Modeling extreme events: time-varying extreme tail shape*

André Lucas,[(a)] *Bernd Schwaab,*[(b)] *Xin Zhang,*[(c)]

[(a)] Vrije Universiteit Amsterdam and Tinbergen Institute

[(b)] European Central Bank, Financial Research

[(c)] Sveriges Riksbank, Research Division

March 2022

Abstract

We propose a dynamic semi-parametric framework to study time variation in tail parameters. The framework builds on the Generalized Pareto Distribution (GPD) for modeling peaks over thresholds as in Extreme Value Theory, but casts the model in a conditional framework to allow for time-variation in the tail parameters. The parameter updates improve the expected Kullback-Leibler divergence between the true unobserved data generating process and its GPD approximation. We illustrate the approach using simulated data and two empirical datasets: daily U.S. equity returns, and 15-minute euro area sovereign bond yield changes.

Keywords: dynamic tail risk, observation-driven models, extreme value theory, stock return tails, Securities Markets Programme (SMP).

JEL classification: *C22, G11.*

*Author information: André Lucas, VU University Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands, Email: a.lucas@vu.nl. Bernd Schwaab, Financial Research, European Central Bank, Sonnemannstrasse 22, 60314 Frankfurt, Germany, email: bernd.schwaab@ecb.int. Xin Zhang, Research Division, Sveriges Riksbank, SE 103 37 Stockholm, Sweden, email: xin.zhang@riksbank.se. The views expressed in this paper are those of the author and they do not necessarily reflect the views or policies of the European Central Bank or Sveriges Riksbank.

1 Introduction

This paper proposes a dynamic semi-parametric framework to study time variation in tail fatness for long univariate time series. The new method builds on ideas from Extreme Value Theory (EVT) and uses a conditional Generalized Pareto Distribution (GPD) with time-varying parameters to approximate the tail beyond a given threshold. The GPD is an appropriate tail approximation for most heavy-tailed densities used in financial economics, econometrics, and the actuarial sciences; see, for example, Embrechts et al. (1997), Coles (2001), McNeil et al. (2010, Chapter 7), and Rocco (2014). As a result, the GPD plays a central role in the study of extremes, comparable to the role the normal distribution plays when studying observations in the center of the distribution, and our framework allows for studying any time-variation in the incidence of such extremes.

The time-varying tail shape and tail scale parameters in our model are driven by the score of the GPD density; see Creal et al. (2013) and Harvey (2013). As a result, the model is observation-driven in the terminology of Cox (1981) and its time-varying parameters are perfectly predictable one step ahead. In addition, the log-likelihood function is known in closed form and allows for parameter estimation and inference via standard maximum likelihood methods. Our results show that our model is able to recover the time-varying tail shape and tail scale parameters well in both simulated and empirical data, as well as EVT-based market risk measures such as Value-at-Risk (VaR) and Expected Shortfall (ES). This is the case even if the model is misspecified or the GPD approximation is not exact. The latter is particularly important in our finite sample setting, where the limiting EVT result of the GPD can only hold approximately given the choice of a finite exceedance threshold in any particular sample.

We illustrate our modeling framework using two applications: U.S. equity log-returns, and changes in euro area sovereign bond yields. Each data set consists of two time series, considered separately. We first consider daily log-returns for a stock index (S&P500) and an individual stock (IBM). The S&P500 log-returns range from July, 3 1962 to December, 31 2020, while the IBM stock returns range from January 2, 1926 to December 31, 2020.

Focusing on the left tail, we find that both GPD parameters vary significantly over time. The tail shape varies between approximately 0.05 and 0.25 for the S&P 500, and between 0.05 and 0.35 for IBM. These values imply a maximum moment of order $1/0.25 = 4$ to $1/0.05 = 20$ for the S&P 500, and 3 to 20 for IBM. Standard error bands around the filtered parameters suggest that the tail shape parameter is almost always far from the thin-tailed (Gumbel) case.

Our second illustration investigates the introduction of control variables for the time-variation in tail shape and tail scale. It studies changes in euro area sovereign yields sampled at a 15-minute frequency. Specifically, we study Italian and Portuguese five-year benchmark bond yields during the extremely adverse sovereign debt crisis between 2010 and 2012. Again, we find that both GPD parameters vary significantly over time. The tail shape varies between 0.1 and 0.4 for Italy and between 0.05 and 0.6 for Portugal, implying moment ranges of 2.5–10 and 1.7–20, respectively, and thus incidences of extreme fat tails. We also find that part of the variation can be explained by including central bank bond purchases as a covariate and we provide a way to translate the impact coefficients into their economically interpretable impact on risk measures such as VaR.

Our paper is closely related to a growing strand of literature on modeling time-variation in EVT tail parameters. Several papers propose methodology to study time variation in the tail index. Davidson and Smith (1990), Coles (2001, Chapter 5.3), and Wang and Tsai (2009), among others, also index the GPD tail parameters with time subscripts and equip them with a parameterized structure. Our approach is different in that their "tail index regression" approach requires conditioning variables that explain (all of) the tail variation. Such variables are not always available. By contrast, our "filtering approach" does not require such conditioning variables, and is arguably better suited for the real-time monitoring of extreme equity or bond market risks. Second, Quintos et al. (2001), Einmahl et al. (2016), Hoga (2017), and Lin and Kao (2018) derive formal tests for a structural break in the tail index. A number of subsequent studies applied such tests to financial time series data. Werner and Upper (2004) identify a break in the tail behavior of high-frequency German Bund future returns. Galbraith and Zernov (2004) argues that certain regulatory changes

in U.S. equity markets have altered the tail index dynamics of equities returns, and Wagner (2005) demonstrates that changes in government bond yields appear to exhibit time-variation in the tail shape for both the U.S. and the euro area. de Haan and Zhou (2021) propose a non-parametric approach to estimating the extreme value index locally. Our paper adds to this strand of literature by proposing a framework that allows us to study both the tail shape and tail scale dynamics directly in a semi-parametric way. Explanatory covariates can be included in the dynamics of both parameters, and likelihood ratio tests are available to test economically relevant hypotheses. Finally, unlike Patton et al. (2019), our tail VaR and ES dynamics explicitly account for fat tail shape beyond a threshold as emerging from EVT. The dynamics based on the score for the GPD contain weights for extreme observations. Such weights are absent in the elicitable score functions of Patton et al. (2019). The resulting dynamics in our model are, as a result, more robust, particularly for the ES.

Whereas de Haan and Zhou (2021) take a non-parametric perspective, the methodological part of this paper is closest to Massacci (2017), who also proposes a dynamic parametric model for the GPD parameters. Our framework is different from Massacci (2017) in that we specify *both* GPD parameters as functions of their respective scores. Massacci (2017), by contrast, uses only the score of the first (tail index) parameter to drive both parameters. This is not optimal in the sense of Blasques et al. (2015), who require that each time-varying parameter is associated to its *own* score. Our paper further differs from Massacci (2017) in that we suggest time-varying EVT thresholds to locate the boundary between the center of the distribution and its tail. This implies that we do not need to assume that the time series at hand has no volatility clustering, nor that we need to pre-filter for such volatility clustering. Absence of conditional heteroscedasticity would be hard to defend for the financial data considered in this paper. Finally, we differ from Massacci (2017) in that we discuss inference on both static and time-varying parameters, provide sufficient conditions for the stationarity and ergodicity of the factor process and observations, explain how to introduce additional conditioning variables into the model and assess their usefulness in economic terms, and provide Monte Carlo evidence on the model's performance in a range of challenging settings.

We proceed as follows. Section 2 presents our statistical model. Section 3 discusses our simulation results. Section 4 illustrates the model using U.S. equity log-returns and changes in euro area sovereign yields. Section 5 concludes. A Web Appendix derives the more technical results and provides further empirical evidence.

2 Statistical model

2.1 Time-varying tail shape and tail scale

2.1.1 Conditional EVT framework

Consider a univariate time series y_t, $t = 1, \ldots, T$, where T denotes the number of observations. This section then describes our model for y_t with time-varying tail shape and scale. We assume y_t is generated by a conditional probability density function (pdf) $g(y_t \mid \mathcal{F}_{t-1})$, where $\mathcal{F}_{t-1} = \{y_{t-1}, y_{t-2}, \ldots, y_1\}$ denotes the information set containing all past data. By keeping the conditional density of y_t in its current general form, we stay close to the semi-parametric nature of an extreme value-based approach and make no modeling assumptions about the *center* of the distribution. Alternatively, we could specify a parametric distribution for y_t with for instance a time-varying conditional location μ_t and scale σ_t. The μ_t and σ_t could then be used to pre-filter the raw y_t. We do not pursue such an approach here. First, modeling the center of the conditional distribution leads away from the focus on the tails only, which is the key aspect in an EVT-based approach. Second, designing an additional model for time-varying location and scale would create another layer of complexity to the model, with subsequent model risk and parameter uncertainty. We therefore keep the general form of $g(y_t \mid \mathcal{F}_{t-1})$ and focus on its *tail* using a dynamic extension of arguments from extreme value theory (EVT), similar to Patton (2006)'s extension of copula theory to the dynamic, observation driven setting.

