Schwaab, Bernd; Zhang, Xin; Lucas, André

Working Paper

Modeling extreme events: time-varying extreme tail shape

Tinbergen Institute Discussion Paper, No. TI 2020-076/III

Provided in Cooperation with:
Tinbergen Institute, Amsterdam and Rotterdam

Suggested Citation: Schwaab, Bernd; Zhang, Xin; Lucas, André (2020) : Modeling extreme events: time-varying extreme tail shape, Tinbergen Institute Discussion Paper, No. TI 2020-076/III, Tinbergen Institute, Amsterdam and Rotterdam

This Version is available at:
https://hdl.handle.net/10419/229696

TI 2020-076/III
Tinbergen Institute Discussion Paper

Modeling extreme events: time-varying extreme tail shape

Bernd Schwaab[1]
Xin Zhang[2]
Andre Lucas[3]

[1] European Central Bank, Financial Research
[2] Sveriges Riksbank, Research Division
[3] Vrije Universiteit Amsterdam, Tinbergen Institute

Modeling extreme events: time-varying extreme tail shape*

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

[(a)] European Central Bank, Financial Research

[(b)] Sveriges Riksbank, Research Division

[(c)] Vrije Universiteit Amsterdam and Tinbergen Institute

Abstract

A dynamic semi-parametric framework is proposed to study time variation in tail fatness of sovereign bond yield changes during the 2010–2012 euro area sovereign debt crisis measured at a high (15-minute) frequency. 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 shape parameters. The score-driven updates used improve the expected Kullback-Leibler divergence between the model and the true data generating process on every step even if the GPD only fits approximately and the model is mis-sepcified, as will be the case in any finite sample. This is confirmed in simulations. Using the model, we find the ECB program had a beneficial impact on extreme upper tail quantiles, leaning against the risk of extremely adverse market outcomes while active.

Keywords: dynamic tail risk, observation-driven models, extreme value theory, European Central Bank (ECB), 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. An earlier version of this paper was circulated as "Tail risk in government bond markets and ECB unconventional policies." Schwaab thanks ECB DG-M for insight and access to high-frequency data on SMP bond purchases. 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 novel semi-parametric framework to study time variation in tail fatness for long univariate time series, applied to high-frequency government bond returns during times of unconventional central bank policies. The new method builds on ideas from Extreme Value Theory (EVT) by using a conditional Generalized Pareto Distribution (GPD) to approximate the tail beyond a given threshold, and endowing this conditional GPD distribution with time-varying parameters. The GPD is an appropriate tail approximation for most heavy-tailed densities used in econometrics and actuarial sciences; see, for example, Davidson and Smith (1990), Embrechts et al. (1997), and McNeil et al. (2010, Chapter 7). 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. Our framework allows us to study the time-variation in tail index parameters associated with time series observations from a wide class of heavy-tailed distributions; see Rocco (2014) for a recent survey of extreme value theory (EVT) methods. We discuss the handling of non-tail time series observations, inference on deterministic and time-varying parameters, and ways to relate time-varying parameters to observed covariates. In this context we also study the effect of time-varying pre-filtering methods possibly applied to the data before the dynamic GPD model is fitted.

In our model, the time-varying tail shape and tail scale parameters of the GPD are driven by the score of the local (time t) objective function; see e.g. Creal et al. (2013) and Harvey (2013). In this approach, the time-varying parameters are perfectly predictable one step ahead. This makes the model observation-driven in the terminology of Cox (1981). The log-likelihood is known in closed form, facilitating parameter estimation and inference via standard maximum likelihood methods. Simulation evidence reveals that our model and estimation approach is able to recover the time-varying tail shape and tail scale parameters sufficiently accurately, as well as EVT-based market risk measures such as Value-at-Risk (VaR) and Expected Shortfall (ES) at high confidence levels (say, 99%). 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 result of the GPD can only hold approximately given the choice of a finite exceedance threshold in any particular sample.

We apply our modeling framework to study the location, scale, and upper tail impact of bond purchases undertaken by the Eurosystem – the European Central Bank (ECB) and its 17 national central banks at the time – during the euro area sovereign debt crisis between 2010 and 2012. We focus on bond purchases within the Eurosystem's Securities Markets Programme (SMP), which targeted sovereign bonds of five euro area countries: Greece, Ireland, Italy, Portugal, and Spain. Based on high-frequency data for five-year benchmark bonds, and explicitly accounting for time-variation in fat tails, we find that purchases lowered the conditional location (mean) of future bond yields by up to -2.9 basis points (bps) per €1 bn of purchases. The impact estimates for the two largest SMP countries, Italy and Spain, are -1.5 bps and -2.6 bps per €1 bn of purchases, respectively. These impact estimates are approximately in line with earlier estimates; see e.g. Eser and Schwaab (2016), Ghysels et al. (2017), and Pooter et al. (2018).

In addition, we find that SMP purchases had a beneficial impact on the extreme upper tail quantiles of yield changes. This suggests that central bank bond purchases lean against the risk of extremely adverse market outcomes while they are active. The beneficial impact is mostly explained by moving the center of the predictive distribution to the left and narrowing it. Beneficial secondary effects come about via the SMP's effect on tail shape and tail scale for large economies such as Spain and Italy. The impact of purchases on tail quantiles is larger than their impact on the conditional location (mean). We estimate that the 97.5% VaR was reduced by 3.8, 6.0, 5.9, 2.1, and 6.9 bps per €1 bn Eurosystem intervention in Spanish, Greek, Irish, Italian, and Portuguese five-year benchmark bonds, respectively. The impact grows with the extremeness of the VaR. We estimate that the 99.5% VaR was reduced, respectively, by 5.1, 10.1, 12.5, 2.9, and 15.4 bps per €1 bn of Eurosystem purchases in the above bonds. The tail impact of the SMP purchases is economically relevant because extreme tail risks alone can force dealer banks and market makers to retreat from supplying liquidity to important segments of the sovereign bond market, particularly when their own

VaR constraints are binding; see Vayanos and Vila (2009) and Adrian and Shin (2010). In turn, malfunctioning sovereign bond markets can impair the transmission of the common monetary policy to all parts of the euro area. Pelizzon et al. (2016) provide evidence that market makers withdrew from trading Italian debt securities in 2011.

Our paper is related to at least two strands of literature. First, several papers propose methodology to study time variation in the tail index. Quintos et al. (2001), Einmahl et al. (2016), and Hoga (2017) 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 (2020) propose a non-parametric approach to estimating the extreme value index locally. Our paper adds to this strand of conditional EVT literature by proposing a model 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 treshold 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.. The resulting dynamics in our model are, as a result, more robust, particularly for the ES.

A second strand of literature assesses the impact of central bank asset purchases on bond yields and yield volatility. For example, Ghysels et al. (2017) study the yield impact of SMP bond purchases by considering bond yields and purchases at 15-minute intervals. In this way they mitigate a bias that unobserved factors could have introduced. The authors estimate that SMP interventions had an impact on the conditional mean of 10-year maturity bonds of between -0.2 and -4.2 bps per €1 bn of purchases. Eser and Schwaab (2016) study yield impact based on daily data. In their framework, identification is based on a panel model

that exploits the cross-sectional dimension of the data. They find that, in addition to large announcement effects, purchases of 1/1000 of the respective outstanding debt had an impact of approximately -3 bps at the five-year maturity. Pooter et al. (2018) use the published weekly data of aggregate SMP purchases to test for an impact on country-specific sovereign bond liquidity premia. The authors find an average impact of -2.3 bps for purchases of 1/1000 of the outstanding debt. Our paper adds to the growing literature on impact identification of central bank asset purchase programs by developing methodology to study their impact on the extreme tail of the distribution.

Whereas de Haan and Zhou (2020) 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 in that we specify both parameters as functions of their respective scores, and adopt a non-diagonal scaling function. We cover inference on both deterministic and time-varying parameters, explain how to introduce additional conditioning variables, and provide Monte Carlo evidence. Owing to a novel autoregressive specification of the EVT threshold following Patton et al. (2019), our model can be fitted to both prefiltered and non-prefiltered time series data.

We proceed as follows. Section 2 presents our statistical model. Section 3 discusses our simulation results. Section 4 studies the tail impact of Eurosystem asset purchases. Section 5 concludes. A Web Appendix derives the score and scaling function for the tail shape model and provides further technical and empirical results.

2 Statistical model

2.1 Time-varying tail shape and scale

This section introduces our model with time-varying tail shape and tail scale for a univariate time series y_t, $t = 1, \ldots, T$, where T denotes the number of observations. We assume

$$y_t = \mu_t + \sigma_t \, \varepsilon_t, \tag{1}$$

where $g(\varepsilon_t \mid \mathcal{F}_{t-1})$ is the conditional probability density function (pdf) of ε_t, μ_t and σ_t are the conditional location and scale of y_t, and $\mathcal{F}_{t-1} = \{y_{t-1}, y_{t-2}, \ldots, y_1\}$ denotes the information set. The parameters μ_t and σ_t can take on many forms ranging from constant values to specifications with autoregressive and conditional volatility dynamics. Key, however, is that these parameters are typically mainly used to describe well the *center* of the distribution. In this paper, by contrast, we concentrate on the *tail* of the distribution using a dynamic extension of arguments from extreme value theory, similar to Patton's (2006) extension of copula theory to the dynamic, observation driven setting.

