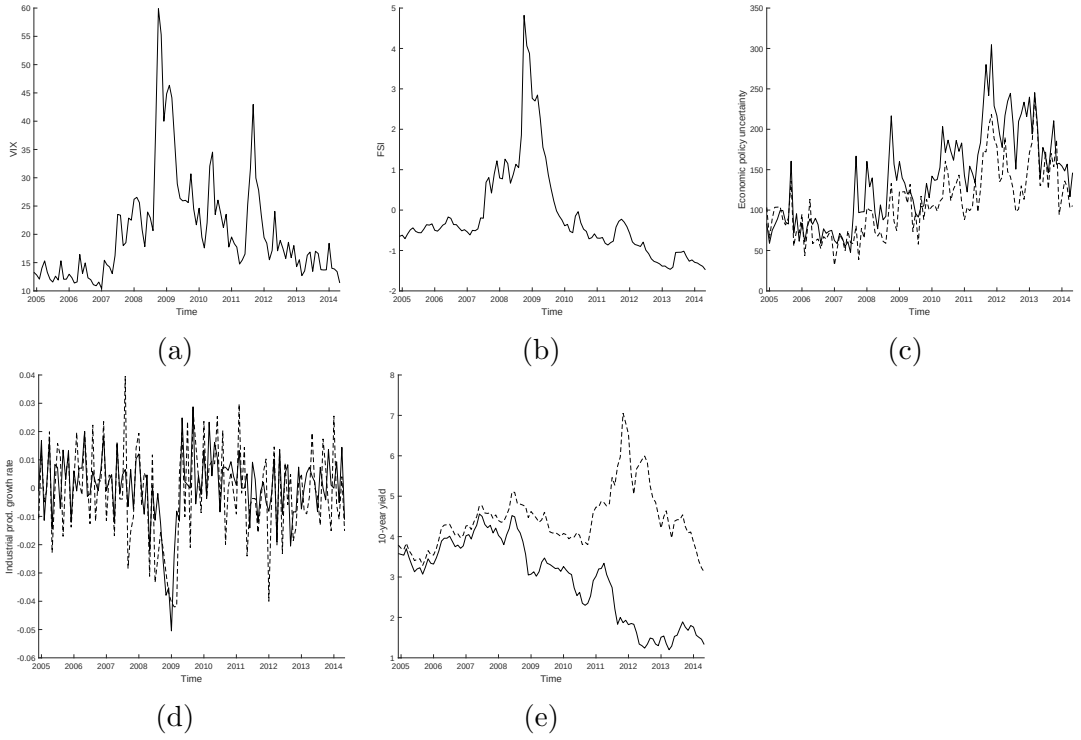


Figure 6: Measures of uncertainty on the financial markets and of the economic context (first row): (a) VIX, (b) FSI and (c) EPU for EU (solid) and Italy (dashed). Second row: (d) Growth rate of the industrial production (monthly basis, solid: EU, dashed: Italy) and (e) 10-year government bond yield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

5 Results

In this section, we apply the proposed model to UniCredit operational losses, using the predictors described in the previous section. The vectors of predictors for the most complex model are given in Table 3. For γ, we restrict the predictors to the ET, 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.

Table 3: Summary of the explanatory variables used in the most complex model, referred as 'Full' in the latter. *cst* refers to a global constant term, $\text{EFRAUD}_{t,i}$ and $\text{CPBP}_{t,i}$ are binary variables indicating if the i^{th} loss at time t belongs to EFRAUD (resp. CPBP) event types, 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 ET variables are ET-specific effects of past counts. EPU_{t-1}^{EU}, EPU_{t-1}^{IT}, $\Delta\text{indprod}_{t-1}^{EU}$, $\Delta\text{indprod}_{t-1}^{IT}$, Δyield_{t-1} and FSI_{t-1} are the EPU for EU and Italy, the growth rate of the monthly industrial production for the same geographical regions, the yield difference of 10-year government bonds between Italy and Germany and the FSI index, respectively.

Variable	Content
s_t	$\log(\text{VIX}_{t-1})$
$x_{t,i}^{\sigma}$	$cst, \text{EFRAUD}_{t,i}, CPBP_{t,i}$ $n_{t-1}^{ET}, n_{t-1}^{ET} \times \text{EFRAUD}_{t,i}, n_{t-1}^{ET} \times \text{CPBP}_{t,i}$ $\text{EPU}_{t-1}^{EU}, \text{EPU}_{t-1}^{IT}, \Delta\text{indprod}_{t-1}^{EU}$ $\Delta\text{indprod}_{t-1}^{IT}, \text{FSI}_{t-1}, \Delta\text{yield}_{t-1}$
$x_{t,i}^{\gamma}$	$cst, \text{EFRAUD}_{t,i}, \text{CPBP}_{t,i}$