We assume $g(y_t \mid \mathcal{F}_{t-1})$ has heavy tails with time-varying tail index $\alpha_t > 0$. A prime example is when $g(y_t \mid \mathcal{F}_{t-1})$ is a univariate Student's t distribution with $\nu_t = \alpha_t$ degrees of freedom. Other examples include the Pareto, inverse gamma, log-gamma, log-logistic, F,

Fréchet, and Burr distribution with one or more time-varying shape parameters. Rather than modeling the (dynamic) tail shape of an arbitrarily chosen parametric family of distributions, we appeal to well-known results from the EVT literature: the conditional cumulative distribution function (cdf) $G(y_t \mid \mathcal{F}_{t-1})$ of y_t can under very general conditions be approximated by $G(y_t \mid \mathcal{F}_{t-1}) \approx G(\tau_t \mid \mathcal{F}_{t-1}) + (1 - G(\tau_t \mid \mathcal{F}_{t-1}))P(x_t; \delta_t, \xi_t)$ with $x_t = y_t - \tau_t$ for sufficiently high thresholds $\tau_t \in \mathbb{R}_+$. More precisely, we have

$$\begin{aligned} &\lim_{\tau_t \to \infty} \sup_{x_t \geq 0} \left| \mathbb{P}\left[Y_t \leq x_t + \tau_t \mid Y_t > \tau_t,\ \mathcal{F}_{t-1} \right] - P(x_t; \delta_t, \xi_t) \right| \\ &\qquad = \lim_{\tau_t \to \infty} \sup_{x_t \geq 0} \left| \frac{G(x_t + \tau \mid \mathcal{F}_{t-1}) - G(\tau_t \mid \mathcal{F}_{t-1})}{1 - G(\tau_t \mid \mathcal{F}_{t-1})} - P(x_t; \delta_t, \xi_t) \right| = 0, \end{aligned} \tag{1}$$

for parameters $\xi_t = \alpha_t^{-1}$ and δ_t, both possibly depending on τ_t, and with Y_t denoting the random variable corresponding to the realization y_t. Here, $P(x_t; \delta_t, \xi_t)$ denotes the cdf of the Generalized Pareto Distribution (GPD), with cdf and pdf given by

$$P(x_t; \delta_t, \xi_t) = 1 - \left(1 + \xi_t \frac{x_t}{\delta_t}\right)^{-\xi_t^{-1}}, \qquad p(x_t; \delta_t, \xi_t) = \delta_t^{-1} \cdot \left(1 + \xi_t \frac{x_t}{\delta_t}\right)^{-\xi_t^{-1} - 1}, \tag{2}$$

respectively (see, for example, McNeil et al., 2010). The quantity $x_t = y_t - \tau_t > 0$ is the so-called peak-over-threshold (POT), or exceedance, of heavy-tailed data y_t over a predetermined threshold τ_t, and $\delta_t > 0$ and $\xi_t > 0$ are the tail scale and tail shape parameter of the GPD, respectively. Most continuous distributions used in statistics and the actuarial sciences lie in the Maximum Domain of Attraction (MDA) of the GPD (see McNeil et al., 2010, Chapter 7.1), meaning that they allow for the above tail shape approximation. By focusing on the tail area directly using EVT arguments, we avoid having to make more ad-hoc assumptions on the parametric form of the tail.

The result in (1) is a limiting result. In any finite sample, the threshold τ_t has to be set to a specific, finite value, such that the GPD approximation will be inexact and the distribution is in that sense misspecified. This will also be the case in our setting. The score-driven updates that we define later on for ξ_t and δ_t, however, still ensure that the expected Kullback-Leibler divergence between the approximate GPD model and the true,

unknown conditional distribution $\mathbb{P}\left[Y_t \leq x_t + \tau_t \mid Y_t > \tau_t,\ \mathcal{F}_{t-1}\right]$ is improved on average at every step for sufficiently small steps, even if the GPD model is (partially) misspecified; see Blasques et al. (2015).

The choice of the threshold τ_t is subject to a well-known bias-efficiency trade-off; see, for instance, McNeil and Frey (2000). In theory, the GPD tail approximation only becomes exact for $\tau_t \to +\infty$. A high threshold, however, also implies a smaller number of exceedances $\varepsilon_t > \tau$, and more estimation error for the parameters of the GPD. Common choices for τ_t from the literature are the 90%, 95%, and 99% empirical quantiles of y_t; see Chavez-Demoulin et al. (2014). In our setting, such choices are less useful as τ_t varies over time. As we explain later, we use the approach of Patton et al. (2019) to set τ_t dynamically in line with the data using one of the above-mentioned quantiles.

2.1.2 Time-varying parameters

A key step in (1)–(2) is that we use the conditional probabilities based on the information set $\mathcal{F}_{t-1}$. As a result, the tail parameters ξ_t and δ_t become time-varying. To capture this time-variation, we model $(\xi_t, \delta_t)'$ using the score-driven dynamics introduced by Creal et al. (2013) and Harvey (2013). In our time series setting, this implies that both δ_t and ξ_t are measurable with respect to $\mathcal{F}_{t-1}$. We ensure positivity of δ_t and ξ_t by using an (element-wise) exponential link function $(\xi_t, \delta_t)' = \exp(f_t)$ for $f_t = (f_t^{\xi}, f_t^{\delta})' \in \mathbb{R}^2$. The transition dynamics for f_t are given by

$$f_{t+1} = \omega + \sum_{i=0}^{q-1} A_i s_{t-i} + \sum_{j=0}^{p-1} B_j f_{t-j}, \tag{3}$$

$$s_t = \mathcal{S}_t \nabla_t, \qquad \nabla_t = \partial \ln p(x_t \mid \mathcal{F}_{t-1}; f_t, \theta)/\partial f_t,$$

where vector $\omega = \omega(\theta)$ and matrices $A_i = A_i(\theta)$ and $B_j = B_j(\theta)$ depend on the static parameter vector θ, which needs to be estimated. The scaling matrix $\mathcal{S}_t$ may depend both on θ, f_t, and $\mathcal{F}_{t-1}$. Effectively, the recursion (3) updates f_t at every point in time via a scaled steepest ascent step to improve the expected fit to the GPD; see Blasques et al. (2015); Creal

et al. (2018). The score of (2) required in (3) is given by

$$\nabla_t = \begin{bmatrix} \xi_t^{-1} \cdot \ln\left(1 + \xi_t\, \delta_t^{-1} x_t\right) - \left(1 + \xi_t^{-1}\right) \dfrac{\xi_t x_t}{\delta_t + \xi_t x_t} \\ \dfrac{x_t - \delta_t}{\delta_t + \xi_t x_t} \end{bmatrix}, \tag{4}$$

where $\ln(\cdot)$ denotes the natural logarithm; see Appendix A.1 for a derivation. We take A_i and B_j as diagonal matrices.

Following Creal et al. (2014) we select the square-root inverse conditional Fisher information of the conditional observation density to scale (4), i.e., $\mathcal{S}_t = L_t'$, with L_t the choleski decomposition of the inverse conditional Fisher information matrix $\mathcal{I}_t = (L_t L_t')^{-1} = \mathbb{E}[\nabla_t \nabla_t' \mid \mathcal{F}_{t-1}; f_t, \theta] = \mathbb{E}[-\partial \nabla_t / \partial f_t' \mid \mathcal{F}_{t-1}; f_t, \theta]$. Compared to so-called inverse information matrix scaling, the current scaling matrix has the advantage that the conditional variance of the scaled score s_t is the unit matrix, i.e., $\mathbb{E}[s_t s_t' \mid \mathcal{F}_{t-1}] = \mathrm{I}_2$. This gives the parameters A_i a more natural interpretation, similar to the standard deviations of the state innovations in a non-linear state-space model. For the GPD, we have

$$L_t = \begin{bmatrix} 1 + \xi_t^{-1} & 0 \\ -1 & \sqrt{1 + 2\xi_t} \end{bmatrix}, \tag{5}$$

see Appendix A.2 for a derivation. Combining terms yields the scaled score

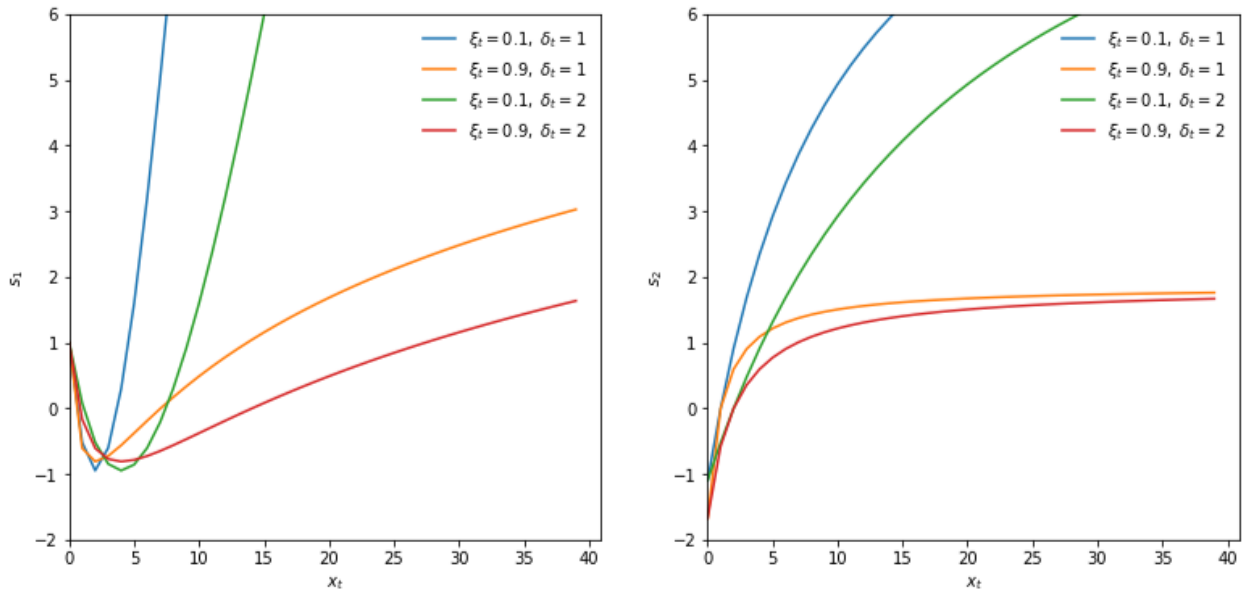
$$s_t = L_t' \nabla_t = \begin{bmatrix} \xi_t^{-2}(1 + \xi_t) \cdot \ln\left(1 + \xi_t\, \delta_t^{-1} x_t\right) + \dfrac{\delta_t - (\xi_t + 3 + \xi_t^{-1}) \cdot x_t}{\delta_t + \xi_t x_t} \\ \sqrt{1 + 2\xi_t} \dfrac{x_t - \delta_t}{\delta_t + \xi_t x_t} \end{bmatrix}. \tag{6}$$

Though the scaled score in (6) seems unstable for ξ_t near zero, the expression actually has a finite left limit equal to $\lim_{\xi_t \downarrow 0} s_{1,t} = 1 - 2\delta_t^{-1} x_t + \frac{1}{2}\delta_t^{-2} x_t^2$.

Figure 1 plots the two elements of (6) as a function of x_t for different values of ξ_t and δ_t. The behavior of the scaled score is intuitive: large x_ts imply that both ξ_t and δ_t are adjusted upwards and tails thus become fatter. The adjustments are largest when the current tail

Figure 1: News impact curves

The first element (left panel) and second element (right panel) of s_t in (6) is plotted against x_t for different values of ξ_t and δ_t.



index ξ_t and tail scale δ_t are low. It is precisely in such cases that observing a large x_t is unlikely. If it occurs nonetheless, the parameters are (strongly) adjusted to account better for similar effects in the future.