We assume the conditional pdf $g(\varepsilon_t \mid \mathcal{F}_{t-1})$ has heavy tails with time-varying tail index $\alpha_t > 0$. A prime example is the univariate Student's t distribution with $\nu_t = \alpha_t$ degrees of freedom. Other examples include the Pareto, inverse gamma, loggamma, log-logistic, F, Fréchet, and Burr distribution with one or more time-varying shape parameters. Rather, however, than modeling the (dynamic) tail shape by an arbitrarily chosen parametric family of distributions, we appeal to well-known results from the extreme value theory (EVT) literature. From EVT, we know that the conditional cumulative distribution function (cdf) $G(\varepsilon_t \mid \mathcal{F}_{t-1})$ of ε_t can under very general conditions be approximated by $G(\varepsilon_t \mid \mathcal{F}_{t-1}) = G(\tau \mid \mathcal{F}_{t-1}) + (1 - G(\tau \mid \mathcal{F}_{t-1}))P(x_t; \delta_t, \xi_t)$ with $x_t = \varepsilon_t - \tau$ for sufficiently high thresholds $\tau \in \mathbb{R}_+$, or more precisely,

$$\begin{aligned} &\lim_{\tau\to\infty} \sup_{e_t \geq \tau} \left| \mathbb{P}\left[\varepsilon_t \leq e_t + \tau \mid \varepsilon_t > \tau,\ \mathcal{F}_{t-1}\right] - P_{\xi_t,\delta_t}(e_t - \tau) \right| \\ &\qquad = \lim_{\tau\to\infty} \sup_{e_t \geq \tau} \left| \frac{G(e_t + \tau \mid \mathcal{F}_{t-1}) - G(e_t \mid \mathcal{F}_{t-1})}{1 - G(e_t \mid \mathcal{F}_{t-1})} - P_{\xi_t,\delta_t}(e_t - \tau) \right| = 0, \end{aligned} \tag{2}$$

for parameters $\xi_t = \alpha_t^{-1}$ and δ_t, both possibly depending on τ. Here, $P(x_t; \delta_t, \xi_t)$ denotes the cdf of the Generalized Pareto Distribution (GPD), with cdf and pdf given by

$$P(x_t; \delta_t, \xi_t) = 1 - \left(1 + \xi_t \frac{x_t}{\delta_t}\right)^{-\xi_t^{-1}}, \qquad p(x_t; \delta_t, \xi_t) = \delta_t^{-1} \cdot \left(1 + \xi_t \frac{x_t}{\delta_t}\right)^{-\xi_t^{-1}-1}, \tag{3}$$

respectively (see, for example, McNeil et al., 2010). The quantity $x_t = \varepsilon_t - \tau > 0$ is the so-called peak-over-threshold (POT), or exceedance, of heavy-tailed data ε_t over a pre-

determined threshold τ, and $\delta_t > 0$ and $\xi_t > 0$ are the 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 (2) 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 $\mathbb{P}\left[\varepsilon_t \leq e_t + \tau \mid \varepsilon_t > \tau,\ \mathcal{F}_{t-1}\right]$ is improved every time for sufficiently small steps, even if the GPD model is misspecified; see Blasques et al. (2015). The choice of the threshold τ is subject to a well-known bias-efficiency tradeoff; see, for instance, McNeil and Frey (2000). In theory, the GPD tail approximation only becomes exact for $\tau \to +\infty$. 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 from the literature are the 90%, 95%, and 99% empirical quantiles of ε_t; see Chavez-Demoulin et al. (2014). We return to the choice, and modeling, of the threshold further below.

A key step in (3) is that we use the conditional probabilities based on the information set $\mathcal{F}_{t-1}$. As a result, the tail shape parameters become time-varying. To capture this time-variation, we model $(\xi_t, \delta_t)'$ using the score-driven (GAS) dynamics introduced by Creal et al. (2013) and Harvey (2013). In our time series setting, that 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 so-called GAS(p, q)-dynamics as

$$f_{t+1} = \omega + \sum_{i=0}^{p-1} A_i s_{t-i} + \sum_{j=0}^{q-1} B_j f_{t-j}, \tag{4}$$
$$s_t = \mathcal{S}_t \nabla_t, \qquad \nabla_t = \partial \ln p(x_t \mid \mathcal{F}_{t-1}; f_t, \theta)/\partial f_t,$$

where vector $\omega = (\omega^\xi, \omega^\delta)' = \omega(\theta)$ and matrices $A_i = A_i(\theta)$ and $B_j = B_j(\theta)$ depend on the deterministic 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 (4) updates f_t at every time point in time via a scaled steepest ascent step to improve the fit to the GPD. The score of (3) required in (4) is given by

$$\nabla_t = \begin{bmatrix} \xi_t^{-1} \cdot \log\left(1 + \xi_t \, \delta_t^{-1} x_t\right) - \left(1 + \xi_t^{-1}\right) \dfrac{\xi_t x_t}{\delta_t + \xi_t x_t} \\ \dfrac{x_t - \delta_t}{\delta_t + \xi_t x_t} \end{bmatrix}, \tag{5}$$

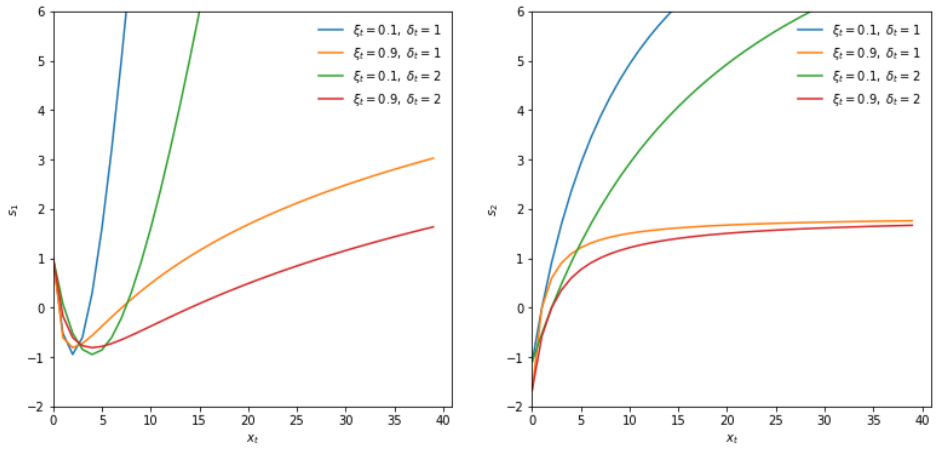
where $\log(\cdot)$ denotes the natural logarithm; see Appendix A.1 for a derivation. We take A_i and B_j as diagonal matrices.

Following Creal et al. (2014) we select the square-root inverse conditional Fisher information of the conditional observation density to scale (5), i.e., $\mathcal{S}_t = L_t'$, with L_t the choleski decomposition of the inverse conditional Fisher information matrix $\mathcal{I}_t = (L_t L_t')^{-1} = \mathbb{E}[\nabla_t \nabla_t' \mid \mathcal{F}_{t-1}; f_t, \theta] = \mathbb{E}[-\partial \nabla_t / \partial f_t' \mid \mathcal{F}_{t-1}; f_t, \theta]$, such that the conditional variance of s_t is equal to the unit matrix. For the GPD, we have

$$L_t = \begin{bmatrix} 1 + \xi_t^{-1} & 0 \\ -1 & \sqrt{1 + 2\xi_t} \end{bmatrix}, \tag{6}$$

Figure 1: News impact curves

The first element (left panel) and second element (right panel) of s_t in (7) is plotted against x_t for different values of ξ_t and δ_t.



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

$$s_t = L_t' \nabla_t = \begin{bmatrix} \xi_t^{-2}(1+\xi_t) \cdot \log\left(1 + \xi_t\, \delta_t^{-1} x_t\right) + \dfrac{\delta_t - (\xi_t + 3 + \xi_t^{-1}) \cdot x_t}{\delta_t + \xi_t x_t} \\ \sqrt{1+2\xi_t}\, \dfrac{x_t - \delta_t}{\delta_t + \xi_t x_t} \end{bmatrix}. \tag{7}$$

Though the scaled score in (7) seems unstable at first sight for ξ_t near zero, the expression actually has a finite 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 (7) 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_t imply that f_t is adjusted upwards. For high realization of x_t the adjustments are greatest when the current tail shape and tail scale are low. The function shapes become increasingly concave as $x \to \infty$ in line with robust updates of the time-varying parameters. This 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 to the VaR exceedance.This can result in noisy or unstable ES estimates. Using the GPD shape as emanating from EVT, Figure 1 shows that ξ_t and δ_t react more modestly to large POT observations. This makes sense, as we expect such 'outliers' to occur more often

for higher values of ξ_t. For extremely high $\xi_t \geq 1$, the ES even ceases to exist. We also note that small realizations of x_t imply downward adjustments of both elements of f_t, up to the point where x_t becomes very small. In that case f_t^{ξ} is adjusted upward, as observations near the center of a fat-tailed distribution signal increased peakedness (=leptokurtosis); see also Lucas and Zhang (2016). The score-driven steps in (7) can thus result in more stable and interpretable parameter paths due to the concavity of the news impact curves.

When there is no tail observation, i.e. $x_t = \varepsilon_t - \tau \leq 0$, then the new observation carries no information about ξ_t and δ_t, and the score is zero as a result. Long consecutive stretches of zero scores can lead to erratic paths for f_t and thus (ξ_t, δ_t). In addition, such stretches of zero scores can be problematic for inference on θ; see Blasques et al. (2018). Both issues can be addressed by taking into account lagged values of the scaled score via the exponentially-weighted moving average specification

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

where $\tilde{s}_t = (1-\lambda)s_t + \lambda\tilde{s}_{t-1}$, $\lambda \in (0,1)$ is an additional parameter to be estimated, and s_t is given by (7). While s_t is most often zero, $\tilde{s}_t$ is not. Clearly, (4) is a special case of (8) for $\lambda \to 0$. Specification (8) leads to a GAS(1,2) specification for f_t,

$$\big(\mathrm{I}_2 - B\,L\big)(1-\lambda)^{-1}\big(1-\lambda L\big)\,f_{t+1} = \omega + As_t,$$

where L is the lag operator. To see this, first rewrite (8) to $(\mathrm{I}_2 - B\,L)f_{t+1} = \omega + A\tilde{s}_t$, and then multiply both sides by $(1-\lambda\,L)/(1-\lambda)$, using $(1-\lambda\,L)\tilde{s}_t = (1-\lambda)s_t$. The smoothing approach in (8) is similar to the approach in Patton (2006) that uses up to ten lags of the driver (in our case the score) to smoothen the dynamics of the time-varying parameter.