5.1 ST models are superior

We investigate the fit provided by subgroups of predictors (EU variables only, Italian variables only, ET variables only and no covariates), with the idea to identify the best model among a set of alternatives that include both ST and non-ST models. In a first step, we look at three criterion: AIC, BIC and censored likelihood score (CLS) in the idea of Diks et al. [2011]. This last criterion is particularly useful to assess the goodness-of-fit of extreme observations, avoiding both selection bias and outcome conditioning (see, e.g. the discussions in Lerch et al. [2017] and Gneiting and Ranjan [2011]). The CLS for a giving censoring threshold α is given by [Diks et al., 2011]

$$\begin{aligned} CLS_{\alpha} = -\sum_{t=1}^{T}\sum_{i=1}^{n_t} \Big\{ & \mathbb{1}(y_{t,i} > \kappa(\alpha)) \log(gpd(y_{t,i}; \hat{\gamma}_{t,i}, \hat{\sigma}_{t,i}))) \\ & + \mathbb{1}(y_{t,i} \leq \kappa(\alpha)) \log(GPD(\kappa(\alpha); \hat{\gamma}_{t,i}, \hat{\sigma}_{t,i}))) \Big\}, \end{aligned} \tag{16}$$

where $\hat{\gamma}_{t,i}$ and $\hat{\sigma}_{t,i}$ are estimated parameters for observation $y_{t,i}$. We choose $\kappa(\alpha)$ as the empirical α-quantile of $\mathbf{y}$ for various values of α, giving special weights to extremely large observations [Hambuckers et al., 2018a]. Additionally, to control for potential in-sample overfitting, we consider a cross-validated version of (16), where we replace $\hat{\gamma}_{t,i}$ and $\hat{\sigma}_{t,i}$ by

their corresponding cross-validated estimators[10]. This criterion is denoted CLS_{α}^{CV}.

Looking first at AIC (Table 4), we see that the ST models fit best the data. The best model appears to be the one with EU variables only (i.e. excluding Italian growth rate and EPU). However, penalizing more for the number of parameter through the use of the BIC, we see that the models without ST effects are preferred. Nevertheless, what matter most in the present application is the goodness-of-fit of extreme observations. Indeed, such observations are the one driving mostly the distribution of the total loss, and for which it is critical to have a good estimation of the distribution to ensure a proper risk assessment. Using the cross-validated CLS with α ranging between .5 and .99 (Figure 7), conclusions are unequivocal: ST models are preferred for every censoring level, whereas proportional differences of the CLS between the best ST model (model 'Full') and non-ST models increase with the threshold. Such a result indicates that extremes are way better modelled via the ST approach, rather than with traditional GP regression models, highlighting the importance of the non-stationary component. In Table 4, for clarity, we only report the difference in CLS for $\alpha = .9$, expressed as a percentage of the CLS for the ST model 'Full'. As robustness checks, we repeat these operations using ET-specific quantiles for the censoring thresholds in equation (16) (Figure 12 in Appendix B). We arrive at the same conclusion.

To test for a significant difference in (in-sample) CLS between the various models, we use the stepwise superior predictive ability (SPA) bootstrap test of Hsu et al. [2010] on the individual censored scores. As benchmark, we use the best non-ST model (i.e. non-ST 'Full'). With this testing procedure, we identify sequentially models with a performance significantly better than the benchmark (Table 5). We find the ST models 'Full', 'EU' and 'Italy' to be significantly better than all non-ST models at the 5% test level. Repeating the same test using this time the ST models 'EU', 'Italy' or 'Full' as benchmarks, respectively, we conclude that no model can beat ST 'Full'. Moreover, ST 'Full' and ST 'EU' are found to be significantly better than ST 'Italy'. These results hold for various values of α between .5 and .95 (see Appendix B for the complete results).

Then, we apply the likelihood ratio test described in Section 2.3 to test for a significant ST effect. It leads us to reject the linearity hypothesis in all cases (Table 6). 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 ST models 'Full' and 'EU' provide significantly better fits in term of CLS than all the other specifications tested.

[10]We 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.

Table 4: AIC, BIC and cross-validated censored likelihood score (CLS) of Diks et al. [2011] for various submodels, with a smooth transition (ST) component (upper panel) or a simple GP regression structure (lower panel). *Cst* refers to a model without explanatory variables, beside the VIX in the ST case. *ET only* refers to a model with only the ET as explanatory variables. *Italy* refers to a model without $\Delta\text{indprod}_{t-1}^{EU}$ and EPU_{t-1}^{EU}, whereas *EU* refers to a model without EPU_{t-1}^{IT} and EPU_{t-1}^{IT}. *Full* refers to the model with the variables given by Table 3.

ST Model	Full	EU	Italy	ET only	Cst
AIC	46,449.93	46,445.17	46,460.92	46,465.48	46,890.46
BIC	46,676.90	46,643.77	46,659.52	46,564.78	46,933.02
$\Delta CLS_{.9}$	0%	0.04%	0.10%	0.19%	1.09%
$\Delta CLS_{.9}^{CV}$	0%	0.06%	0.18%	0.27%	0.83%
No ST	**Full**	**EU**	**Italy**	**ET only**	**Cst**
AIC	46.491.48	46,487.87	46,494.12	46,487.01	46,917.34
BIC	46,597.87	46,580.08	46,586.33	46,529.56	46,931.53
$\Delta CLS_{.9}$	.39%	.39%	.43%	.46%	1.36%
$\Delta CLS_{.9}^{CV}$	.25%	.24%	.27%	.29%	1.04%