The news impact curves become increasingly concave for lower values of ξ_t and δ_t. For such parameters, large x_t values are already more likely due to the fat-tailed nature of the GPD itself. As a result, the parameters need to be adjusted less if a large x_t actually materializes. This resembles the well-known robustness properties of score-driven updates in the context of time-varying volatility modeling; see Creal et al. (2013) and Harvey (2013). It also distinguishes our current set-up sharply from an approach directly based on quantile functions; see Patton et al. (2019) and Catania and Luati (2019), in particular for risk measures such as ES. In Patton et al. (2019), ES reacts linearly rather than concavely to the VaR exceedance. This can result in noisy or unstable ES estimates. Figure 1 illustrates that our score-driven approach is less susceptible to such instabilities and can therefore result in more interpretable parameter paths.

We also note that small realizations of x_t imply downward adjustments of both ξ_t and

δ_t, up to the point where x_t becomes very small. For very small $x_t > 0$, ξ_t is adjusted upward: observations near the center of a fat-tailed distribution signal increased peakedness (=leptokurtosis) and thus higher ξ_t; compare Lucas and Zhang (2016) for a similar effect in the Student's t setting.

When there is no tail observation, i.e. $x_t = y_t - \tau_t \leq 0$, the new observation carries no information about ξ_t and δ_t; see McNeil et al. (2010, Chapter 7). In such cases we set the score to zero and continue to use (3) to update f_t. Long consecutive stretches of zero scores could potentially lead to erratic paths for f_t and thus (ξ_t, δ_t). If so, this issue can be addressed by taking into account lagged values of the scaled score via an exponentially-weighted moving average specification

$$\begin{aligned} f_{t+1} &= \omega + A\tilde{s}_t + Bf_t, \qquad \tilde{s}_t = (1-\lambda)s_t + \lambda\tilde{s}_{t-1}, \qquad \Rightarrow \\ f_{t+1} &= (1-\lambda)\omega + (1-\lambda)As_t + (\lambda \mathrm{I}_2 + B)f_t - \lambda B f_{t-1}, \end{aligned} \tag{7}$$

where $\lambda \in [0, 1)$ is an additional parameter to be estimated or fixed ex-ante; see Appendix A.3 for the derivation. This alters the values of $(p, q) = (1, 1)$ in equation (3) to $(p, q) = (2, 1)$. While s_t is most often zero, $\tilde{s}_t$ is not. Clearly, (3) is a special case of (7) for $\lambda = 0$. The smoothing approach in (7) is similar to the approach in Patton (2006) who uses up to ten lags of the driver (in our case the score) to smooth the dynamics of the time-varying parameter.

The transition equation for f_t can be extended further if additional conditioning variables are available by respecifying (7) as

$$f_{t+1} = \omega + A\tilde{s}_t + Bf_t + Cz_t, \tag{8}$$

where all explanatory variables are stacked into the vector z_t, and C is a conformable matrix of impact coefficients that needs to be estimated.

2.1.3 Time-varying thresholds

For the dynamic thresholds τ_t, we use the specification suggested by Patton et al. (2019),

$$\tau_{t+1} = \tau_t + a^\tau \cdot (1\{y_t > \tau_t\} - (1-\kappa)), \tag{9}$$

where a^τ is a parameter to be estimated, $1-\kappa$ denotes a sufficiently small tail probability corresponding to the dynamic quantile τ_t. We initialize τ_t by the unconditional κ-quantile of y_t. The recursive specification (9) is a martingale since $\mathbb{E}[1\{y_t > \tau_t\} \mid \mathcal{F}_{t-1}, \theta] = (1-\kappa)$. The threshold τ_t responds to quantile exceedances in an intuitive way by locally increasing the corresponding quantile. The transition equation for τ_t can easily be extended to include exogenous variables z_t as in

$$\tau_{t+1} = \tau_t + a^\tau \cdot (1\{y_t > \tau_t\} - (1-\kappa)) + c^{\tau\prime} \cdot z_t, \tag{10}$$

for a suitable vector of coefficients c^τ.

2.1.4 Interpretation of time-varying parameters

It is important to briefly comment on parameter interpretability. The tail shape parameter ξ_t can always be interpreted as observation y_t's contemporaneous inverse tail index α_t^{-1}. By contrast, the estimated tail scale parameter δ_t need not have a straightforward interpretation in terms of y_t's scale σ_t. For example, assume that y_t has a GPD distribution with time-varying tail shape parameter α_t^{-1} and scale σ_t. Web Appendix B.1 demonstrates that the tail probability $\mathbb{P}[Y_t \leq y_t + \tau_t \mid Y_t > \tau_t, \mathcal{F}_{t-1}]$ then also has a GPD shape (exactly, not only approximately). The tail shape parameter is the same as that of the *center*: $\xi_t = \alpha_t^{-1}$. However, the tail scale parameter δ_t is very different from the scale parameter σ_t that applies to the center, in particular $\delta_t = \sigma_t + \alpha_t^{-1} \cdot \tau_t$. As a result, δ_t increases with the threshold τ_t, varies positively with the tail shape parameter ξ_t, and, importantly, should not be expected to provide a consistent estimate of the scale parameter σ_t that applies to the *center* of the distribution of y_t. A similar result can be derived if y_t were Student's t-distributed with scale

σ_t and degrees of freedom parameter α_t when the tail probability $\mathbb{P}[Y_t \leq y_t + \tau_t \mid Y_t > \tau_t, \mathcal{F}_{t-1}]$ only has an approximate GPD shape; see Web Appendix B.2. We return to this issue in our simulation Section 3, where we consider pseudo-true values for ξ_t and δ_t to benchmark how well the model performs in terms of tracking an unknown data generating process.

2.2 Stationarity

The score-driven dynamics for ξ_t and δ_t are highly non-linear. Still the structure of the model allows us to obtain clear conditions for stationarity and ergodicity. The result is given in the following theorem.

Theorem 1. *If x_t is generated by the GPD with score-driven time varying parameters $f_t = (\ln \xi_t, \ln \delta_t)$ as in equations (3)–(7), then f_t and x_t are stationary and ergodic (SE) if*

$$\mathbb{E} \ln \sup_{\xi \geq 0} \Big| b^{\xi} + a^{\xi}(1-\lambda) \cdot \big((3\xi^{-1} + 2\xi^{-2})(1 - \exp(-\xi E)) - (2\xi + 3 + \xi^{-1}) E \exp(-\xi E) - \xi^{-1} E \big) \Big| < 0, \tag{11}$$

$$|b^{\delta}| < 1, \tag{12}$$

where E is a standard exponentially-distributed random variable.

The proof can be found in Web Appendix C and is based on the contraction conditions in Bougerol (1993), similar to the work of Blasques et al. (2022). Interestingly, the condition for δ_t in (12) is simple and standard. The condition for ξ_t is somewhat more involved and bears resemblance to the condition of Nelson's triangle for GARCH processes; see for instance Francq and Zakoian (2019). A sketch of the region is given in Web Appendix C.

Three items deserve comment. First, the conditions in Theorem 1 are *sufficient* only. The ξ_t and δ_t processes could thus continue to behave well if the conditions were not met. We investigate this in our simulation section. Second, for a full theoretical asymptotic analysis of the model and the estimation procedure, one would also need to establish filter invertibility and the existence of specific moments of x_t, ξ_t, and δ_t and their derivatives with respect to the static parameters; compare the full analysis in Blasques et al. (2022). We leave these

topics for future research. Finally, the stationarity and ergodicity of x_t and f_t are not always needed to prove consistency and asymptotic normality of the static parameter estimates; see for instance the work of Li et al. (2018) in the integrated volatility (iGARCH) context.

2.3 Parameter estimation

Parameter estimates can be obtained by standard numerical maximization of the log-likelihood function. Observation-driven time series models such as (1) – (10) have the appealing feature that the log-likelihood is known in closed form. For a given set of time series observations $x_t = 1\{y_t > \tau_t\} \cdot (y_t - \tau_t)$ for $t = 1, \ldots, T$, the vector of unknown parameters θ can be estimated by maximizing the log-likelihood function of the GPD with respect to θ. The average log-likelihood function is given by

$$\begin{aligned}\mathcal{L}(\theta|\mathcal{F}_T) &= \frac{1}{T^*}\sum_{t=1}^{T} 1\{x_t > 0\} \cdot \ln p(x_t; \delta_t, \xi_t) \\ &= \frac{1}{T^*}\sum_{t=1}^{T} 1\{x_t > 0\} \cdot \left[-\ln(\delta_t) - \left(1 + \frac{1}{\xi_t}\right) \ln\left(1 + \xi_t \frac{x_t}{\delta_t}\right)\right], \end{aligned} \tag{13}$$

where $T^* = \sum_{t=1}^{T} 1\{x_t > 0\}$ is the number of POT values in the sample. Maximization of (13) can be carried out using a conveniently chosen quasi-Newton optimization method.

Blasques et al. (2022) provide conditions under which the maximum likelihood estimator of θ is consistent and asymptotically normally distributed within the class of score-driven models. They also prove that (quasi-)maximum likelihood estimation of θ can remain consistent (to pseudo-true values) and asymptotically normal even if the score-driven model is misspecified. This is reassuring since the GPD is never exact for any finite value of $\tau_t < \infty$. In the presence of misspecification, score updates continue to minimize the local Kullback-Leibler divergence between the true conditional density and the model-implied conditional density, and remain optimal in this sense; see Blasques et al. (2015). The asymptotic covariance matrix $W = \text{Var}(\hat{\theta})$ takes its usual sandwich form; see e.g. Davidson and MacKinnon (2004, Ch. 10) and Blasques et al. (2022).

The autoregressive parameter a^τ in (9) cannot be estimated using (13). Another objective

function is needed in this case. We suggest using the average quantile regression check function of Koenker (2005, Ch. 3). The optimization problem can be formulated as

$$\min_{\{a^\tau\}} \frac{1}{T}\sum_{t=1}^{T} \rho_\kappa\left(y_t - \tau_t\right) \iff \min_{\{a^\tau\}} \frac{1}{T}\sum_{t=1}^{T} (y_t - \tau_t)\left(\kappa - 1\{y_t < \tau_t\}\right)$$
$$\iff \max_{\{a^\tau\}} \frac{1}{T}\sum_{t=1}^{T} (y_t - \tau_t)\left((1-\kappa) - 1\{y_t > \tau_t\}\right), \tag{14}$$

where $\rho_\kappa(u_t) = u_t\left(\kappa - 1\{u_t < 0\}\right)$, and τ_t evolves as in (9). See also Engle and Manganelli (2004) and Catania and Luati (2019) for the use of this objective function in a different dynamic context. In practice, we estimate all thresholds τ_t via (14) before maximizing (13).

2.4 Confidence bands for tail shape and tail scale

Given the maximum likelihood estimate $\hat{\theta}$, confidence (or standard error) bands around $\hat{f}_t = f_t(\hat{\theta})$ allow us to visualize the impact of estimation uncertainty on the filtered estimates thus on $\hat{\xi}_t = \xi_t(\hat{\theta})$ and $\hat{\delta}_t = \delta_t(\hat{\theta})$. Quantifying the uncertainty of the estimated parameter paths is important as classical EVT estimators of time-invariant tail shape parameters are already typically associated with sizeable standard errors; see e.g. Hill (1975) and Huisman et al. (2001).