The transition equation for f_t can be extended further if additional conditioning variables are available. For example, central bank sovereign bond purchases may help explain the time-variation in the tail shape and tail scale parameters associated with changes in sovereign bond yields; see Section 4. Such additional variables can be taken into account in a straightforward

way via the modified transition equation,

$$f_{t+1} = \omega + A\tilde{s}_t + Bf_t + C \cdot z_t, \tag{9}$$

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

We consider three different ways to set the relevant thresholds. The thresholds can be either time-invariant (τ) or time-varying (τ_t). The construction of the thresholds can be important in practice because τ_t determines whether an observation lies in the tail, and, if so, what is the magnitude of the exceedance $x_t = \varepsilon_t - \tau_t > 0$. The κ-quantile $Q^{\kappa}_{1:T}(\{\varepsilon_1, \ldots, \varepsilon_T\})$ associated with the full sample is an obvious first candidate, $\kappa \in (0, 1)$. In this case, $\tau = Q^{\kappa}_{1:T}(\{\varepsilon_1, \ldots, \varepsilon_T\})$ is time-invariant. Alternatively, we can compute the quantile recursively up to time t and set $\tau_t = Q^{\kappa}_{1:t}(\{\varepsilon_1, \ldots, \varepsilon_t\})$, such that τ_t is time-varying. Finally, we consider a dynamic specification as suggested by Patton et al. (2019), according to which

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

where a^{τ} is a parameter to be estimated, and $\tau_1 = Q^{\kappa}_{1:T}$ is used to initialize the process. The recursive specification (10) is a martingale since $\mathbb{E}[1\{\varepsilon_t > \tau_t\} \mid \mathcal{F}_{t-1}, \theta] = (1 - \kappa)$. The threshold τ_t can now respond to changes in the underlying location, scale, and higher-order moments of ε_t in a straightforward way. This is particularly relevant if the data y_t is not pre-filtered based on an appropriate location–scale model in a first step, for instance if we set $\mu_t = 0$ and $\sigma_t = 1$ in (1), thus modeling the conditional extreme tail shape of y_t directly.

We close this section with a brief comment on parameter interpretability. The tail shape parameter ξ_t can always be interpretated as observation y_t's contemporaneous inverse tail index α_t^{-1}. By contrast, the estimated scale parameter δ_t need not have a straightforward interpretation in terms of y_t's conditional variance. For example, assume that y_t were GPD distributed with time-varying tail shape parameter α_t^{-1} and scale σ_t. We can then show that the derived POT x_t also has an exact GPD-distribution, with the *same* tail shape parameter

$\xi_t = \alpha_t^{-1}$, but a *different* scale parameter $\delta_{t,\tau} = \sigma_t + \alpha_t^{-1} \cdot \tau$; see Web Appendix B.1 for details. As a result, $\delta_{t,\tau}$ increases with the threshold, varies positively with the tail shape parameter ξ_t, and, importantly, should not be expected to provide a consistent estimate of σ_t. A similar result can be derived if the time series data y_t were Student's t-distributed with scale σ_t and tail index α_t; see Web Appendix B.2. We return to this issue in our simulation Section 3, where we consider pseudo-true values of both parameters to benchmark how well the model can estimates these.

2.2 Confidence bands for tail shape and scale

Confidence (or standard error) bands allow us to visualize the impact of estimation uncertainty associated with the maximum likelihood estimate $\hat{\theta}$ on the filtered estimates $\hat{f}_t(\hat{\theta})$, and, by extension, also $(\hat{\xi}_t, \hat{\delta}_t)' = \exp(\hat{f}_t(\hat{\theta}))$. Quantifying the uncertainty about these 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 $V_t = \text{Var}(\hat{f}_t)$. 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 C. In our empirical application below, however, the linear approximations are typically insufficient to capture the uncertainty in the highly non-linear dynamics for some countries. 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 $\text{N}(\hat{\theta}, \hat{W})$, where $\hat{W}$ is the estimated covariance matrix of $\hat{\theta}$, e.g., a sandwich covariance matrix estimator. If the finite-sample distribution of $\hat{\theta}$ were known, that could be used instead. For each draw $\hat{\theta}^s$ we now run the filter for f_t from $t = 1$ to $t = T$, thus obtaining S 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 for f_t. The simulated bands can now be obtained directly by calculating the appropriate percentiles

for each t over the S draws of the paths $\hat{f}_t^s$ for $s = 1, \ldots, S$.

2.3 Parameter estimation

Parameter estimates can be obtained in a standard way by numerically maximizing the log-likelihood function. Observation-driven time series models such as (3) – (10) are attractive because the log-likelihood is known in closed form. For a given set of time series observations $x_1, \ldots, x_T$, the vector of unknown parameters θ can be estimated by maximizing the log-likelihood function with respect to θ. The average log-likelihood function is given by

$$\begin{aligned}\mathcal{L}(\theta|\mathcal{F}_T) &= (T^*)^{-1} \sum_{t=1}^{T} 1\{x_t > 0\} \cdot \ln p(x_t; \delta_t, \xi_t) \\ &= (T^*)^{-1} \sum_{t=1}^{T} 1\{x_t > 0\} \cdot \left[-\ln(\delta_t) - \left(1 + \frac{1}{\xi_t}\right) \ln\left(1 + \xi_t \frac{x_t}{\delta_t}\right) \right], \end{aligned} \tag{11}$$

where $T^* = \sum_{t=1}^{T} 1\{x_t > 0\}$ is the number of POT values in the sample. Maximization of (11) can be carried out using a conveniently chosen quasi-Newton optimization method.

Blasques et al. (2020) provide conditions under which the maximum likelihood estimator of θ is consistent and asymptotically normally distributed within the class of correctly-specified 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 in terms of $\ln p(x_t; f_t)$. This is reassuring since the GPD is never exact for any finite value of $\tau < \infty$. 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 $W = \mathrm{Var}(\hat{\theta})$ then takes its usual sandwich form; see e.g. Davidson and MacKinnon (2004, Ch. 10) and Blasques et al. (2020).

The autoregressive parameter a^τ in (10) cannot be estimated using (11). 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

$$\min_{\{a^\tau\}} T^{-1} \sum_{t=1}^{T} \rho_\kappa \left(\epsilon_t - \tau_t\right) \iff \min_{\{a^\tau\}} T^{-1} \sum_{t=1}^{T} (\epsilon_t - \tau_t) \left(\kappa - 1\{\epsilon_t < \tau_t\}\right)$$
$$\iff \max_{\{a^\tau\}} T^{-1} \sum_{t=1}^{T} (\epsilon_t - \tau_t) \left((1-\kappa) - 1\{\epsilon_t > \tau_t\}\right), \tag{12}$$

where $\rho_\kappa(u_t) = u_t \left(\kappa - 1\{u_t < 0\}\right)$, and τ_t evolves as in (10). 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 (12) before maximizing (11).[1]

2.4 A conditional location–scale–df model

This section introduces a score-driven location–scale–df model that can be used to pre-filter univariate time series data y_t that is arbitrarily fat-tailed, where df denotes the degrees of freedom. The model modifies the setting of Lucas and Zhang (2016) with a Student's t distribution with time-varying volatility and degrees of freedom parameters to a setting that also allows for a time-varying location μ_t parameter and to more extreme tails ($\nu_t < 2$), in which case the volatility no longer exists, but a time varying scale parameter $\sigma_t > 0$ does exist. Since all parameters are time-varying, using this model minimizes the risk of mistaking time-variation in the center of the distribution for time-variation in the tail, and vice versa. The restriction $\nu_t > 0$ aligns closely with the assumption $\alpha_t > 0$ and $\xi_t > 0$ in Section 2.1. For the purposes of pre-filtering, in this section y_t is assumed to be generated by

$$y_t \sim \mathrm{t}(y_t; \mu_t, \sigma_t, \nu_t), \tag{13}$$

[1] Numerical gradient-based optimizers, such as e.g. MaxBFGS, may only indicate weak convergence at the optimum of (12). This is due to the piecewise linear objective function. The optimizer at hand may not be suited for such a function, and will end up in a kink. This is not a problem, assuming we are not interested in standard errors for a^τ. Alternatively the interior point algorithm of Koenker and Park (1996) could be used.

where $\mu_t = \mathbb{E}[y_t \mid \mathcal{F}_{t-1}]$ if $\nu_t > 1$, and $\sqrt{\nu_t/(\nu_t-2)}\,\sigma_t$ is the conditional volatility of y_t if $\nu_t > 2$. All time-varying parameters are modeled in a score-driven way as

$$\mu_{t+1} = \omega^\mu + a^\mu s_t^\mu + b^\mu \mu_t + c^\mu z_t + d^\mu y_t, \tag{14}$$

$$\ln \sigma_{t+1} = \omega^\sigma + a^\sigma s_t^\sigma + b^\sigma \ln \sigma_t + c^\sigma z_t + d^\sigma 1\{y_t > \mu_t\} s_t^{\text{Lev}}, \tag{15}$$

$$\nu_{t+1} = \omega^\nu + a^\nu s_t^\nu + b^\nu \nu_t + c^\nu z_t, \tag{16}$$

where $\omega^{(\cdot)}$, $a^{(\cdot)}$, $b^{(\cdot)}$, $c^{(\cdot)}$, and $d^{(\cdot)}$ are scalar parameters to be estimated, and z_t is a vector of additional conditioning variables which may be available. The required scaled scores are

$$s_t^\mu = \frac{(\nu_t+3)(y_t-\mu_t)}{\nu_t + \sigma_t^{-2}(y_t-\mu_t)^2}, \tag{17}$$

$$s_t^\sigma = \frac{\nu_t+3}{2\nu_t} \cdot \left(\frac{(\nu_t+1)(y_t-\mu_t)^2}{\nu_t\sigma_t^2 + (y_t-\mu_t)^2} - 1 \right), \tag{18}$$

$$s_t^\nu = \frac{1}{2}\left[\frac{\nu_t}{4}\gamma''\left(\frac{\nu_t+1}{2}\right) - \frac{\nu_t}{4}\gamma''\left(\frac{\nu_t}{2}\right) + \frac{1}{2}\frac{\nu_t+5}{(\nu_t+1)(\nu_t+3)} \right]^{-1} \left[\frac{1}{\nu_t} + \gamma'\left(\frac{\nu_t}{2}\right) - \gamma'\left(\frac{\nu_t+1}{2}\right) + \ln\left(1 + \frac{(y_t-\mu_t)^2}{\nu_t\sigma_t^2}\right) - \frac{\nu_t+1}{\nu_t}\frac{(y_t-\mu_t)^2}{\nu_t\sigma_t^2 + (y_t-\mu_t)^2} \right], \tag{19}$$

where the functions $\gamma'(x)$ and $\gamma''(x)$ are the first and second derivatives of the log-gamma function. We refer to Web Appendix D for a derivation of (17) – (19).