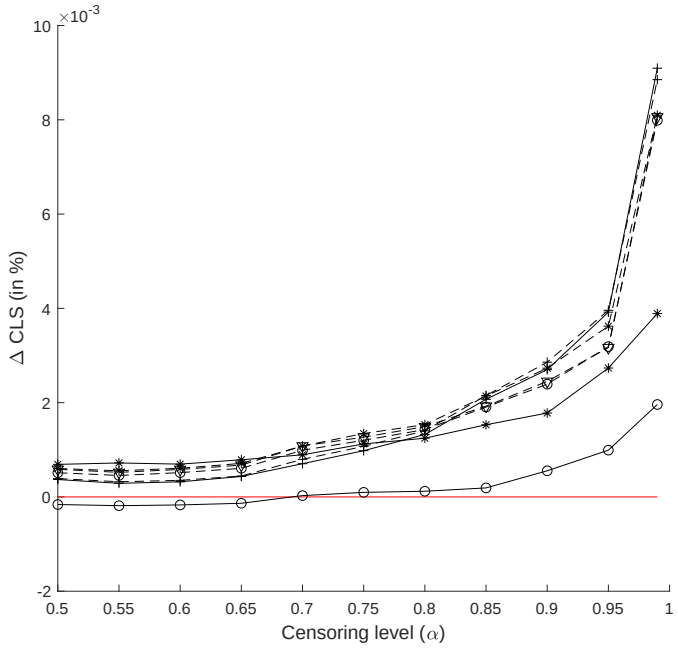


Figure 7: Differential in cross-validated CLS, as a percentage of CLS_{α}^{CV} for the ST model 'Full'. Positive differences imply a less favorable fit for the alternative model. Solid (resp. dashed): ST (resp. non-ST) models. $+$: 'ET only'; $*$: 'Italy', $\circ$: 'EU', $\triangledown$: 'Full'. Red: ST model 'Full'. Models 'Cst' have been ommited for readability.

Table 5: 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 having 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 implementaton 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 6: P-value of the likelihood ratio tests. Left panel: the null model (on the horizontal line) is a ST model, and is tested against an alternative ST model (first colum). Right panel: the null model is a classical GP regression model, and is tested against an alternative ST model. For these tests, due to the unidentification issue under H0 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	-	-	-	$< .01$	-	-	-	$< .01$	$< .01$

Another important question consists in knowing if the ST model provides better Value-at-Risk (VaR) i.e. quantile estimate, of the total loss distribution[11]. To assess the various severity models on a ground independent from a frequency model, we compute quantile

[11]The quantile at level α of Y, Q_α, is defined such that $\mathbb{P}(Y \leq Q_\alpha) = \alpha$.

estimates conditional on knowing the exact number of losses taking place over each time period. That is, for each model 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%) [12]. Estimated VaR at level .975 for various models are plotted over the total loss in Figure 8. The performance 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*. An exception is defined as a time period where 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:

$$H_0 : \mathbb{E}(I_t(\alpha_1)) = \alpha \textbf{ and } \mathbb{E}(I_t(\alpha_2)) = \alpha_2, \tag{17}$$

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 similarly defined). Defining $N(\alpha) = \sum_{t=1}^{T} I_t(\alpha)$ as the number of exception for level α, we test (17) relying on the likelihood ratio test statistic given by

$$\begin{aligned} LR_{MUC}(\alpha_1, \alpha_2) = &-2\log\left[\alpha_1^{T-N(\alpha_1)}(\alpha_2-\alpha_1)^{N(\alpha_1)-N(\alpha_2)}(1-\alpha_2)^{N(\alpha_2)}\right] \\ &+ 2\log\left[\left(\frac{N_0}{T}\right)^{N_0}\left(\frac{N_1}{T}\right)^{N_1}\left(\frac{N(_2)}{T}\right)^{N(\alpha_2)}\right], \end{aligned} \tag{18}$$

with $N_0 = T - N(\alpha_1)$ and $N_1 = N(\alpha_1) - N(\alpha_2)$. The finite sample distribution of (18) is obtained via simulations for $T = 113$, using the procedure proposed in Section 2.3 of Colletaz et al. [2013] [13]. Table 7 shows the results of the LR_{MUC} test obtained from the observed number of exceptions and super exceptions for all models and various VaR levels (see Appendix B, Table 10 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 level .99, whereas this is not the case for the VaR of all ST models. Except for $\alpha_1 = .95$ and $\alpha_2 = .975$, the correct specification of the ST model 'Full' and 'EU' is never rejected at the 5% test level. Remark that the ST model 'ET only' seems to perform quite well in this exercise, too. However, it comes at the price of larger fluctuations in the VaR estimates far in the tail, which is costly and unpractical for banks. For example, the computation of the ratio of mean, median, standard error and maximum value of the VaR at level $\alpha = .99$ between the ST models 'Full' and 'ET only' reveals that the former is more parsimonious: the mean and median VaR are 10% and 15% smaller, the variance is reduced by 12% whereas the

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

[13] In this procedure, we simulate 1,000,000 time series of length $T = 113$ from a U(0,1) distribution, and compute how many times the r.v. are larger than α_1 and α_2. Then, the distribution of (18) is obtained from the empirical distribution.

maximum VaR is shrunk by 24%[14]. Hence, the ST model 'Full' seems more appropriate, for a bank to infer its operational risk capital reserves. Figure 8 displays the total loss and the VaR at level .975 for the ST models 'Full' and 'ET only', as well as the non-ST model 'Full'. It illustrates that the alternatives to the ST model 'Full' imply either excessively large or too small VaR, leading to a costly insurance mechanism or unwanted exceptions.