Our confidence bands are based on the variance of $\hat{f}_t$, which we denote by $V_t = \mathrm{Var}(\hat{f}_t)$. There exist two possible ways to construct these bands. Delta-method-based bands can be devised using a linear approximation of the non-linear transition function for f_t, thus extending Blasques et al. (2016, Section 3.2) to the case of multiple lags. We provide the equations in Web Appendix D. In our empirical illustrations below, however, the linear approximations are typically insufficient to capture the uncertainty in the highly non-linear and persistent dynamics of $\hat{f}_t$; compare Figure 1. As a result, delta-method-based bands can become unstable. Therefore, we instead use simulation-based bands as in Blasques et al. (2016, Section 3.3).

Simulation-based confidence bands build on the asymptotic normality of $\hat{\theta}$. In particular,

we draw S parameter values $\hat{\theta}^s$, $s = 1, \ldots, S$ from the distribution $\mathrm{N}(\hat{\theta}, \hat{W})$, where $\hat{W}$ is the estimated covariance matrix of $\hat{\theta}$ as obtained via the sandwich covariance matrix estimator or via a bootstrapping procedure. If the finite-sample distribution of $\hat{\theta}$ were known, that could be used instead. For each draw $\hat{\theta}^s$ we run the filter $\hat{f}_t^s = f_t(\hat{\theta}^s)$ for $t = 1, \ldots, T$. This way, we obtain S time-varying parameter paths $\hat{f}_t^s$ for $s = 1, \ldots, S$ and $t = 1, \ldots, T$. These paths account automatically for all non-linearities in the dynamics of f_t. We obtain the pointwise simulated uncertainty bands of $\hat{f}_t$ by directly calculating the appropriate percentiles over the S draws of the $\hat{f}_t^s$ for each t.

2.5 Market risk measures

Market risk measurement is a major application of EVT methods in practice; see Manganelli and Engle (2004) and McNeil et al. (2010). We consider the conditional VaR and conditional ES as measures of one-step-ahead market risk. The GPD approximation (1)–(2) yields useful closed-form estimators of the VaR and ES for high upper quantiles $\gamma > G(\tau_t \mid \mathcal{F}_{t-1})$; see McNeil and Frey (2000) and Rocco (2014). We can estimate the $1-\gamma$ tail probability of y_t based on the GPD cdf for x_t, obtaining

$$\mathrm{VaR}^{\gamma}(y_t \mid \mathcal{F}_{t-1}, \theta) = \tau_t + \delta_t \xi_t^{-1} \left[\left(\frac{1-\gamma}{t^*/t} \right)^{-\xi_t} - 1 \right], \tag{15}$$

where t^* is the number of observations with $x_t > 0$ up to time t, i.e., the number of observations y_s for $s = 1, \ldots, t$ for which $y_s > \tau_s$. Put differently, t^*/t is an estimator of the tail probability $G(\tau_t \mid \mathcal{F}_{t-1})$.

The conditional ES is the average conditional VaR in the tail across all quantiles γ (see McNeil et al., 2010, Chapter 2), provided $\xi_t < 1$. The closed-form expression is

$$\begin{aligned} \mathrm{ES}^{\gamma}(y_t \mid \mathcal{F}_{t-1}, \theta) &= \frac{1}{1-\gamma} \int_{\gamma}^{1} \mathrm{VaR}^{\tilde{\gamma}}(y_t \mid \mathcal{F}_{t-1}, \theta) \mathrm{d}\tilde{\gamma} \\ &= \frac{\mathrm{VaR}^{\gamma}(y_t \mid \mathcal{F}_{t-1}, \theta)}{1-\xi_t} + \frac{\delta_t - \xi_t \tau_t}{1-\xi_t}, \end{aligned} \tag{16}$$

see Web Appendix E for a derivation of (15) and (16). The $\mathrm{ES}^{\gamma}(y_t \mid \mathcal{F}_{t-1}, \theta)$ is strictly higher

than the $\text{VaR}^{\gamma}(y_t \mid \mathcal{F}_{t-1}, \theta)$ at the same confidence level, as it "looks further into the tail." It can be shown that the ratio $\text{ES}^{\gamma}(y_t \mid \mathcal{F}_{t-1}, \theta)/\text{VaR}^{\gamma}(y_t \mid \mathcal{F}_{t-1}, \theta)$ increases monotonically in ξ_t for $\gamma \to 1$, indicating that expected losses beyond the VaR become increasingly worse for heavier-tailed (higher ξ_t) distributions. Maximum likelihood estimators of the conditional VaR and conditional ES can be obtained by inserting filtered estimates of ξ_t and δ_t into (15) and (16), respectively.

3 Simulation study

3.1 Simulation design

This section investigates the ability of our dynamic EVT model to simultaneously recover (i) the time variation in tail shape and tail scale parameters ξ_t and δ_t, (ii) EVT-based market risk measures $\text{VaR}^{\gamma}(y_t; \tau_t, \xi_t, \delta_t)$ and $\text{ES}^{\gamma}(y_t; \tau_t, \xi_t, \delta_t)$ at high confidence levels such as $\gamma = 99\%$, and (iii) parameter estimates and their standard errors for all static parameters collected in θ. We do so using two sets of data generating processes (DGPs). In these experiments we track the performance of our score-driven modeling approach when putting it into a variety of challenging settings, such as when the conditional density is only approximately correct, the time-varying parameter process is misspecified, and/or when both tail parameters follow similar paths. We investigate whether the EVT-based market risk estimates that follow from the model are reliable in such cases. Finally, we study inference on the static parameters when the autoregressive parameters b^{ξ} and b^{δ} are close or equal to one.

DGP 1: Our first simulation design considers $D = 2$ different densities (GPD and t), $P = 4$ different parameter paths for tail shape and scale, and $H = 3$ different ways to obtain the appropriate thresholds τ_t. This yields $2 \times 4 \times 3 = 24$ simulation experiments. In each experiment, we draw $S = 100$ univariate simulation samples of length $T = 25,000$. We focus on the upper $1 - \kappa = 5\%$ tail. As a result, approximately $25,000 \cdot 0.05 = 1,250$ observations are available in each simulation to compute informative POTs $x_t > 0$.

We first simulate y_t from a $\text{GPD}(\alpha_t^{-1}, \sigma_t)$ distribution with time-varying tail shape α_t^{-1}

and scale σ_t. We then consider a Student's t distribution with time-varying scale σ_t and degrees of freedom α_t. In the GPD case, our score-driven model uses the exact conditional density for x_t, while in the Student's t case the GPD conditional density for x_t is only approximately correct (given finite thresholds τ_t in any given sample); see Web Appendix B. POT values x_t are obtained as $x_t = y_t - \tau_t$.

We consider four different paths for the tail shape α_t^{-1} and scale σ_t parameters. For both the GPD and Student's t densities we consider

(1) Constant: $\alpha_t^{-1} = 0.5$, $\sigma_t = 1$;

(2) Sine and constant: $\alpha_t^{-1} = 0.5 + 0.3\sin(4\pi t/T)$, $\sigma_t = 1$;

(3) Slow sine and frequent sine: $\alpha_t^{-1} = 0.5 + 0.3\sin(4\pi t/T)$, $\sigma_t = 1 + 0.5\sin(16\pi t/T)$;

(4) Synchronized sines: $\alpha_t^{-1} = 0.5 + 0.3\sin(4\pi t/T)$, $\sigma_t = 1 + 0.5\sin(4\pi t/T)$.

Path (1) considers the special case of time-invariant tail shape and scale parameters. Naturally, we would want our dynamic framework to cover constant parameters as a special case. Path (2) allows the tail shape to vary considerably between 0.2 and 0.8, while keeping the scale (volatility) of the data constant. Path (3) stipulates that both parameters vary over time. Finally, Path (4) considers the case of synchronized variation in both parameters. This last setting may be particularly challenging for two reasons. First, the tail observations occur most frequently when both tail shape and scale are high, making it potentially difficult to disentangle the two effects. Second, less information about the tail is available when both parameters are low simultaneously.

We consider three ways to construct the thresholds τ_t. First, we use the true time-varying 95%–quantile based on our knowledge of the true density and of α_t and σ_t. This constitutes an infeasible best benchmark. Second, we construct τ_t as the 95%–quantile of the expanding window of data up to time t, i.e. $\tau_t = Q_{1:t}^{0.95}(\{y_1, \ldots, y_t\})$. Finally, we use the recursive specification (9), initialized at the full-sample quantile $\tau_1 = Q_{1:T}^{0.95}$.

Our main evaluation metric for evaluating model performance is the root mean squared error RMSE $= \frac{1}{S}\sum_{s=1}^{S}\sqrt{\frac{1}{T}\sum_{t=1}^{T}(\hat{\xi}_t^s - \bar{\xi}_t^s)^2}$, where $\hat{\xi}_t^s$ is the estimated tail shape parameter in simulation s, $\bar{\xi}_t^s$ is the corresponding (pseudo-)true tail shape, $s = 1, \ldots, S$ denotes the

simulation run, and $t = 1, \ldots, T$ is the number of observations in each draw. The RMSE for the tail scale parameter δ_t is obtained analogously. The pseudo-true values $\bar{\xi}_t^s$ and $\bar{\delta}_t^s$ are obtained by numerically minimizing the Kullback-Leibler divergence between the GPD and the data generating process beyond the true time-varying 95% quantile τ_t. As the true conditional density is known at all times in a simulation setting, these pseudo-true benchmarks are easily computed numerically for each s and t. We note that particularly the GPD tail scale parameter $\bar{\delta}_t$ may have very different dynamics from σ_t, as it combines dynamics in α_t and σ_t via the EVT limiting expression in (1).