The "leverage" term $d^\sigma \cdot 1\{y_t > \mu_t\} s_t^{\text{Lev}}$ in (15) allows $\ln \sigma_{t+1}$ to be higher (or lower, depending on the sign of d^σ) when y_t is above its location μ_t. The term $s_t^{\text{Lev}} = s_t^\sigma(y_t) - s_t^\sigma(\mu_t)$ is constructed such that the score is continuous at μ_t. Leverage specifications are often found to be valuable in many empirical applications; see e.g. Engle and Patton (2001). The deterministic parameters in (14) – (16) can be estimated by (quasi-)maximum likelihood methods in line with the discussion in Section 2.3.

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 (2) – (3) yields

useful closed-form estimators of the VaR and ES for high upper quantiles $\gamma > G(\tau \mid \mathcal{F}_{t-1})$; 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

$$\begin{aligned}
\text{VaR}^{\gamma}(\epsilon_t \mid \mathcal{F}_{t-1}, \theta) &= \tau_t + \delta_t \xi_t^{-1} \left[\left(\frac{1-\gamma}{t^*/t} \right)^{-\xi_t} - 1 \right], \\
\text{VaR}^{\gamma}(y_t \mid \mathcal{F}_{t-1}, \theta) &= \mu_t + \sigma_t \text{VaR}^{\gamma}(\epsilon_t \mid \mathcal{F}_{t-1}, \theta),
\end{aligned} \tag{20}$$

where μ_t and σ_t are defined below (1), and t^* is the number of observations of $x_t > 0$ 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 $\kappa_t = G(\tau_t \mid \mathcal{F}_{t-1})$.

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 expressions are

$$\begin{aligned}
\text{ES}^{\gamma}(\epsilon_t \mid \mathcal{F}_{t-1}, \theta) &= \frac{1}{1-\gamma} \int_{\gamma}^{1} \text{VaR}^{\tilde{\gamma}}(\epsilon_t \mid \mathcal{F}_{t-1}, \theta) \mathrm{d}\tilde{\gamma} \\
&= \frac{\text{VaR}^{\gamma}(\epsilon_t \mid \mathcal{F}_{t-1}, \theta)}{1-\xi_t} + \frac{\delta_t - \xi_t \tau_t}{1-\xi_t}, \\
\text{ES}^{\gamma}(y_t \mid \mathcal{F}_{t-1}, \theta) &= \mu_t + \sigma_t \text{ES}^{\gamma}(\epsilon_t \mid \mathcal{F}_{t-1}, \theta);
\end{aligned} \tag{21}$$

see Web Appendix E for a derivation of (20) – (21). The $\text{ES}^{\gamma}(y_t|\cdot)$ is strictly higher than the $\text{VaR}^{\gamma}(y_t|\cdot)$ at the same confidence level, as it "looks further into the tail." It can be shown that the ratio $\text{ES}^{\gamma}(y_t|\cdot)/\text{VaR}^{\gamma}(y_t|\cdot)$ 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 μ_t, σ_t, ξ_t and δ_t into (20) and (21).

For later reference, the sensitivity of $\text{VaR}^{\gamma}(y_t)$ to bond purchases z_{t-1} is given by

$$\frac{\mathrm{d}\text{VaR}^{\gamma}(y_t)}{\mathrm{d}z_{t-1}} = \frac{\partial \text{VaR}}{\partial \mu_t} \frac{\mathrm{d}\mu_t}{\mathrm{d}z_{t-1}} + \frac{\partial \text{VaR}}{\partial \sigma_t} \frac{\mathrm{d}\sigma_t}{\mathrm{d}\ln\sigma_t} \frac{\mathrm{d}\ln\sigma_t}{\mathrm{d}z_{t-1}} + \frac{\partial \text{VaR}}{\partial \delta_t} \frac{\mathrm{d}\delta_t}{\mathrm{d}f_t^{\delta}} \frac{\mathrm{d}f_t^{\delta}}{\mathrm{d}z_{t-1}} + \frac{\partial \text{VaR}}{\partial \xi_t} \frac{\mathrm{d}\xi_t}{\mathrm{d}f_t^{\xi}} \frac{\mathrm{d}f_t^{\xi}}{\mathrm{d}z_{t-1}}.$$

The expression is intuitive: extreme upper quantiles can change if bond purchases z_{t-1} affect

the conditional location μ_t, the conditional scale σ_t, the tail scale δ_t, or the tail shape ξ_t. The derivative is given by

$$\begin{aligned}\frac{\mathrm{dVaR}^{\gamma}(y_t)}{\mathrm{d}z_{t-1}} =& c^{\mu} + \sigma_t \mathrm{VaR}^{\gamma}(\epsilon_t) c^{\sigma} + \sigma_t \left(\mathrm{VaR}^{\gamma}(\epsilon_t) - \tau_t\right) c^{\delta} \\ & - \sigma_t \left\{ \mathrm{VaR}^{\gamma}(\epsilon_t) - \tau_t + \delta_t \left(\frac{1-\gamma}{t^*/t}\right)^{-\xi_t} \ln\left(\frac{1-\gamma}{t^*/t}\right) \right\} c^{\xi}, \end{aligned} \tag{22}$$

where μ_t and σ_t are given by (14) and (15), f_t^{ξ} and f_t^{δ} are given by (9) with $C = (c^{\delta}, c^{\xi})'$.

3 Simulation study

This section studies the question whether our score-driven modeling approach can reliably recover the time series variation in tail shape and tail scale in a variety of potentially challenging settings. In addition, we are interested in how to best choose the thresholds τ_t, as well as the accuracy of EVT-based market risk measures when used in combination with our modeling approach.

3.1 Simulation design

Our simulation design considers $D = 2$ different densities (GPD and t), $P = 4$ different parameter paths for tail shape and tail 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 $x_t > 0$. The time series dimension T is chosen to resemble that of the empirical data considered in Section 4.

GPD and t-densities: We first simulate y_t from a GPD distribution with time-varying tail shape α_t^{-1} and tail scale σ_t, $y_t \sim \mathrm{GPD}(\alpha_t^{-1}, \sigma_t)$. We then consider a Student's t distribution with time-varying scale σ_t and degrees of freedom α_t, $y_t \sim \mathrm{t}(0, \sigma_t, \alpha_t)$. POT values x_t are obtained as $x_t = y_t - \tau_t$.

Parameter paths: We consider four different paths for the tail shape α_t^{-1} and tail scale σ_t parameters. For both GPD and 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: $\alpha_t^{-1} = 0.5 + 0.3\sin(4\pi t/T)$, $\sigma_t = 1 + 0.5\sin(4\pi t/T)$.

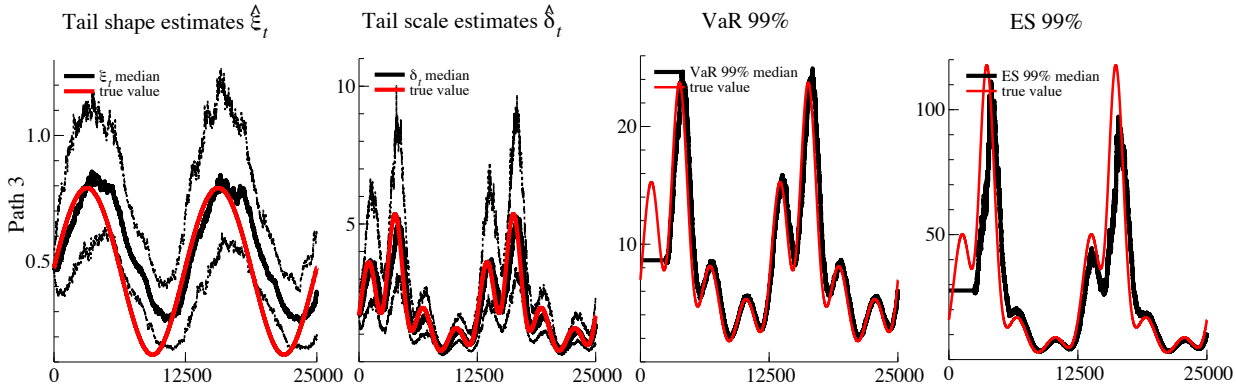
Path (1) considers the special case of time-invariant tail shape and scale parameters. Naturally, we would want our dynamic framework to cover constant parameters as a special case. Path (2) allows the tail shape to vary considerably between 0.2 and 0.8, while keeping the scale (volatility) of the data constant. This parameter path corresponds to the empirical practice of working with volatility pre-filtered data. Path (3) stipulates that both parameters vary over time. Finally, Path (4) considers the case of synchronized variation in both parameters. This 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.