Table 7: Results of the multiple LR test of Colletaz et al. [2013], using various values for α_1 and α_2. In parenthes, p-values obtained from a simulation procedure with 10^6 samples of length $T = 113$.

Multivariate unconditional coverage test										
	ST models					non-ST models				
(α_1, α_2)	Full	EU	Italy	ET only	Cst	Full	EU	Italy	ET only	Cst
(.95, .975)	4.54	4.54	9.05	1.9	4.543	6.10	7.12	6.10	6.10	7.12
	(.039)	(.039)	(.003)	(.314)	(.042)	(.016)	(.009)	(.016)	(.016)	(.009)
(.95, .99)	2.89	2.89	4.25	2.29	3.63	5.94	5.94	5.94	5.94	5.94
	(.101)	(.101)	(.056)	(.156)	(.086)	(.019)	(.019)	(.019)	(.019)	(0.019)
(.975, .99)	0.88	0.88	0.27	0.84	6.75	2.21	2.47	2.21	2.21	6.57
	(.168)	(.168)	(.457)	(.233)	(.009)	(.116)	(.069)	(.116)	(.116)	(.011)
(.975, .995)	0.30	0.30	1.17	0.45	4.51	1.40	0.50	2.45	2.45	6.67
	(.301)	(.301)	(.16)	(.301)	(.033)	(.123)	(.236)	(.08)	(.08)	(.007)
(.99, .995)	0.55	0.55	0.55	0.06	0.02	2.49	2.49	2.49	2.49	2.49
	(.235)	(.235)	(.235)	(.533)	(.900)	(.024)	(.024)	(.024)	(.024)	(.024)

[14]Comparing the ST model 'Full' with its non-ST counterpart, we find that the median VaR at level .99 is 15.5% smaller for the ST model.

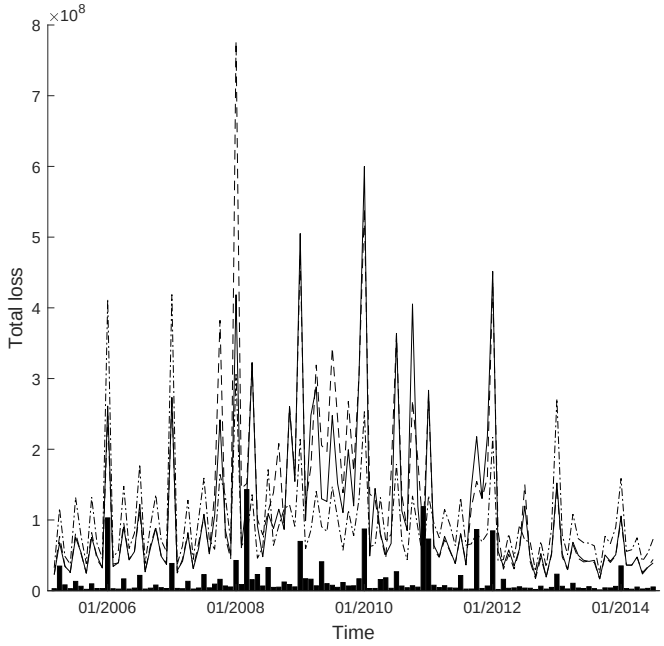


Figure 8: Comparison between various VaR estimates at level .975 and the total loss distribution (black bar chart). Solid: ST model 'Full'. Dashed: ST model 'ET only'. Dashed-dotted: non-ST model 'Full'.

5.2 Economic determinants of operational risks severity

In light of these various performance indicators, we continue our analysis by discussing the estimated coefficients obtained with the ST model 'Full'. Estimates and associated confidence intervals are displayed in Table 9. We find coefficients significantly different from zero in both limiting regimes. We also test for differences in regression coefficients across limiting regimes, and found significant differences (Table 9 in Appendix B). Remember that the distribution parameters are given by (equations (11) and (12)):

$$\begin{aligned}
\gamma(\mathbf{x}_{t,i}^{\gamma}, s_t) &= h\left((\mathbf{x}_{t,i}^{\gamma})^T \boldsymbol{\beta}_2^{\gamma}\right) + \phi(s_t)\left(h\left((\mathbf{x}_{t,i}^{\gamma})^T \boldsymbol{\beta}_1^{\gamma}\right) - h\left((\mathbf{x}_{t,i}^{\gamma})^T \boldsymbol{\beta}_2^{\gamma}\right)\right), \\
\sigma(\mathbf{x}_{t,i}^{\sigma}, s_t) &= h\left((\mathbf{x}_{t,i}^{\sigma})^T \boldsymbol{\beta}_2^{\sigma}\right) + \phi(s_t)\left(h\left((\mathbf{x}_{t,i}^{\sigma})^T \boldsymbol{\beta}_1^{\sigma}\right) - h\left((\mathbf{x}_{t,i}^{\sigma})^T \boldsymbol{\beta}_2^{\sigma}\right)\right).
\end{aligned}$$

Thus, β_2^{σ} and β_2^{γ} summarize the complete dependence structure in period of low uncertainty, i.e. when the VIX is below 20. Indeed, as shown on Figure 9, panel (b), the transition function ϕ is around zero for VIX values up to that threshold. Similarly, for values of the VIX around 30 and above, ϕ is equal to one and $(\beta_2^{\sigma}, \beta_2^{\gamma})$ govern the dependence. Figure 9, panel (a) shows that low uncertainty periods cover roughly the time frames 2005M1-2007M7 and 2012M7-2014M6. The main high uncertainty periods are roughly 2007M8-2008M4, 2008M10-2010M2 and 2011M8-2012M1.