DGP 2: In our second design, we investigate the effect of covariates and unit root type dynamics on the dynamic parameter paths and the static parameter estimates and their standard errors. We simulate 100 samples from a GPD$(x_t; \xi_t, \delta_t)$ density, with T=25,000 observations each, thus abstracting from any misspecification effects. The factor $f_t = (\ln \xi_t, \ln \delta_t)'$ follows the transition equation

$$f_{t+1} = \omega + A s_t + B f_t + C z_t, \tag{17}$$

where matrices ω, A, B, and C take four different sets of values. As a first case, we consider a slowly mean-reverting factor process with $\omega = (0.50, 1.00)'$, $A = \text{diag}(0.03, 0.07)$, $B = \text{diag}(0.98, 0.98)$, and $C = 0$. The second case considers an integrated factor process: $B = \text{I}_2$, while ω, A, and C remain unchanged. A third and fourth case are identical to the first and second case, except that now $C = (-3, -1.5)'$ in (8) for an observed variable z_t. As our z_t we use the central bank purchases of Italian sovereign bonds as considered in our second application in Section 4.2. We refer to cases two and four as "iGAS" (integrated GAS) models, in line with the non-stationary factor paths implied by $B = \text{I}_2$. We know from the volatility literature that such non-stationary behavior can still yield consistent estimates of the remaining static parameters; see for instance Li et al. (2018). In our context, we are interested in what happens to the estimates of the static parameter and their standard errors when we mistakenly impose $B = \text{I}_2$ when $B < \text{I}_2$, and vice versa.

Table 1: RMSE outcomes for DGP1

Root mean squared error (RMSE) statistics for two different distributions (GPD and t, in columns) and for four different parameter paths for tail shape ξ_t and tail scale δ_t (paths (1) – (4), in rows). Thresholds τ_t, $\hat{\tau}_t$, and $\hat{\tau}_t^*$ denote, respectively, *i*) the infeasible true time-varying threshold, *ii*) the empirical quantile associated with an expanding window of observations $y_1, \ldots, y_t$, and *iii*) the estimated conditional quantile using (14) and a suitably calibrated $a^\tau = 0.25$ to speed up the computations. We consider 100 simulations for each DGP, and a time series of $25,000$ observations in each simulation. Model performance is measured by the RMSE from the true $\bar{\xi}_t$ and $\bar{\delta}_t$ in each draw.

Model	GPD(τ_t) (infeasible)	GPD($\hat{\tau}_t$)	GPD($\hat{\tau}_t^*$)	t(τ_t) (infeasible)	t($\hat{\tau}_t$)	t($\hat{\tau}_t^*$)
			RMSE $\hat{\xi}_t$			
(1)	0.000	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
(2)	0.171	0.177	0.178	0.182	0.188	0.189
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
(3)	0.182	0.188	0.189	0.190	0.197	0.197
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
(4)	0.177	0.186	0.183	0.188	0.195	0.192
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
			RMSE $\hat{\delta}_t$			
(1)	0.005	0.014	0.068	0.005	0.010	0.034
	(0.003)	(0.006)	(0.013)	(0.002)	(0.004)	(0.006)
(2)	1.646	1.774	1.753	0.580	0.589	0.588
	(0.034)	(0.040)	(0.036)	(0.013)	(0.012)	(0.013)
(3)	2.421	2.913	2.813	0.836	0.960	0.924
	(0.054)	(0.054)	(0.049)	(0.015)	(0.020)	(0.017)
(4)	2.608	2.904	2.844	0.925	0.970	0.964
	(0.057)	(0.059)	(0.059)	(0.020)	(0.020)	(0.022)

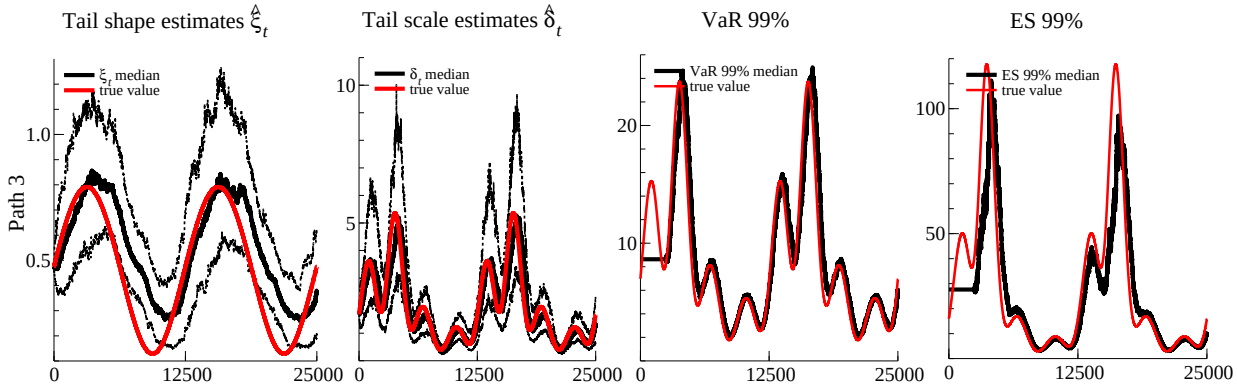
3.2 Simulation results

DGP 1: For the first DGP setting, we are particularly interested in two issues: first, what is the effect of increasing misspecification by moving from a GPD to a Student's t density in the data generating process of y_t, and second, how accurately can we recover high-confidence market risk measures when the conditional GPD density is only approximately correct.

Table 1 presents the results for DGP1. It reports RMSE statistics for tail shape $\hat{\xi}_t$ and tail scale $\hat{\delta}_t$. Figure 2 provides a representative example of the simulation outcomes where we compare median estimated parameter paths for $\hat{\xi}_t$, $\hat{\xi}_t$, $\widehat{\text{VaR}}^{0.99}$, and $\widehat{\text{ES}}^{0.99}$ to their (pseudo-)true values. Full results are found in Web Appendix F. The true parameters in Figure 2 follow Path 3 from Section 3.1, and time-varying thresholds are estimated based on the recursive specification (9) and objective function (14).

Figure 2: Simulation results: an example

Time series data is here generated as $y_t \sim \text{t}(0, \sigma_t, \alpha_t)$, where $\alpha_t^{-1} = 0.5 + 0.3\sin(4\pi t/T)$ and $\sigma_t = 1 + 0.5\sin(16\pi t/T)$. This is Path 3 in Section 3.1. Pseudo-true parameter values are reported in solid red. The four panels report estimates of ξ_t, δ_t, VaR_t, and ES_t, respectively. Median filtered values are plotted in solid black. The first two panels also indicate the lower 5% and upper 95% quantiles of the estimates (black dots). The time-varying threshold $\hat{\tau}_t$ is estimated based on the recursive specification (9) in conjunction with the objective function (14).



We focus on three main findings. First, all models seem to work well in recovering the true underlying ξ_t and δ_t dynamics. The median estimates in Figure 2 tend to be close to their (pseudo-)true values. The full results in the Web Appendix confirm this. Even the highly non-linear patters of δ_t are recovered well. The model also captures the peaks of ξ_t, which correspond to the episodes with the fattest tails. The model needs some time to recognize that the extreme tail has become more benign, i.e., that ξ_t has gone down. The good fit is corroborated by Table 1. We also note that both estimation methods for τ_t only loose about 10% RMSE for ξ_t and δ_t compared to the use of the true (infeasible) τ_t.

Second, when comparing the results for ξ_t and δ_t based on the recursive estimate $\hat{\tau}_t$ and the dynamic estimate τ_t^* of Patton et al. (2019), Table 1 shows differences are mostly small and insignificant. If there is no time-variation (Path 1), the estimates of $\hat{\delta}_t$ based on recursive $\hat{\tau}_t$ fare slightly better (as expected). The converse is true for if the true parameters vary over time (Paths 2–4).

Third, Figure 2 shows that EVT-based market risk measures such as high-confidence level ($\gamma = 0.99$) VaRs and ESs tend to be estimated sufficiently accurately with our dynamic EVT approach. Again, this is confirmed by the full results in Web Appendix F. Both the

Table 2: RMSE outcomes for DGP2

The entries in the table are the RMSEs associated with the filtered tail parameters ξ_t and δ_t and with the estimates of the static parameters a^ξ and a^δ, based on simulations. Top panel: We simulate from iGAS or GAS models (columns 2–5 and 6–9) and estimate back both iGAS and GAS models (rows 4 and 5). Bottom panel: We simulate as before, but also include an additional explanatory covariate z_t in both the DGP and empirical model. These extended models are labeled iGAS-X and GAS-X.

Model	DGP							
	$\hat{\xi}_t$	$\hat{\delta}_t$	a^ξ	a^δ	$\hat{\xi}_t$	$\hat{\delta}_t$	a^ξ	a^δ
	iGAS				GAS			
iGAS	0.047	0.192	0.005	0.008	0.097	0.287	0.008	0.011
GAS	0.124	0.171	0.018	0.014	0.052	0.052	0.012	0.010
	iGAS-X				GAS-X			
iGAS-X	0.071	0.176	0.008	0.009	0.172	0.167	0.010	0.011
GAS-X	0.125	0.204	0.016	0.010	0.061	0.056	0.013	0.011

low and high frequency dynamics of VaR and ES are captured well. There only appears some under-estimation of the ES at its highest peak, where tails are extremely fat. Overall, we conclude that the model captures the dynamics of the tails accurately, even in cases where the model does not coincide with the true, unobserved data generating process and the model is thus misspecified.

DGP 2: We now turn to the results for DGP2. Table 2 presents RMSEs associated with the time-varying parameters ξ_t and δ_t and the static parameters a^ξ and a^δ. We consider two settings: with (bottom panel) and without (top panel) a covariate. Within each of these, we consider a stationary (GAS) and unit root (iGAS) DGP, as well as a (possibly) stationary and unit root model specification. Figure F.3 in Web Appendix F.2 provides more results in the form of representative draws of $\hat{\xi}_t$ and $\hat{\delta}_t$ for each of the four cases, and results on standard error estimates for a^ξ and a^δ.

We focus on two main results. First, Table 2 suggests that both the GAS and the iGAS models work well if they are correctly specified (iGAS row and iGAS column, or GAS row and GAS column, etc.). In particular, the estimated $\hat{\xi}_t$ and $\hat{\delta}_t$ tend to be closely aligned to their true values. Also the (slightly) misspecified cases of a GAS model for an iGAS DGP and vice versa continue to work reasonably well: ξ_t and δ_t remain close to their true paths; see also Web Appendix F.

Second, when investigating the standard errors of the static parameter estimates, Table F.1 in Web Appendix F suggests that, while parameter point estimates are close to their true values, the usual asymptotic standard error estimates based on the inverse Hessian or the sandwich estimates are not necessarily reliable in the two iGAS cases. In our set-up, these common estimates of the standard errors are typically too large, providing too conservative inference. A bootstrap procedure tailored to integrated processes can be used to avoid this problem; see for instance Boswijk et al. (2021). We provide such a bootstrapping scheme in Web Appendix G and apply it in our second empirical illustration in Section 4.2.