Different thresholds: We consider three 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. $\tau_t = Q_{1:t}^{0.95}(\{\varepsilon_1, \ldots, \varepsilon_t\})$. Finally, we use the recursive specification (10), with a^τ fixed at 0.25, and initialized at $\tau_1 = Q_{1:T}^{0.95}$.

Evaluation metrics: Our main metric for evaluating model performance is the root mean squared error RMSE $= \frac{1}{S}\sum_{s=1}^{S}\sqrt{\frac{1}{T}\sum_{t=1}^{T}(\hat{\xi}_{st} - \bar{\xi}_{st})^2}$, where $\hat{\xi}_{st}$ is the estimated tail shape parameter in simulation s, $\bar{\xi}_{st}$ 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, RMSE $= \frac{1}{S}\sum_{s=1}^{S}\sqrt{\frac{1}{T}\sum_{t=1}^{T}(\hat{\delta}_{st} - \bar{\delta}_{st})^2}$, where $\bar{\delta}_{st}$ denotes the pseudo-true value of δ_{st}. The pseudo-true

Figure 2: Simulation results: a representative 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 (10) in conjunction with the objective function (12).



values $\bar{\xi}_{st}$ and $\bar{\delta}_{st}$ 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. We note that particularly the GPD 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 (2).

3.2 Simulation results

Table 1 presents root mean squared error (RMSE) statistics for tail shape $\hat{\xi}_{s,t}$, tail scale $\hat{\delta}_{s,t}$, and Value-at-Risk $\widehat{\text{VaR}}_{s,t}$ estimates. Figures F.1 and F.2 in Web Appendix F.1 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. Figure 2 is a representative example of the simulation outcomes when y_t is generated by a Student's t distribution.

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 Figures F.1 and F.2 tend to be close to their (pseudo-)true values. Particularly the sometimes highly non-linear patters

Table 1: Simulation RMSE results

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 *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 (12) with $a^\tau = 0.25$, respectively. 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. For VaR, model performance is measured in relative terms as RMSE rescaled by the squared VaR_t.

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 $\hat{\xi}_{s,t}$			
(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 $\hat{\delta}_{s,t}$			
(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)
			RMSE $\widehat{\text{VaR}}_{s,t}$			
(1)	0.001	0.003	0.016	0.124	0.124	0.149
	(0.001)	(0.002)	(0.003)	(0.001)	(0.001)	(0.002)
(2)	0.924	0.987	0.964	0.249	0.243	0.257
	(0.027)	(0.032)	(0.031)	(0.003)	(0.003)	(0.003)
(3)	1.063	1.304	1.209	0.322	0.344	0.349
	(0.025)	(0.041)	(0.033)	(0.004)	(0.005)	(0.004)
(4)	1.020	1.120	1.083	0.302	0.297	0.319
	(0.027)	(0.028)	(0.028)	(0.003)	(0.003)	(0.003)

of δ_t are recovered well. The model also captures well the peaks of ξ_t, so 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. 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 recursive estimate $\hat{\tau}_t$ versus the dynamic τ_t^* of Patton et al.

(2019) in Table 1, differences are mostly small and insignificant. If there is no time-variation (path (1)), the recursive estimate does slightly better, as expected. The converse is true for δ_t if the true parameters vary over time.

Third, Figure 2 as well as Figures F.1 and F.2 in Web Appendix F.1 corroborate that our EVT-based market risk measures, such as VaR and ES at a high confidence level $\gamma = 0.99$, tend to be estimated sufficiently accurately when used in combination with our modeling approach. The low and high frequency dynamics of the VaR and ES are both captured well. There only appears some under-estimation of the ES at its very peak where tails are extremely fat. Overall, we conclude that the model captures well the dynamics of the tails, even if the model does not coincide with the data generating process and is therefore misspecified.

4 The tail impact of Eurosystem asset purchases

4.1 Data

4.1.1 High-frequency data on bond yields

We obtain high-frequency data on changes in euro area sovereign bond yields from Thomson Reuters/Datastream, focusing on Spanish (EN), Greek (GR), Irish (IE), Italian (IT), and Portuguese (PT) five-year sovereign benchmark bonds. These market segments were among the most affected by the euro area debt crisis; see e.g. ECB (2014). SMP bond purchases undertaken during the debt crisis predominantly targeted the two- to ten-year maturity bracket, with the five-year maturity approximately in the middle of that spectrum. We focus on the impact on five-year benchmark bonds for this reason. We model the midpoint between ask and bid prices. Bond prices are expressed in yields-to-maturity and are obtained from continuous dealer quotes.

Our sample ranges from 04 January 2010 to 31 December 2012, covering the most intense phase of the euro area sovereign debt crisis. The bond yields are sampled at the 15-minute frequency between 8AM and 6PM. Following Ghysels et al. (2017) we do not

consider overnight changes in yield, such that the first 15-minute interval covers 8AM to 8:15AM. This yields 40 intra-daily observations per trading day. This yields 40 intra-daily observations per day, with $T \approx 3 \times 260 \times 40 \approx 31,000$ observations per country.

The Greek data are an exception. Greek bonds experienced a credit event on 09 March 2012. In January and February 2012 the five-year benchmark bond continued trading, infrequently and at low prices, until approximately one week before the credit event. Our Greek data sample ends on 02 March 2012 for this reason. We include the Greek pre-default data as a truly extreme case, allowing us to "stress-test" our EVT estimation methodology.

Figure F.3 in the Web Appendix F.2 plots the yield-to-maturity of our five benchmark bond yields in levels and in first differences. All five yields exhibited large and sudden moves, leading to volatility clustering and extreme realizations of yield changes during the euro area sovereign debt crisis.

Table 2 provides summary statistics for changes in our five benchmark bond yields sampled at the 15-minute frequency. All time series have significant non-Gaussian features under standard tests and significance levels. In particular, we note the non-zero skewness and large values of kurtosis for almost all time series in the sample. Yield changes are covariance stationary according to standard unit root (ADF) tests. Most yield changes are below one bps in absolute value. This suggests that the data are not only heavy-tailed, but also extremely peaked around zero in the center. The pronounced non-Gaussian data features strongly suggest a non-Gaussian empirical framework for modeling conditional location, dispersion, and higher-order moments.

4.1.2 High–frequency data on Eurosystem bond purchases

We study the impact of SMP bond purchases between 2010 and 2012 for five euro area countries: Greece, Ireland, Italy, Portugal, and Spain. At the end of our sample, the Eurosystem held €99.0 bn in Italian sovereign bonds, €30.8 bn in Greek debt, €43.7 bn in Spanish debt, €21.6 bn in Portuguese debt, and €13.6 bn in Irish bonds; see the ECB (2013)'s Annual Report. The SMP's daily cross-country breakdown of the purchase data is still confidential at the time of writing. We use the country-specific data on SMP purchases when studying

Table 2: Data descriptive statistics

Summary statistics for changes in five-year sovereign benchmark bond yields measured in percentage points. Columns labeled EN, GR, IE, IT, and PT refer to Spanish, Greek, Irish, Italian, and Portuguese five-year benchmark bond yields. The sample ranges from 04 January 2010 to 28 December 2012. The Greek sample ends on 02 March 2012. Reported p-values for skewness and kurtosis refer to D'Agostino et al. (1990)'s test. The last row reports the fraction of yield changes smaller than one basis point in absolute value.

	EN	GR	IE	IT	PT
Median	0.00	0.00	0.00	0.00	0.00
Std. dev.	0.02	0.46	0.06	0.03	0.08
Minimum	-0.74	-20.73	-0.91	-0.39	-1.15
Maximum	0.47	14.77	1.45	0.43	1.20
Skewness	-42.29	-104.76	34.11	14.91	12.40
Skew. p-value	0.00	0.00	0.00	0.00	0.00
Kurtosis	357.94	195.05	301.26	293.40	279.44
Kurt. p-value	0.00	0.00	0.00	0.00	0.00
Fraction $y_t < 1$ bp	81%	77%	81%	81%	77%

the impact of the program.

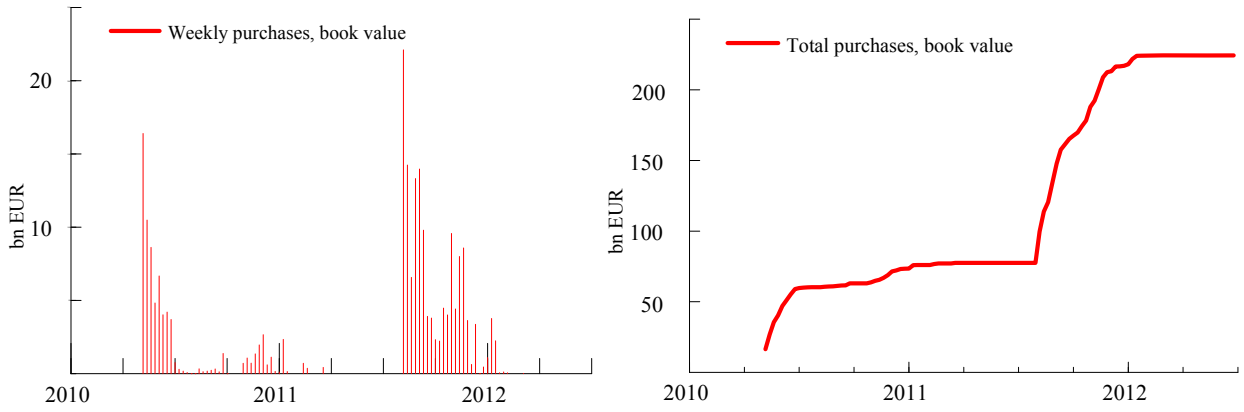
The SMP had the objective of helping to restore the monetary policy transmission mechanism by addressing the malfunctioning of certain government bond markets. The SMP consisted of interventions in the form of outright secondary market purchases. Implicit in the concept of malfunctioning markets is the notion that government bond yields can be unjustifiably high and volatile.