In the low uncertainty period, σ is rather constant: we find few effects of the covariates, with the notable exception of the FSI and the industrial production growth rate. The positive signs of the regression coefficients are consistent with previous studies [Cope et al., 2012, Hambuckers et al., 2018b] suggesting that an increase in economic activity, via a mechanic increase in the size of the transactions, leads to bigger operational losses

and more extreme losses. Similarly, a high level of financial stress can be related to an increase in liquidity issues, which in turn increase the likelihood of extreme losses. Past counts seem to not have any effect on the severity distribution either in these conditions. Looking at differences across ET, CPBP has the highest ET-specific constant, followed by EFRAUD then EDPM. Regarding γ, we observe discrepancies across ET as well, with EDPM having the heaviest tail, followed by CPBP then EFRAUD. CPBP is thus found to be a riskier ET than EFRAUD, a result in line with Cope et al. [2012]

Now, looking at the high uncertainty regime, we observe important changes: first, the value for γ increases for every ET, indicating a higher likelihood of extreme losses, *ceteris paribus*, in a highly uncertain context. This is consistent with previous findings in Chernobai and Yildrim [2008], Cope et al. [2012], Hambuckers et al. [2018b] and Hambuckers et al. [2018a] that connect high uncertainty with laxer monetary policies followed by more risky investments [Bekaert et al., 2013, Boubaker et al., 2017, Delis and Kouretas, 2011]. Another explanation is the impact of unexpected price variations of derivatives on losses stemming from delivery, pricing, selling issues or IT system crashes.

Second, n_{t-1}^{ET} exhibits a significant and negative effect. We observe some discrepancies across ET as well, probably due to scaling differences in the number of losses. Overall, it suggests a self-inhibiting mechanism, where recent events lower the likelihood of extreme losses. This effect might be due to the reviews of internal processes, reinforced controls and various adjustments in the monitoring procedures that usally 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 due 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] regarding 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, G-index and so forth can prove useful to model the severity distribution.

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 might also lead to laxer monetary policies, incentivizing banks to take more risky investments for which operational mistakes are more costly. The observed effect can also be related to the counter-cyclicality of insurance frauds found in Dionne and Wang [2013], where the severity of external fraud losses is higher in recession times. Here, however, we can distinguish between periods of high or low uncertainty. Regarding FSI_{t-1}, in a situation of high uncertainty, an increase in financial instability could be responsible for a shift in risk preferences, pushing towards less risky investment. This phenomenon comes

as a counterbalance of the effect of $\Delta\text{indprod}_{t-1}^{EU}$. Then, for Δyield_{t-1}, a decrease in yield means either an increase in Germany's yield or a decrease in Italian's yield larger than the opposite side of the spread. Hence, a decrease in spread can be associated to relatively cheaper financing conditions from an Italian point of view. This phenomenon can be related as well with an incentivizing context for riskier investments. For example, Bruno and Shin [2015] find evidence of monetary policy spillovers due to cross-border capital flows, and highlight that a decrease in funding costs will affect decisions by the banks regarding how much exposures to take on, leading to greater risks. A similar effect is suggested in Angeloni et al. [2015]. UniCredit being a large European bank active in various countries, we can expect it to be quite sensitive to this channel. Lastly, EPU_{t-1}^{EU} seems positively related with the likelihood of large losses, indicating that anticipated economic instability correlates positively with extreme losses.

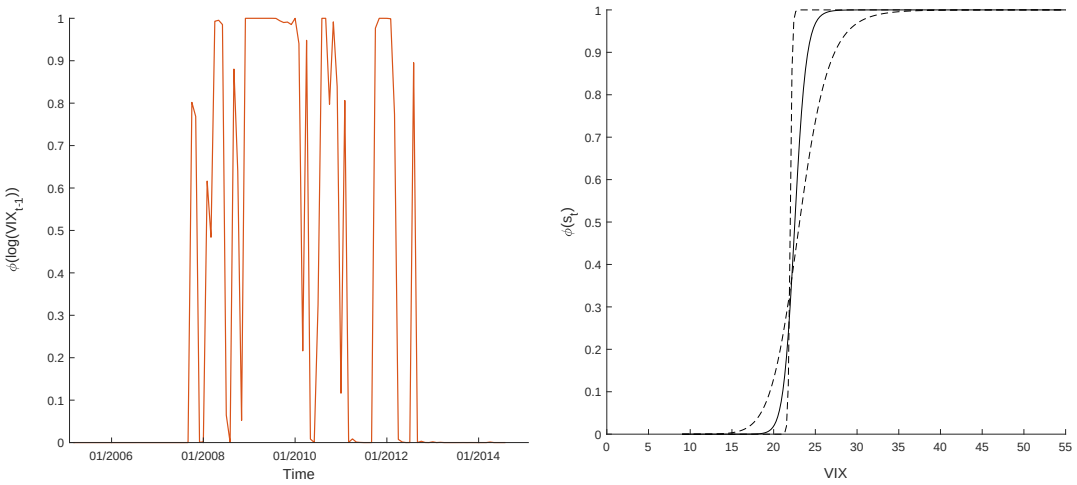
Between these two limiting regimes, the transition component ϕ allows capturing intermediate situations, where uncertainty modulates the effect of the covariates. Especially, it gradually increases γ along an increase in uncertainty. From a theoretical perspective, it also allows the mixing of the different 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 10 displays the time-varying effect of a change in n_{t-1}^{ET} on σ, as a function of the VIX[15]. We can deduce from these graphs regions where σ is more sensitive to changes in n_{t-1}^{ET} i.e. when the VIX is high.