4 Empirical illustrations

4.1 First illustration: equity log-returns

4.1.1 Data

To illustrate our approach, we obtain end-of-day prices for the S&P500 index and for IBM stock as two easily and publicly available series from the CRSP database. The S&P500 data range from July, 3 1962 to December, 31 2020, yielding 14,726 daily observations. The IBM stock data range from January 2, 1926 to December 31, 2020, yielding 25,028 daily observations. To model the adverse left tail of equity log-returns we consider *negative* log-returns $y_t = -100 \times (\ln p_t - \ln p_{t-1})$, with p_t the price level, before applying our methodology as described in Section 2. The VaR and ES for log-returns are then given by the negative of (15) and (16), respectively.

4.1.2 Static parameter estimates

We rely on the time-variation in the thresholds τ_t to accommodate time-variation in any parameters describing the center of the distribution. The thresholds evolve over time according to (9), at $(1-\kappa) = 10\%$, and are initialized at the $\kappa = 90\%$ empirical quantile of y_t. We initialize $f_t = (\ln \xi_t, \ln \delta_t)'$ at $f_0 = (\mathrm{I}_2 - B)^{-1}\,\omega$. (Alternatively, f_0 could be based on a static GPD model fitted to, say, the first year of y_t; $\hat{\theta}$ and $\hat{f}_t$ then remain similar.)

Table 3: Parameter estimates

Parameter estimates for the dynamic tail shape model. The second and third columns refer to the first application (equity log-returns of the S&P500 index and IBM stock). The estimation samples range from 3 July 1962 to 31 December 2020 for the S&P500 index, and from 2 January 1926 to 31 December 2020 for IBM stock. The remaining columns refer to the second application (changes in sovereign yields). Columns labeled IT 5y and PT 5y refer to yield changes for Italy and Portuguese five-year benchmark bonds sampled at the 15-minute frequency. The estimation samples range from 4 January 2010, 9AM, to 31 December 2012, 5PM. Standard error estimates are in round brackets, p-values are in square brackets. Standard errors and p-values are based on a sandwich covariance matrix estimator for the first application, and on a bootstrap procedure for the second application.

	First illustration		Second illustration			
	S&P500	IBM	IT 5y yield		PT 5y yield	
ω^{ξ}	-0.001	-0.001				
	(0.001)	(0.000)				
	[0.116]	[0.036]				
ω^{δ}	-0.008	-0.006				
	(0.002)	(0.002)				
	[0.000]	[0.001]				
a^{ξ}	0.024	0.023	0.019	0.019	0.037	0.042
	(0.009)	(0.004)	(0.002)	(0.002)	(0.002)	(0.005)
	[0.004]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
a^{δ}	0.139	0.126	0.081	0.078	0.099	0.095
	(0.013)	(0.017)	(0.003)	(0.003)	(0.005)	(0.004)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
b^{ξ}	0.9996	0.9997				
	(0.000)	(0.001)				
	[0.000]	[0.000]				
b^{δ}	0.991	0.988				
	(0.002)	(0.004)				
	[0.000]	[0.000]				
c^{ξ}				-6.187		-21.629
				(1.763)		(8.299)
				[0.000]		[0.009]
c^{δ}				-1.114		-14.964
				(0.675)		(3.774)
				[0.099]		[0.000]
λ	0	0	$0.05^{1/32}$	$0.05^{1/32}$	$0.05^{1/32}$	$0.05^{1/32}$
a^{τ}	0.231	0.306	0.341	0.341	0.566	0.566
c^{τ}			-0.041	-0.041	0.040	0.040
T	14,726	25,028	24,416	24,416	24,576	24,576
T^*	1,474	2,504	2,447	2,447	2,457	2,457
loglik	-21,895.1	-62,683.4	-81,057.0	-81,030.1	-198,641.8	-198,605.8
AIC	43,802.2	125,378.8	162,122.1	162,068.2	397,291.6	397,219.6
BIC	43,847.8	125,427.6	162,154.5	162,100.2	397,324.0	397,252.0

The first two columns in Table 3 present our estimates of the static parameters of model (1) – (9). Parameters a^{ξ} and a^{δ} are statistically significant at any reasonable significance level, for both S&P500 and IBM. These parameters can be interpreted as the average size of the scores driving $\ln \xi_t$ and $\ln \delta_t$, respectively; see the statements above equation (5). Parameters b^{ξ} and b^{δ} are close to one for both series, implying that shocks to each time-varying parameter die out only slowly (over months and years).

4.1.3 Tail parameter estimates

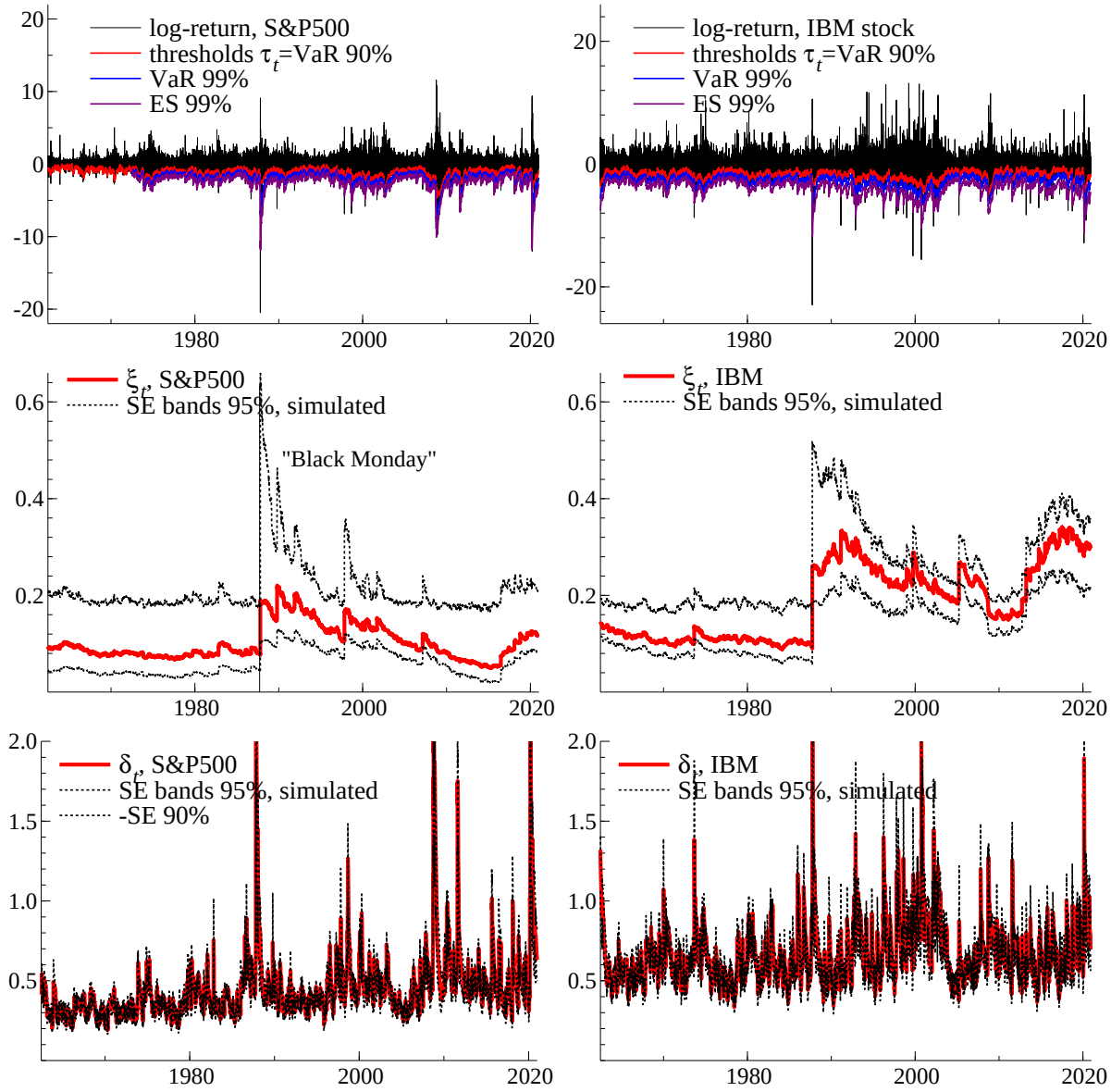
Figure 3 presents the raw log-returns (top panels) along with filtered estimates of ξ_t and δ_t (middle and bottom panels). The filtered tail shape varies between approximately 0.05 and 0.25 for the S&P 500 index, and between approximately 0.05 and 0.35 for IBM. The filtered tail scales vary roughly between 0.2 and 2.0 for the S&P 500 and IBM. The standard error bands around each filtered parameter suggest that both are reasonably precisely estimated, and that the tail shape parameter is often far from zero.

The filtered estimates of ξ_t and δ_t suggest that one-off, extremely negative returns affect the filtered tail shape more than the filtered tail scale. By contrast, longer lasting crises appear to affect the tail scale more than the tail shape. For example, the 1987 stock market crash, also known as "Black Monday" (October 19, 1987), considerably increases $\hat{\xi}_t$ but not $\hat{\delta}_t$. Similarly, the 2010 "flash-crash" on May 6, 2010 increases $\hat{\xi}_t$ more than $\hat{\delta}_t$. By contrast, the global financial crisis between 2008 and 2009, and the Covid-19 pandemic recession in early 2020, both temporarily increase $\hat{\delta}_t$ while leaving $\hat{\xi}_t$ less affected.

The patterns of $\hat{\xi}_t$ and $\hat{\delta}_t$ are intuitive even though a numerical check on the first condition in (11) suggests that the estimates do not satisfy the *sufficient* conditions for stationarity and ergodicity of the model. As mentioned earlier, stationarity of the model is not necessarily a required condition for the consistency and asymptotic normality of the remaining parameters in the model; compare Li et al. (2018). The distance of $(\hat{a}^{\xi}, \hat{b}^{\xi})$ to the SE region is modest for both log-returns. We confirm that, if a^{ξ} and b^{ξ} were restricted to lie inside the SE region, then the time-varying parameter paths presented in Figure 3 would remain similar, and the new $\hat{a}^{\xi}$ and $\hat{a}^{\delta}$ would remain statistically significant for both log-returns. Similarly, if ω in

Figure 3: Filtered tail parameters for S&P500 (left) and IBM (right) log-returns

Top panels: daily log-returns for the S&P500 index (left) and IBM common stock (right). Middle and bottom panels: filtered tail shape (ξ_t, middle) and tail scale (δ_t, bottom) parameters. The thresholds τ_t are reported at a 90% confidence level; Value-at-Risk (VaR) and Expected Shortfall (ES) are plotted at an extreme 99% confidence level (top panels). The thresholds τ_t, VaR, and ES are mirrored at the horizontal axis to correspond to log-returns (instead of percentage losses). The estimation samples range from July 3, 1962 to December, 31 2020 for the S&P500 index, and from January, 2 1926 to December, 31 2020 for the IBM stock. The reported samples range from July 3, 1962 to December, 31 2020.