Figure 3 plots weekly total SMP purchases across countries as well as their accumulated book value over time. Approximately €214 billion (bn) of bonds were acquired within the SMP between 2010 and early 2012. The SMP was announced on 10 May 2010 and initially focused on Greek, Irish, and Portuguese debt securities. The program was extended to include Italian and Spanish bonds on 8 August 2011. The SMP was replaced by the Outright Monetary Transactions (OMTs) program on 6 September 2012; see Cœuré (2013). Visibly, the purchase data are unevenly spread over time. Between 10 May 2010 and Spring 2012 there are long periods during which the SMP was open but inactive.

The SMP purchase data are time-stamped, allowing us to construct time series data z_t of country-specific SMP purchases at the high (15-minute) frequency. The 15-minute frequency is chosen because 15 minutes is the regulatory limit for the recording of trades

Figure 3: Weekly and total SMP purchase amounts.

The figure plots the book value of settled SMP purchases as of the end of a given week. We report weekly purchases across countries (left panel) as well as the cumulative amounts (right panel). Maturing amounts are excluded.



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.

4.2 Location–scale–df model estimates

This section applies the location–scale–df model of Section 2.4 to study changes in the yield-to-maturity of five-year sovereign benchmark bonds as discussed in Section 4.1.1. We are particularly interested in each series' location, scale, and degrees of freedom, and how these respond to Eurosystem bond purchases.

We apply the model to the raw data series after removing a (negligible) intra-daily pattern via dummy variable regression. We introduce two simplifications to the general specification. First, preliminary analyses suggest that the location parameters are approximately time-invariant, such that a^μ and b^μ are close to zero. We proceed by imposing this restriction. Note that the specification for the mean still includes $d^\mu \cdot y_{t-1}$ to accommodate a potentially negative serial correlation at the 15-minute frequency; see e.g. Roll (1984). Second, we find that the persistence (b^σ and b^ν) in volatility and degrees of freedom parameters is very high. We therefore set $\omega^\sigma = \omega^\nu = 0$ and $b^\sigma = b^\nu = 1$, thus adopting the EWMA restricted score

Table 3: Parameter estimates for the location–scale–df model

Parameter estimates for the univariate location–scale–df model (13). Rows labeled EN, GR, IE, IT, and PT refer to Spanish, Greek, Irish, Italian, and Portuguese five-year benchmark bond yields. The estimation sample ranges from 04 January 2010 to 28 December 2012 for all countries except Greece. Standard error estimates are in round brackets and are taken from a sandwich covariance matrix. P-values are provided in square brackets.

	EN	GR	IE	IT	PT
ω^μ	0.013	-0.038	0.003	-0.003	-0.016
	(0.007)	(0.024)	(0.005)	(0.007)	(0.006)
	[0.046]	[0.108]	[0.641]	[0.618]	[0.004]
c^μ	-2.623	-2.856	0.017	-1.479	-0.053
(SMP)	(0.941)	(2.483)	(1.594)	(0.552)	(2.068)
	[0.005]	[0.250]	[0.992]	[0.007]	[0.980]
d^μ	-0.039	-0.000	-0.010	-0.029	-0.004
(AR1)	(0.007)	(0.000)	(0.002)	(0.010)	(0.001)
	[0.000]	[0.200]	[0.000]	[0.004]	[0.003]
a^σ	0.107	0.141	0.135	0.124	0.089
	(0.015)	(0.011)	(0.012)	(0.013)	(0.011)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
c^σ	-0.126	-0.055	-0.461	-0.049	-0.441
(SMP)	(0.089)	(0.115)	(0.324)	(0.050)	(0.228)
	[0.158]	[0.635]	[0.155]	[0.325]	[0.053]
d^σ	0.004	-0.004	-0.000	0.005	-0.000
(LEV)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
	[0.006]	[0.001]	[0.885]	[0.002]	[0.641]
a^ν	0.004	0.018	0.007	0.006	0.008
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
c^ν	0.031	0.042	0.022	0.013	0.000
(SMP)	(0.014)	(0.027)	(0.035)	(0.007)	(0.028)
	[0.033]	[0.122]	[0.530]	[0.052]	[0.993]
loglik	-56226.2	-68788.4	-68584.7	-56218.6	-78164.0
AIC	112468.4	137592.9	137185.4	112453.2	156343.9
BIC	112534.9	137656.8	137252.0	112519.8	156410.6

dynamics of Lucas and Zhang (2016) for the scale and df parameters. A comparison of model selection criteria (AIC, BIC) across model specifications confirms these choices. With these simplifications in place, Table 3 now presents the parameter estimates.

We focus on three findings in Table 3. First, SMP bond purchases tended to lower the conditional location of future bond yields for most countries. The estimate of c^μ is negative for four out of five countries, and is statistically significantly negative for two of them. The estimated impacts for the two largest SMP countries, Italy and Spain, are -1.5 bps and -2.6 bps per €1 bn of purchases, respectively. The highest impact per €1 bn is observed for

Greek bonds, at -2.9 bps per €1 bn of purchases. Greek bonds were the most illiquid at the time. The estimates of c^μ for IE and PT are smaller in magnitude and not significant. Since the yields are modeled in first differences and $a^\mu = b^\mu = 0$, these impacts are associated with long-lasting (permanent) changes in yield *levels*. Overall, our c^μ estimates are in line with those obtained by Eser and Schwaab (2016) based on daily data and factor modeling techniques, and approximately similar but less dispersed than those obtained by Ghysels et al. (2017) based on high-frequency data and VAR/GARCH modeling techniques.

Second, our parameter estimates for c^σ suggest a reduction in scale (volatility) following SMP bond purchases. The point estimates are all negative, although none are statistically significant at a 5% confidence level. Sizeable standard error estimates for c^σ are intuitive because the SMP intervention data is scarce even at the 15-minute frequency and the log-scale is subject to pronounced time series variation.

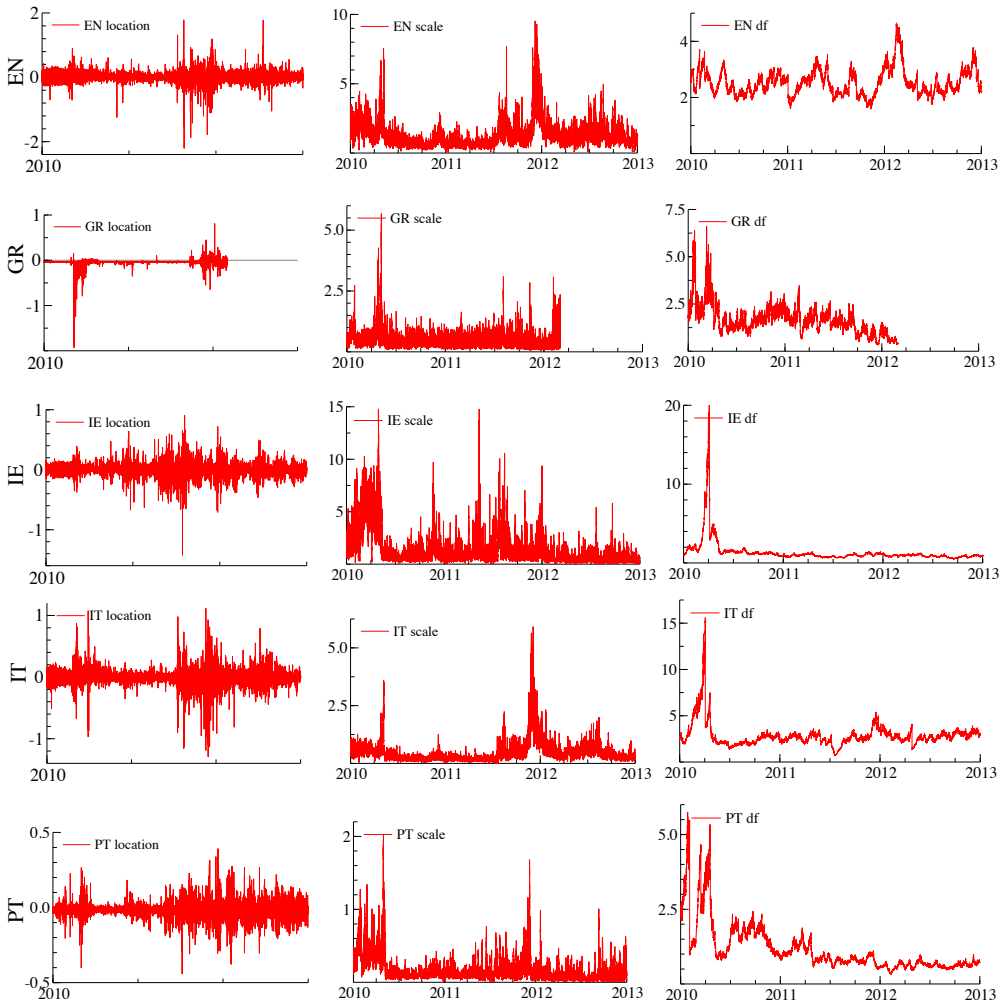
Third, the point estimates of c^ν are all positive, and statistically significant in one case (EN, with IT a borderline case). As a result, the time-varying degrees of freedom ν_t tend to increase following SMP bond purchases, suggesting an increasingly "Gaussian" tail shape when the central bank is active as a buyer-of-last-resort. Taken together, the estimates of $c^\mu < 0$, $c^\sigma < 0$, and $c^\nu > 0$ suggest an overall beneficial, market–stabilizing impact of the bond purchases on sovereign bond yields.

The remaining parameter estimates in Table 3 can be interpreted as follows. The autoregressive coefficients d^μ associated with lagged y_{t-1} are all negative. The negative autocorrelation is in line with severely illiquid markets for all five sovereign bonds during our sample; see Roll (1984). The intercept terms ω^μ are small and statistically different from zero only in two out of five cases (for EN and PT). The leverage terms d^σ are positive and statistically significant for EN and IT. In these cases an increase in yield has a greater influence on future log–scale than a decrease.