Figure 11 shows the estimated γ and σ over time for the different ET. We display the quantile at level 90% as a measure of the risk (bottom row). As highlighted previously, we observe a higher likelihood of extremes during the financial crisis (end of 2008) and the debt crises (2011 and 2012).

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

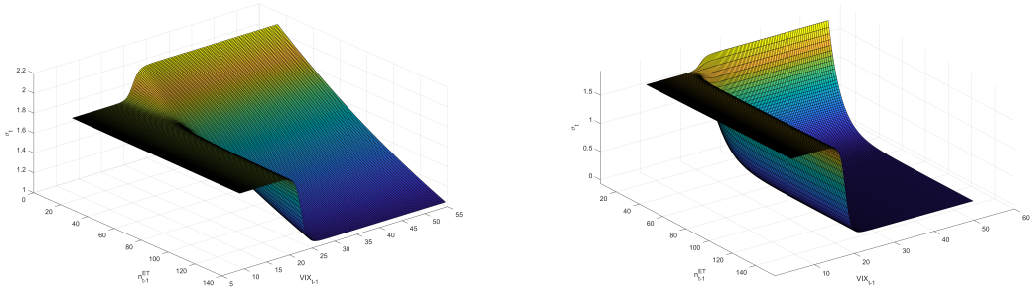
Table 8: Estimated regression coefficient obtained from the ST model 'Full'. In parentheses, estimated standard errors of the estimates. * (resp. ** and ***) denotes coefficients found significantly different from zero at the 5% (resp. 1%) test level.

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)
Cst	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)	-	-
$CPBP_{t,i}$	.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)	-	-	-	-
Δyield_{t-1}	-24.22***	3.66	-	-	-	-
	(6.99)	(4.15)	-	-	-	-

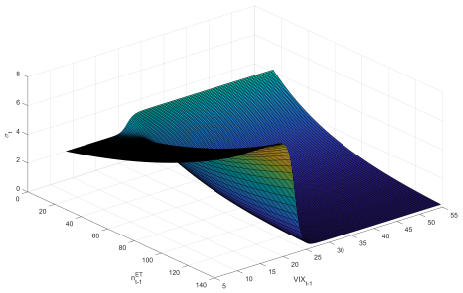


(a) (b)

Figure 9: Value of the transition function ϕ (a) over time and (b) as a function of the VIX (right panel). Dashed lines denote 50% asymptotic confidence intervals.



EFRAUD EDPM



CPBP

Figure 10: 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.