(3) were chosen as $\omega = \omega_1 + \omega_2 \cdot 1\{t > \text{October 19, 1987}\}$, then $\hat{a}^{\xi}$ and $\hat{b}^{\xi}$ would again lie inside the SE region for IBM.

Figure 3 also plots estimates of VaR_t and ES_t at a 99.0% confidence level. We checked that indeed 1.0% of the log-returns lie beyond $\widehat{\text{VaR}}_t$ in the case of the S&P 500 index (0.9% for IBM stock). The average value of $-y_t$ conditional on it exceeding its VaR is -3.50% for the S&P 500, and -4.73% for IBM. These values are approximately in line with the time series average of $\widehat{\text{ES}}_t$ at -3.14% for the S&P 500 and -4.67% for IBM.

4.2 Second illustration: changes in sovereign yields

4.2.1 High-frequency data

Sovereign yields

In our second application we illustrate the inclusion of explanatory variables in the dynamics of ξ_t and δ_t as in equation (8). To do so, we study whether there was any tail risk impact of central bank asset purchases on changes in Italian (IT) and Portuguese (PT) five-year bond yields between 2010 and 2012. Both Italy and Portugal were in the "epicenter" of the existential euro area sovereign debt crisis at that time; see e.g. ECB (2013), ECB (2014), Eser and Schwaab (2016), Ghysels et al. (2017), and Pooter et al. (2018). Italy is an example for a large euro area country that was affected by the crisis relatively late (in mid-2011), and that benefited from Eurosystem bond purchases only during a relatively short period of time, between August 2011 and March 2012. Portugal, by contrast, is an example for a smaller euro area country that was affected relatively early (already in 2010), and that benefited from Eurosystem bond purchases more uniformly over time, between May 2010 and March 2012.

Eurosystem bond purchases undertaken during the sovereign debt crisis predominantly targeted the two- to ten-year maturity bracket, with the five-year maturity approximately in the middle of that spectrum; see e.g. Eser and Schwaab (2016). We consider the impact on five-year benchmark bonds for this reason.

The bond yields are sampled at the 15-minute frequency, between 9AM and 5PM, and

are obtained from continuous dealer quotes. We do not consider overnight changes in yield, such that the first 15-minute interval covers 9AM to 9:15AM. Our sample ranges from 04 January 2010 to 31 December 2012. This yields 32 intra-daily observations per trading day, with $T \approx 3 \times 260 \times 32 \approx 25,000$ observations per country.

Central bank bond purchases

To provide "depth and liquidity" to euro area bond markets that were strongly affected by the sovereign debt crisis, the Eurosystem acquired sovereign bonds within its Securities Markets Programme (SMP) between 2010 and 2012, including Italian and Portuguese bonds. The SMP had the objective of helping to restore the monetary policy transmission mechanism by addressing the malfunctioning of certain government bond markets. Implicit in the concept of malfunctioning markets is the notion that government bond yields can be unjustifiably high and volatile. Italian and Portuguese yields exhibited large and sudden moves in our sample, leading to extreme realizations of yield changes; see the top row of Figure 4. Both time series have significant non-Gaussian features under standard tests and significance levels (not reported). At the end of our sample, the Eurosystem held €99.0 bn in Italian sovereign bonds and €21.6 bn in Portuguese bonds; see the ECB (2013)'s Annual Report.

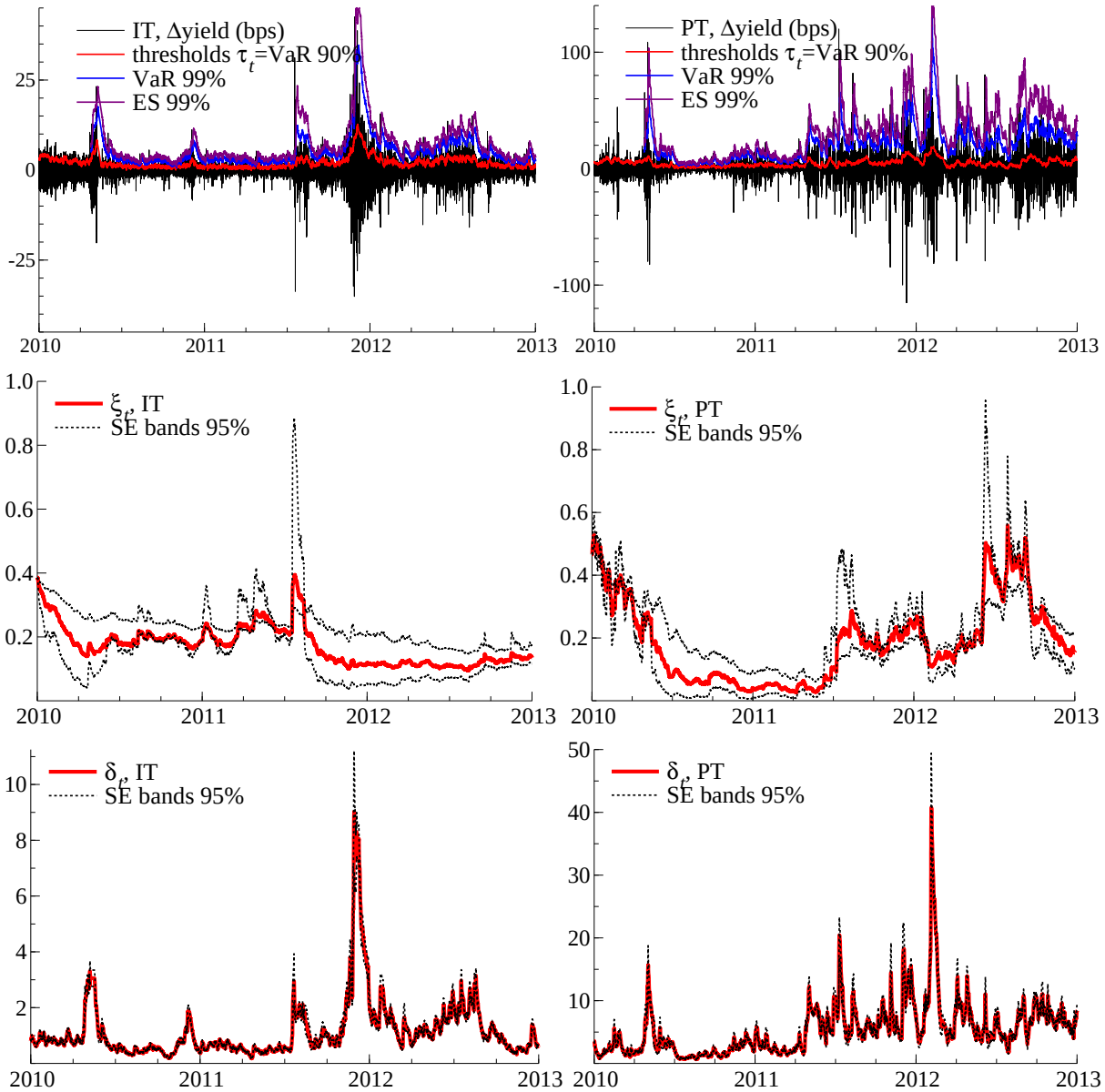
We construct time series data z_t of country-specific SMP purchases at the high (15-minute) frequency. The 15-minute frequency corresponds to the allowed latitude in the recording of trades by the Eurosystem. Observations z_t contain all sovereign bond purchases at par (nominal) value between $t-1$ and t for the respective country, not only purchases of the five-year benchmark bond. We including these as an additional conditioning variable in (8) to see whether they mitigated extreme tail behavior or not.

4.2.2 Static parameter estimates

Changes in the tail shape and tail scale parameters are highly persistent for the high-frequency data. Rather than estimating the persistence parameters, we therefore simplify the estimation by setting $\omega^{\xi} = \omega^{\delta} = 0$ and $b^{\xi} = b^{\delta} = 1$ as in Lucas and Zhang (2016), Lucas et al. (2019), and Joao et al. (2022). The choice $b^{\xi} = b^{\delta} = 1$ implies that both EVT

Figure 4: The tail shape and tail scale estimates

Top row: Five-year sovereign benchmark bond yields for Italy (IT, left column) and Portugal (PT, right column) between 2010 and 2012. Middle row: filtered tail shape (ξ_t) parameter. Bottom row: filtered tail scale (δ_t) parameter. Standard error bands are simulated at a 95% confidence level.



tail parameters remain at their most recently-filtered values, via (8), unless new information arrives from the tail (when $y_t > \tau_t$, or $z_t \neq 0$). This invariance of f_t to information coming from the center of $g(y_t \mid \mathcal{F}_{t-1})$ aligns with intuition from "static" EVT; see e.g. McNeil et al. (2010, Chapter 7). We also use the smoothing parameter λ in (6). As there are only log-likelihood contribution for the GPD when $x_t = z_t - \tau_t > 0$ is not missing, λ is hard to identify empirically. We therefore fix it at $\lambda = 0.05^{1/32} \approx 0.91$, such that most smoothing materializes within 32 15-minute intervals, i.e., one day. Fixing it at reasonable alternative values has little effect on our empirical findings.

Columns three to six of Table 3 present our estimates of the static parameters for the model (1) – (10). Columns three and five refer to a baseline model without central bank purchases. Given the potential issues with common Hessian-based standard errors as reported in our second simulation experiment in Section 3, we report bootstrapped standard errors for the static parameters using the procedure outlined in Web Appendix G; compare also Boswijk et al. (2021).

Parameters a^ξ and a^δ suggest pronounced and statistically significant time series variation in both the tail shape ξ_t and tail scale δ_t parameters, both of which are captured by our time-varying parameter model. The impact parameters c^ξ of bond purchases on tail shape are estimated to be negative in both cases. The estimates of c^δ are negative as well. We also note that bond purchases reduce the time-varying threshold $\hat{\tau}_t$ in the case of Italy. The increases in log-likelihood fit (26.9 points for IT, and 36.0 points for PT) are substantial. A comparison of model selection criteria (AIC, BIC) further supports the inclusion of central bank asset purchases as a useful covariate to explain each time series' extreme tail dynamics, thus illustrating how conditioning variables can help describe the evolution of extreme tail parameters.

4.2.3 Tail parameter estimates and VaR impact

Figure 4 plots the corresponding filtered estimates for time-varying tail shape ξ_t and tail scale δ_t. Time series variation is present and pronounced in both parameters. The filtered tail shape varies between approximately 0.1 and 0.4 for Italian yields, and between 0.05 and

0.6 for Portuguese yields. The filtered tail scale varies between approximately 1 and 10.0 for Italian yields, and between approximately 1 and 40 for Portuguese yields. The standard error bands around each time-varying parameter suggest that all are reasonably precisely estimated, and that the tail shape parameter is often far from the Gumbel case of $\xi_t = 0$.