Figure 4.2 plots all time-varying parameters μ_t, σ_t, and ν_t. The conditional location parameters μ_t (left column) tend to be small and rarely exceed one bp in absolute value. The observed time-variation is due to the inclusion of the lagged term y_{t-1} and bond purchases z_{t-1}. High values for the conditional scale σ_t (middle column) are visible for Greece, Ireland

Figure 4: Filtered location, scale, and degrees of freedom parameters

Filtered location (first column), scale (second column) and degrees of freedom (third column) parameters associated with the location–scale–df model introduced in Section 2.4. Rows labeled EN, GR, IE, IT, and PT refer to Spanish, Greek, Irish, Italian, and Portuguese five-year benchmark bond yields. Greek bonds discontinued trading after 02 March 2012, and experienced a credit event on 09 March 2012.



and Portugal in 2010, and for Spain and Italy in late 2011. The conditional df parameters ν_t (right column) suggest that the conditional distribution is profoundly heavy-tailed, even after allowing for time-variation in the location and scale parameters. The df parameter associated with the Greek data declines almost monotonically until the credit event in March 2012.

Table 4: Parameter estimates

Parameter estimates for the extended (with SMP purchases z_t) tail shape model. Columns labeled EN, GR, IE, IT, and PT refer to Spanish, Greek, Irish, Italian, and Portuguese five-year benchmark bond yields. The estimation sample ranges from 04 January 2010 to 28 December 2012 for all countries except Greece. Standard error estimates are in round brackets and are constructed from a sandwich covariance matrix. P-values are in square brackets.

	EN	GR	IE	IT	PT
α^{ξ}	0.006	0.032	0.021	0.055	0.026
	(0.006)	(0.006)	(0.007)	(0.012)	(0.011)
	[0.291]	[0.000]	[0.005]	[0.000]	[0.023]
α^{δ}	0.027	0.144	0.078	0.027	0.103
	(0.007)	(0.026)	(0.014)	(0.006)	(0.015)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
c^{ξ}	0.001	0.005	0.033	-0.009	-0.033
	(0.006)	(0.014)	(0.045)	(0.010)	(0.029)
	[0.847]	[0.705]	[0.467]	[0.376]	[0.258]
c^{δ}	-0.013	-0.031	0.060	-0.011	0.107
	(0.014)	(0.034)	(0.077)	(0.005)	(0.083)
	[0.344]	[0.362]	[0.433]	[0.036]	[0.196]
a^{τ}	0.010	0.312	0.143	0.027	0.263
T	30279	21839	30799	30519	30719
T^*	3003	2223	3093	3041	3084
loglik	-102027.1	-152861.1	-226080.5	-103050.5	-306523.8
AIC	204062.1	305730.2	452168.9	206109.0	613055.7
BIC	204095.4	305762.2	452202.3	206142.3	613089.0

4.3 Tail shape and tail scale estimates

This section discusses our time-varying tail shape (ξ_t) and tail scale (δ_t) estimates. We focus on results obtained from pre-filtered data, where we used the fitted location–scale–df model from Section 4.2 to clean y_t from location and scale effects. Web Appendix G presents the analogous tail shape and scale estimates from raw bond yield data, allowing us to compare the two approaches; see Section 4.4. Our main results are based on POT observations $x_t = (y_t - \hat{\mu}_t)/\hat{\sigma}_t - \hat{\tau}_t$ if $(y_t - \hat{\mu}_t)/\hat{\sigma}_t > \hat{\tau}_t$ and $x_t = missing$ otherwise, where $\hat{\mu}_t$ and $\hat{\sigma}_t$ are the location and scale estimates as reported in Figure 4.2, and where $\hat{\tau}_t$ was obtained using the autoregressive specification (10) in conjunction with the objective function (12).

Preliminary analyses suggest that changes in the tail shape and scale parameters are highly persistent for our high-frequency data. We thus set $\omega^{\xi} = \omega^{\delta} = 0$ and $b^{\xi} = b^{\delta} = 1$ to simplify the model, again adopting the EWMA restricted score dynamics of Lucas and

Zhang (2016) for these parameters. We also set the smoothing parameter $\lambda = 0$, see (7), given the absence of mean reversion in $\ln \xi_t$ and $\ln \delta_t$ at the 15-minute frequency.[2] We allow $C \neq 0$ such that SMP bond purchases z_{t-1} can impact both $\ln \xi_t$ and $\ln \delta_t$ via their impact coefficients c^ξ and c^δ. A comparison of unreported model selection criteria (AIC, BIC) across different model specifications supports these choices.

Table 4 presents tail shape estimates based on pre-filtered data. Parameters a^ξ and a^δ can be interpreted as the standard deviations of the scores driving $\ln \xi_t$ and $\ln \delta_t$, respectively; see the statements above (6). The associated estimates suggest pronounced time series variation in both parameters. The SMP impact parameters c^ξ are estimated negatively in two out of five cases, but are not statistically significant according to their t-values. Estimates of c^δ are negative in three out of five cases, and are significantly negative in one case (IT). As a result, most of the tail impact of SMP purchases appears to have come about through its impact on the center of the distribution (μ_t, σ_t) and not on its tail shape (δ_t, ξ_t).

Figure 5 plots the corresponding filtered estimates for time-varying tail shape ξ_t and tail scale δ_t. Blue bars indicate the approximate timing of SMP purchases in the respective markets. Time series variation is present and pronounced in both tail shape and tail scale parameters. The heaviest tail is estimated for Greek bonds during the weeks preceding the credit event on 09 March 2012. The tail shape parameter can be above one, suggesting that no conditional mean, variance, and ES exist at such times. The other estimates for ξ_t typically vary between zero and one. Estimates above one can occur but are temporary and rare. Time-variation in δ_t is pronounced as well.

Table 5 addresses the question how market risk measures responded on average to €1 bn of central bank bond purchases. The total impact is decomposed into the impact on the conditional location, scale, tail scale, and tail shape by setting the non-active summands in (22) to zero. The estimates corroborate that most of the SMP's effect on extreme market risk came from its impact on location and scale, and thus from its impact on the center of the distribution. We estimate that the 97.5% VaR was reduced by 3.8, 6.0, 5.9, 2.1, and 6.9 bps

[2]The smoothing parameter λ is hard to estimate numerically given the absence of mean reversion in our high-frequency data. Fixing it to reasonable alternative values has little effect on our empirical findings.

Figure 5: Filtered tail shape and tail scale estimates

Filtered ξ_t and δ_t estimates for Spanish (EN), Greek (GR), Irish (IE), Italian (IT), and Portuguese (PT) five-year sovereign benchmark bond yields between 2010 and 2012. The sample for Greek bonds is shorter as these bonds discontinued trading on 02 March 2012 and experienced a credit event on 09 March 2012. Standard error bands are simulated at a 95% confidence level. Blue bars indicate the approximate timing of SMP purchases in the respective markets. The SMP amounts are first aggregated over all five SMP countries, then smoothed using a centered one-week moving average. The resulting common time series is reported in the respective panel if the SMP was active for the respective market segment at the time; see Section 4.1.2 for the respective announcement days. The scaling of the purchase amounts is omitted and differs between left and right panels only for visibility.