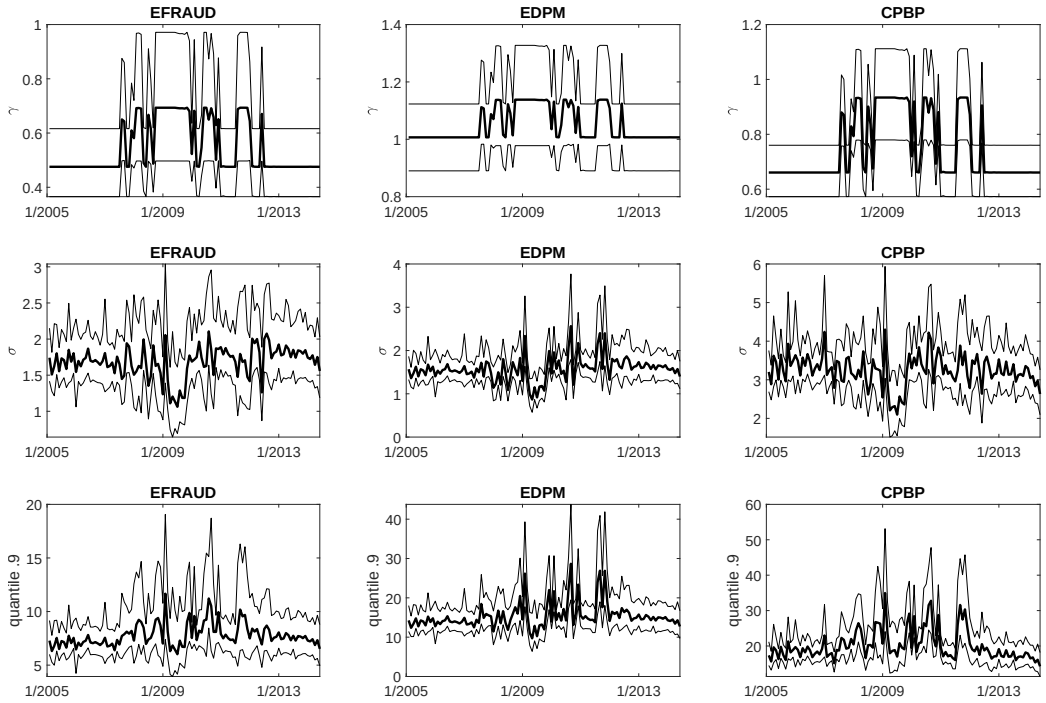


Figure 11: From top to bottom: estimates of γ, σ and the 90% quantile, respectively. Thick line: estimated parameters over time. Thin line: 99% confidence intervals obtained from the posterior simulation procedure of Wood [2006], for a fixed transition function.

6 Conclusion

In this paper, we focus on the issue of connecting the severity distribution of operational losses with the uncertainty on the financial markets and the economic context of a financial institution. In particular, we are interested in the link between past number of losses (i.e. the frequency process) and the severity, with the idea that it proxies the intensity of internal controls.

Acknowledging the potential time-varying effect of changes in economic context due to changes in uncertainty level, we introduce a smooth transition generalized Pareto (ST-GPD) regression approach. This model has the advantage to account for a high probability of extremes and to allow for changing effects of explanatory variables, over time, via a transition function driven by an uncertainty measure. We highlight in a simulation study the good properties of the proposed estimation and testing procedures.

Then, we conduct an empirical study of the dynamics driving operational losses severity at UniCredit. As our transition variable, we use the VIX. We discover that smooth transition structures are highly relevant to improve the goodness-of-fit of extremes. In addition, we suggest that two different mechanisms are at work: in low uncertainty periods, the severity is mostly driven by economic growth and financial instability. On the other end, in high uncertainty periods, we observe a strong inhibition effects suggesting that monitoring and supervision processes at UniCredit following operational events mitigate the likelihood of extremes in subsequent periods. In addition, we find situations synonym

of laxer monetary policies and financing conditions to significantly increase the risk of extreme losses.

Our findings have several policy implications from a risk management perspective: first, the time-varying dependence structure needs to be accounted for, to provide adequate risk estimates for the total loss distribution. Neglecting this effect can lead to too many VaR breaches or too much capital to be reserved. Second, measures of internal control seriousness provide useful indications for the future likelihood of extreme operational events. Therefore, including such indicators in an early-warning system would presumably allow a finer management of risk levels. In particular, evidence of self-inhibition mechanism could ensure counter-cyclical capital requirement, since it accounts for the reduction of risks that follows operational events. Lastly, a general uncertainty measure like the VIX seems to convey some information regarding the strength of the connection between financial factors and the severity of operational events. Thus, a regular re-assessment of operational risks along changes in this uncertainty seems to be needed for proper measurements and monitoring.

References

T. Adrian and H.S. Shin. Liquidity and leverage. *Journal of Financial Intermediation*, 19 (3):418–437, 2010.

D.W.K Andrews and W. Plöberger. Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica*, 62(6):1382–1414, 1994.

A. Ang and A. Timmermann. Regime changes and financial markets. *Annual Review of Financial Economics*, 4(1):313–337, 2012.

I. Angeloni, E. Faia, and M.L. Duca. Monetary policy and risk taking. *Journal of Economic Dynamics and Control*, 52(C):285 – 307, 2015.

F. Aue and M. Kalkbrener. LDA at work: Deutsche Banks approach to quantifying operational risk. *Journal of Operational Risk*, 1(4):49–93, 2006.

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.

R. Becker, W. Enders, and S. Hurn. A general test for time dependence in parameters. *Journal of Applied Econometrics*, 19(7):899–906, 2004.

G. Bekaert and M. Hoerova. The VIX, the variance premium and stock market volatility. *Journal of Econometrics*, 183(2):181–192, 2014.

G. Bekaert, M. Hoerova, and Duca M.L. Risk, uncertainty and monetary policy. *Journal of Monetary Economics*, 60(7):771–788, 2013.

S. 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 STARGARCH 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.

A Chapelle, Y. Crama, G. Hübner, and J.-P. Peters. Practical methods for measuring and managing operational risk in the financial sector: a clinical study. *Journal of Banking & Finance*, 32(6):1049–1061, 2008.

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

P. Chelley-Steeley, N. Lambertides, and C.S. Savva. Illiquidity shocks and the comovement between stocks: New evidence using smooth transition. *Journal of Empirical Finance*, 23:1 – 15, 2013.

A. Chernobai and Y. Yildrim. The dynamics of operational loss clustering. *Journal of Banking & Finance*, 32(12):2655–2666, 2008.

A. Chernobai, P. Jorion, and F. Yu. The derminants of Operational Risk in U.S. financial institutions. *Journal of Financial and Quantitative Analysis*, 46(8):1683–1725, 2011.

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. Chuwonganant. 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. Macroenvironmental 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 Fontenouvelle, 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. Klupperlberg, and T. Mikosch. *Modelling extremal events for insurance and finance*. Springer - Verlag, Berlin, 1997.

C. Florackis, G. Giorgioni, A. Kostakis, and C. Milas. On stock market illiquidity and real-time GDP growth. *Journal of International Money and Finance*, 44:210 – 229, 2014.