As was clear from Table 3, the coefficients c^ξ and c^δ measure the impact of bond purchases on the tail behavior of bond changes. As these parameters are difficult to interpret by themselves, we show in Web Appendix H how they can be translated into an impact on VaR via its link to τ_t, δ_t, and ξ_t. For Italy, we obtain a total VaR impact of 0.0 (tail threshold) $-$ 3.5 (tail scale) $-$ 1.9 (tail shape) $= -5.4$ bps for a €1 bn Eurosystem intervention. For Portugal, we obtain a larger impact of 0.0 (tail threshold) $-$ 133.9 (tail scale) $-$ 24.0 (tail shape) $= -157.9$ bps. These point estimates are of course subject to substantial estimation uncertainty; see Table 3. The 95% confidence intervals for VaR impact can be bootstrapped along with the parameters, and are $[-9.5, -0.7]$ for IT and $[-234.8, -73.4]$ for PT. The stronger impact for Portugal than for Italy is likely due to a €1 bn intervention constituting a larger share of the overall market.

5 Conclusion

We introduced a semi-parametric modeling framework to study time variation in tail parameters for long univariate time series. To this end we modeled the time variation in the shape and scale parameters of the Generalized Pareto Distribution, which approximates the tail of most heavy-tailed densities used in econometrics and the actuarial sciences. We discussed the handling of non-tail time series observations, inference on static and time-varying parameters, and how to relate tail variation to observed covariates if such variables are available. The model therefore complements and extends recent work based on different methodologies, such as the non-parametric approach to tail index variation of de Haan and Zhou (2021), the time-varying quantile (and ES) approaches of Patton et al. (2019) and Catania and Luati (2019), and the parametric modeling approach of Massacci (2017). We applied the model to study variation in the left tail of U.S. equity log-returns, and in the right tail of changes

in euro area sovereign bond yields measured at a high frequency. In the latter case we also studied the impact of Eurosystem bond purchases, concluding that these had a beneficial impact on tail parameters, leaning against the risk of extremely adverse market outcomes while active.

Evidently, our model for time-varying tail parameters is focussed on capturing marginal features. In many applications it may also be of interest to study the time-varying nature of joint extremes; see e.g. Castro-Camilo et al. (2018), Escobar-Bach et al. (2018), and Mhalla et al. (2019). In terms of our first illustration, for example, one could wonder whether extremely negative log-returns for the S&P 500 and IBM stock were more dependent at certain points in time. We leave such research for future work; but see also Lucas et al. (2014), Patton and Oh (2018), and Hautsch and Herrera (2020) in this regard.

References

Blasques, F., S. J. Koopman, K. Lasak, and A. Lucas (2016). In-sample confidence bands and out-of-sample forecast bands for time-varying parameters in observation-driven models. *International Journal of Forecasting 32*, 875–887.

Blasques, F., S. J. Koopman, and A. Lucas (2015). Information theoretic optimality of observation driven time series models for continuous responses. *Biometrika 102*(2), 325–343.

Blasques, F., J. van Brummelen, S. J. Koopman, and A. Lucas (2022). Maximum likelihood estimation for Generalized Autoregressive Score models. *Journal of Econometrics*, (in press).

Boswijk, H. P., G. Cavaliere, I. Georgiev, and A. Rahbek (2021). Bootstrapping non-stationary stochastic volatility. *Journal of Econometrics 224*(1), 161–180.

Bougerol, P. (1993). Kalman filtering with random coefficients and contractions. *SIAM Journal on Control and Optimization 31*(4), 942–959.

Castro-Camilo, D., M. de Carvalho, and J. Wadsworth (2018). Time-varying extreme value dependence with application to leading European stock markets. *Annals of Applied Statistics 12*(1), 283–309.

Catania, L. and A. Luati (2019). Semiparametric modeling of multiple quantiles. *Available at SSRN 3494995*.

Chavez-Demoulin, V., P. Embrechts, and S. Sardy (2014). Extreme-quantile tracking for financial time series. *Journal of Econometrics 181*(1), 44–52.

Coles, S. (2001). *An introduction to statistical modeling of extreme values*. Springer Press, London.

Cox, D. R. (1981). Statistical analysis of time series: some recent developments. *Scandinavian Journal of Statistics 8*, 93–115.

Creal, D., S. J. Koopman, and A. Lucas (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics 28*(5), 777–795.

Creal, D., S. J. Koopman, A. Lucas, and M. Zamojski (2018). Generalized autoregressive method of moments.

Creal, D., B. Schwaab, S. J. Koopman, and A. Lucas (2014). An observation driven mixed measurement dynamic factor model with application to credit risk. *The Review of Economics and Statistics 96(5)*, 898915.

Davidson, A. C. and R. L. Smith (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Association, Series B 52*(3), 393–442.

Davidson, R. and J. G. MacKinnon (2004). *Econometric theory and methods*. Oxford University press.

de Haan, L. and C. Zhou (2021). Trends in extreme value indices. *Journal of the Americal Statistical Association 116*(535), 1265–1279.

ECB (2013). European Central Bank, Annual Report for 2012.

ECB (2014). The determinants of euro area sovereign bond yield spreads during the crisis. ECB Monthly Bulletin article, May 2014.

Einmahl, J., L. de Haan, and C. Zhou (2016). Statistics of heteroscedastic extremes. *Journal of the Royal Statistical Society, Series B 78*, 31–51.

Embrechts, P., C. Klüppelberg, and T. Mikosch (1997). *Modelling extremal events for insurance and finance*. Springer Verlag, Berlin.

Engle, R. F. and S. Manganelli (2004). CAViaR: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics 22*(4), 367–381.

Escobar-Bach, M., Y. Goegebeur, and A. Guillou (2018). Local robust estimation of the Pickands dependence function. *Annals of Statistics 46*(6A), 2806–2843.

Eser, F. and B. Schwaab (2016). Evaluating the impact of unconventional monetary policy measures: Empirical evidence from the ECB's Securities Markets Programme. *Journal of Financial Economics 119*(1), 147–167.

Francq, C. and J.-M. Zakoian (2019). *GARCH models: structure, statistical inference and financial applications*. John Wiley & Sons.

Galbraith, J. W. and S. Zernov (2004). Circuit breakers and the tail index of equity returns. *Journal of Financial Econometrics 2*(1), 109–129.

Ghysels, E., J. Idier, S. Manganelli, and O. Vergote (2017). A high frequency assessment of the ECB Securities Markets Programme. *Journal of European Economic Association 15(1)*, 218–243.

Harvey, A. C. (2013). *Dynamic models for volatility and heavy tails with applications to financial and economic time series*. Cambridge University Press.

Hautsch, N. and R. Herrera (2020). Multivariate dynamic intensity peaks-over-threshold models. *Journal of Applied Econometrics 35*, 248–272.

Hill, B. (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics 3*(5), 1163–1174.

Hoga, Y. (2017). Testing for changes in (Extreme) VaR. *Econometrics Journal 20*, 23–51.

Huisman, R., K. Koedijk, C. Kool, and F. Palm (2001). Tail-index estimates in small samples. *Journal of Business & Economic Statistics 19*(1), 208–216.

Joao, I. C., A. Lucas, J. Schaumburg, and B. Schwaab (2022). Dynamic clustering of multivariate panel data. *Journal of Econometrics, forthcoming*.

Koenker, R. (2005). *Quantile Regression*. Cambridge: Cambridge University Press.

Li, D., X. Zhang, K. Zhu, and S. Ling (2018). The ZD-GARCH model: A new way to study heteroscedasticity. *Journal of Econometrics 202*, 1–17.

Lin, C.-H. and T.-C. Kao (2018). Multiple structural changes in the tail behavior: evidence from stock market futures returns. *Nonlinear Analysis: Real World Applications 9*, 1702–1713.

Lucas, A., J. Schaumburg, and B. Schwaab (2019). Bank business models at zero interest rates. *Journal of Business & Economic Statistics 37*(3), 542–555.

Lucas, A., B. Schwaab, and X. Zhang (2014). Conditional euro area sovereign default risk. *Journal of Business and Economics Statistics 32*(2), 271–284.

Lucas, A. and X. Zhang (2016). Score driven exponentially weighted moving average and value-at-risk forecasting. *International Journal of Forecasting 32*, 293–302.

Manganelli, S. and R. F. Engle (2004). A comparison of value at risk models in finance. In G. Szego (Ed.), *Risk Measures for the 21st Century*. Wiley Finance.

Massacci, D. (2017). Tail risk dynamics in stock returns: Links to the macroeconomy and global markets connectedness. *Management Science 63*(9), 112–132.

McNeil, A. and R. Frey (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: An Extreme Value approach. *Journal of Empirical Finance 7*(3-4), 271–300.

McNeil, A. J., R. Frey, and P. Embrechts (2010). *Quantitative risk management: Concepts, techniques, and tools*. Princeton University press.

Mhalla, L., M. de Carvalho, and V. Chavez-Demoulin (2019). Regression-type models for extremal dependence. *Scandinavian Journal of Statistics 46*(4), 1141–1167.

Patton, A. (2006). Modelling asymmetric exchange rate dependence. *International Economic Review 47*(2), 527–556.

Patton, A. J. and D. H. Oh (2018). Time-varying systemic risk: evidence from a dynamic copula model of CDS spreads. *Journal of Business & Economic Statistics 36*(2), 181–195.

Patton, A. J., J. F. Ziegel, and R. Chen (2019). Dynamic semiparametric models for Expected Shortfall (and Value-at-Risk). *Journal of Econometrics 211*(2), 388–413.

Pooter, M. D., R. F. Martin, and S. Pruitt (2018). The liquidity effects of official bond market intervention. *Journal of Financial and Quantitative Analysis 53*(1), 243–268.

Quintos, C., Z. Fan, and P. C. Phillips (2001). Structural change tests in tail behaviour and the asian crisis. *The Review of Economic Studies 68*(3), 633–663.

Rocco, M. (2014). Extreme value theory in finance: A survey. *Journal of Economic Surveys 28*(1), 82–108.

Wagner, N. (2005). Autoregressive conditional tail behavior and results on government bond yield spreads. *International Review of Financial Analysis 14*(2), 247–261.

Wang, H. and C.-L. Tsai (2009). Tail Index Regression. *Journal of the American Statistical Association 104*(487), 1233–1240.

Werner, T. and C. Upper (2004). Time variation in the tail behavior of bund future returns. *Journal of Futures Markets 24*(4), 387–398.