E.W. Frees and E.A. Valdez. Hierarchical insurance claims modeling. *Journal of the American Statistical Association*, 89(425):208–218, 2008.

E.W. Frees and E.A. Valdez. Multivariate Frequency-Severity Regression Models in Insurance. *Risks*, 4(1):4, 2016.

J. Garrido, C. Genest, and J. Schulz. Generalized linear models for dependent frequency and severity of insurance claims. *Insurance: Mathematics and Economics*, 70:205–215, 2016.

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.

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. Testing for Linearity. *Journal of Economic Surveys*, 13(5):551–576, 1999.

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.

L. Kilian and M.P. Taylor. Why is it so difficult to beat the random walk forecast of exchange rates? *Journal of International Economics*, 60(1):85 – 107, 2003.

F. Klaassen. Improving GARCH volatility forecasts with regime-switching GARCH. *Empirical Economics*, 27(2):363, 2002.

K.L. Kliesen and D.C. Smith. Measuring Financial Market Stress. *Economic SYNOPSES*, (2), 2010.

M. Kratz. Normex, a new method for evaluating the distribution of aggregated heavy tailed risks. *Extremes*, 17(4):661–691, 2014.

R. Langrock, T. Kneib, R. Glennie, and T. Michelot. Markov-switching generalized additive models. *Statistics and Computing*, 27(1):259–270, 2017.

S. Lee, M.H. Seo, and Y. Shin. Testing for Threshold Effects in Regression Models. *Journal of the American Statistical Association*, 89(425):208–218, 2011.

S. Lerch, T.L. Thorarinsdottir, F. Ravazzolo, and T. Gneiting. Forecasters dilemma: Extreme events and forecast evaluation. *Statistical Science*, 32(1):106–127, 2017.

R. Lukkonen, P. Saikkonen, and T. Teräsvirta. Testing Linearity Against Smooth Transition Autoregressive Models. *Biometrika*, 75(3):491–499, 1988.

M. Pesaran, D. Pettenuzzo, and A. Timmermann. Forecasting time series subject to multiple structural breaks. *Review of Economic Studies*, 73(4):1057–1084, 2006.

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.

M. Power. The invention of operational risk. *Review of International Political Economy*, 12(4):577–599, 2005.

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

H. Rootzen, J. Segers, and J.L. Wadsworth. Multivariate peaks over thresholds models. *Extremes*, 21(1):115–145, 2018.

P. Sands, G. Liao, and Y. Ma. Rethinking operational risk capital requirements. *Journal of Financial Regulation*, 4(1):1–34, 2018.

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.

A. Soprano, B. Crielaard, F. Piacenza, and D. Ruspantini. *Measuring Operational and Reputational Risk: A Practitioner's Approach.* Wiley, 2009.

T. Teräsvirta. Speficifation, 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.

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 Extensions

B Additional results

In this appendix, we provide additional results, highlighting the robustness of our findings to the specification of the different tests. First, we use an ET-specific quantile to compute the CLS statistics (Figure 12). The ranking stays unchanged, and we observe the same increase of the gap between non-ST and ST models when the censoring threshold increases.

Then, Table 9 displays the results of the test for significant differences across limiting regression models, whereas in Table 10 we provide the number of VaR breaches suffered with the various models.

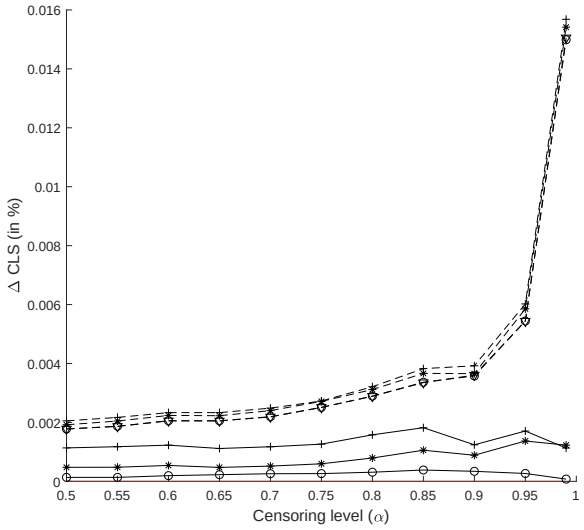


Figure 12: $\kappa(\alpha)$ is computed here as an ET-specific quantile of order α. Differential in CLS (given by equation (16)), as a percentage of the CLS for the ST model 'Full'. Positive differences imply a less favorable fit for the alternative model. Solid (resp. dashed): ST (resp. non-ST) models. $+$: 'ET only'; $*$: 'Italy', $\circ$: 'EU', $\triangledown$: 'Full'. Red: ST model 'Full'.

Table 9: Wald test of a significant difference between regression coefficients of the different limiting regimes. We report the observed difference and correspponding 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***	0.123
	(0.39)	(0.07)
$\text{EFRAUD}_{t,i}$	-0.991**	0.254
	(.5)	(0.18)
$CPBP_{t,i}$	-0.554	0.224*
	(.43)	(.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 10: Number of exceptions, i.e. number of time periods where the historical total loss is larger than the estimated quantile at level α.

	Number of exceptions									
	ST models					non-ST models				
α	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