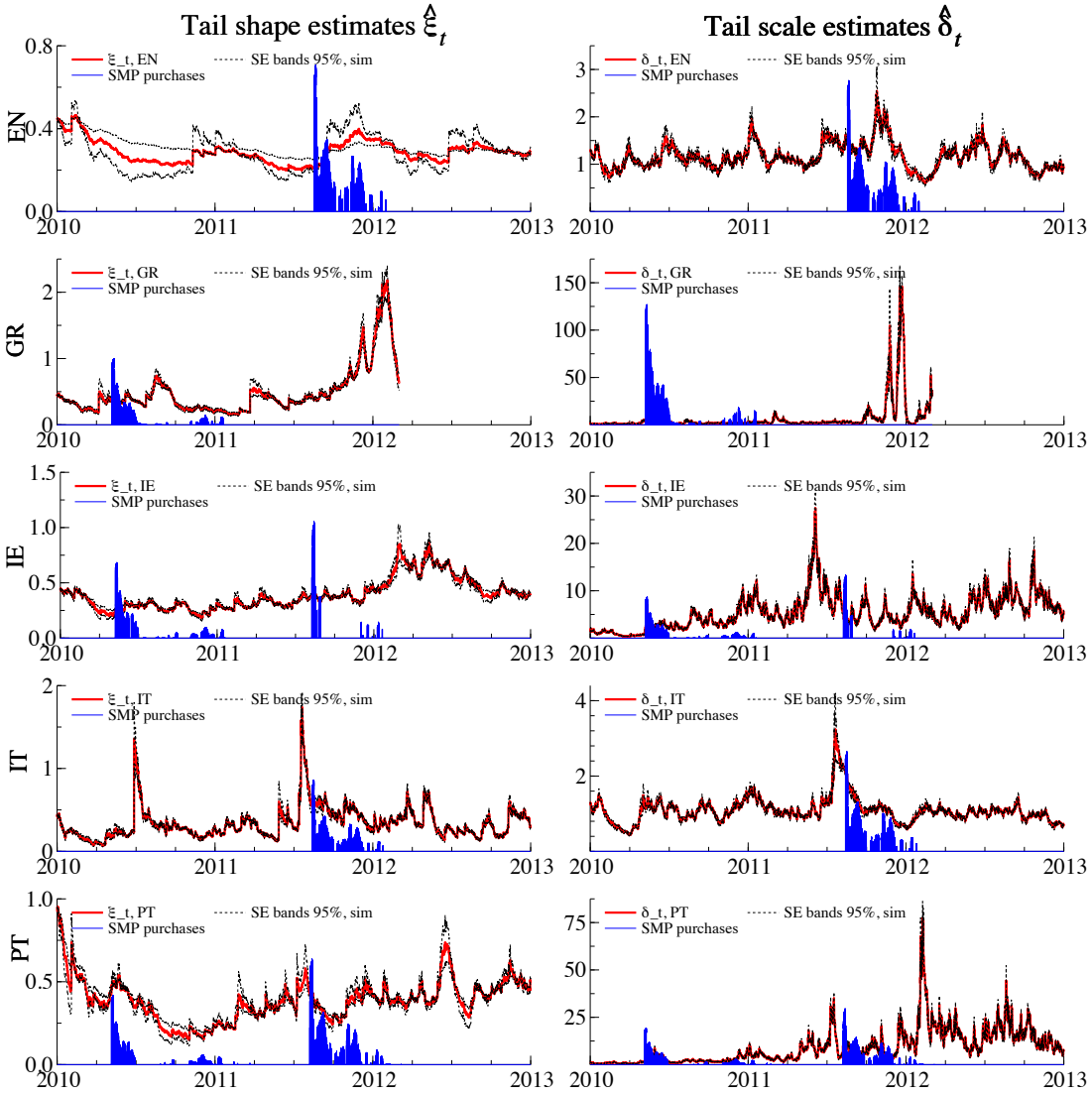


Table 5: Impact of €1 bn of SMP purchases on 97.5% VaR and 99.5% VaR

The table reports the impact of €1 bn of SMP bond purchases on the 97.5% and 99.5% VaR. The total impact is estimated as $(1/\sum_t^T z_t)\sum_t^T (\mathrm{dVaR}^\gamma(y_t)/\mathrm{d}z_{t-1})\, z_{t-1}$, where $\mathrm{dVaR}^\gamma(y_t)/\mathrm{d}z_{t-1}$ is given by (22). The total impact is decomposed into the impact on the conditional location μ_t, scale σ_t, tail scale δ_t, and tail shape ξ_t by setting the non-active summands in (22) to zero. Columns labeled EN, GR, IE, IT, and PT refer to Spanish, Greek, Irish, Italian, and Portuguese five-year benchmark bond yields.

	97.5% VaR				
	EN	GR	IE	IT	PT
μ_t	-2.623	-2.856	0.017	-1.479	-0.053
σ_t	-1.155	-2.439	-6.570	-0.537	-8.006
δ_t	-0.068	-0.702	0.587	-0.042	1.216
ξ_t	0.001	0.037	0.068	-0.013	-0.040
Total	-3.845	-5.961	-5.899	-2.071	-6.883

	99.5% VaR				
	EN	GR	IE	IT	PT
μ_t	-2.623	-2.856	0.017	-1.479	-0.053
σ_t	-2.302	-5.361	-14.604	-1.163	-18.578
δ_t	-0.190	-2.120	1.638	-0.131	3.556
ξ_t	0.009	0.260	0.443	-0.096	-0.278
Total	-5.106	-10.077	-12.507	-2.870	-15.352

per €1 bn Eurosystem intervention in Spanish, Greek, Irish, Italian, and Portuguese five-year benchmark bonds, respectively. The impact grows with the extremeness of the VaR. The 99.5% VaR estimate is reduced, respectively, by 5.1, 10.1, 12.5, 2.9, and 15.4 bps per €1 bn of Eurosystem purchases in the above bonds. The table also shows that these improvements were obtained without worsening the tail risk. If anything, additional beneficial secondary effects came about via the SMP's effect on tail shape and tail scale parameters for large economies such as Spain and Italy.[3]

4.4 Tail shape and tail scale estimates from raw data

We conclude our empirical study with a discussion of tail shape and tail scale estimates from raw (un-prefiltered) data y_t. The dynamic tail shape and tail scale model of Section 2

[3] Additional, beneficial SMP announcement effects are not taken into account in Table 5. This is because both the 09 May 2010 and 08 August 2011 SMP announcements occurred when markets were closed, and are therefore not part of our sample. The VaR impact estimates are conservative in this sense.

could be robust to omitted variation in the center of the distribution $g(y_t \mid \mathcal{F}_{t-1})$. This is because of two effects. First, the autoregressive specification of τ_t via (10) implies that τ_t can adjust to time variation in the center of the distribution. The resulting exceedances $\tilde{x}_t = y_t - \tilde{\tau}_t$ from unfiltered data can therefore in practice still be close to the exceedances $x_t = (y_t - \mu_t)/\sigma_t - \tau_t$ from pre-filtered data. Second, the dynamic specification of δ_t via (7) implies that the tail scale could mop up omitted time-variation in σ_t, leaving ξ_t free to fit the time-variation in tail shape.

Web Appendix G discusses our tail shape and tail scale estimates obtained from POTs $\tilde{x}_t = y_t - \tilde{\tau}_t$, along with the model's deterministic parameters. The deterministic parameter estimates a^τ and a^δ are now uniformly higher than those reported in Table 4. This is intuitive, as both τ_t and δ_t need to adjust more strongly to also pick up omitted variation from the center of the distribution. The estimates of c^δ, the SMP impact on tail scale, are now all negative, and statistically significantly negative at a 5% confidence level in two cases (IE, PT). This is intuitive, as δ_t now not only captures dispersion in the tail, but to some extent also in the center. The point estimates of c^ξ are not statistically significant. This is in line with our earlier conclusion that most of the SMP's beneficial impact on extreme tail quantiles (97.5% and 99.5% VaR) does not come via its impact on tail shape and tail scale but instead via its impact on location and scale in the center of the distribution. The estimates $\hat{\xi}_t$ and $\hat{\delta}_t$ from un-prefiltered data are more volatile, and visibly different from the estimates from pre-filtered data as reported in Figure 5. This suggests that the score-driven updates of τ_t and δ_t do not fully absorb all variation in μ_t and σ_t for our data at hand. We therefore prefer the estimates based on appropriately prefiltered data.

5 Conclusion

We introduced a semi-parametric conditional extreme value theory (EVT) modeling framework to study time variation in tail fatness 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 deterministic and time-varying parameters, and how to relate tail variation to observed covariates. 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 (2020), 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 the impact of bond purchases within the Eurosystem's SMP between 2010 and 2012 on the extreme upper tail of sovereign bond yield changes measured at a high frequency, concluding that the program had a beneficial impact on extreme tail quantiles, leaning against the risk of extremely adverse market outcomes while active.

References

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

Blasques, F., P. Gorgi, and S. J. Koopman (2018). Missing observations in observation-driven time series models. *Tinbergen Institute Discussion Paper 2018-013/III*, 1–39.

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., S. J. Koopman, and A. Lucas (2020). Meximum likelihood estimation for score-driven models. *Tinbergen Institute discussion papers 2014-029/III, update July 2020*, 1–54.

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.

Cœuré, B. (2013). Outright Monetary Transactions, one year on. Speech at the conference "The ECB and its OMT programme," Berlin, 2 September 2013.

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., 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.

D'Agostino, R. B., A. Belanger, and R. B. D'Agostino Jr (1990). A suggestion for using powerful and informative tests of normality. *The American Statistician 44*(4), 316–321.

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 (2020). Trends in extreme value indices. *Journal of the Americal Statistical Association, forthcoming*.

ECB (2013). European Central Bank Annual Report 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.

Engle, R. F. and A. J. Patton (2001). What good is a volatility model? *Quantitative Finance 1*(2), 237–245.

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.

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.

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 and Economic Statistics 19*(1), 208–216.

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

Koenker, R. and B. J. Park (1996). An interior point algorithm for nonlinear quantile regression. *Journal of Econometrics 71*(1-2), 265–283.

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.

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

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.

Pelizzon, L., M. Subrahmanyam, D. Tomio, and J. Uno (2016). Sovereign credit risk, liquidity, and ECB intervention: Deus Ex Machina? *Journal of Financial Economics 122*(1), 44–52.

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.

Roll, R. (1984). A simple implicit measure of the effective bid-ask spread in an efficient market. *The Journal of Finance 39*(4), 1127–1139.

Vayanos, D. and J.-L. Vila (2009). A preferred habitat model of the term structure of interest rates. NBER Working Paper 15487.

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

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

Web Appendix to

Modeling extreme events:

time-varying extreme tail shape*

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

[(a)] European Central Bank, Financial Research

[(b)] Sveriges Riksbank, Research Division

[(c)] Vrije Universiteit Amsterdam and Tinbergen Institute

*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.

A Appendix: GPD score and scaling functions

A.1 The GPD score function

This section derives the score (5). Recall the GPD pdf as

$$p(x_t;\delta_t,\xi_t)=\frac{1}{\delta_t}\left(1+\xi_t\frac{x_t}{\delta_t}\right)^{-\frac{1}{\xi_t}-1}.$$

with log-likelihood contribution

$$l_t=\ln p(x_t;\delta_t,\xi_t)=-\ln(\delta_t)-\left(1+\frac{1}{\xi_t}\right)\ln\left(1+\xi_t\frac{x_t}{\delta_t}\right),$$

where $\delta_t>0$, $\xi_t>0$, and $x_t>0$. Using $\xi_t=\exp(f_{1t})$, the first element of the score is obtained as

$$\begin{aligned}\nabla_{1t} &= \frac{\partial l(x_t;\delta_t,\xi_t)}{\partial f_{1t}}=\frac{\partial l(x_t;\delta_t,\xi_t)}{\partial \xi_t}\cdot\frac{\mathrm{d}\xi_t}{\mathrm{d}f_{1t}},\\ \frac{\partial l(x_t;\delta_t,\xi_t)}{\partial \xi_t} &= \frac{1}{\xi_t^2}\ln\left(1+\xi_t\frac{x_t}{\delta_t}\right)-\left(1+\frac{1}{\xi_t}\right)\frac{x_t}{\delta_t+\xi_t x_t},\\ \frac{\mathrm{d}\xi_t}{\mathrm{d}f_{1t}} &= \exp(f_{1t})=\xi_t.\end{aligned}$$

Similarly, for $\delta_t=\exp(f_{2t})$, the second element of the score is obtained as

$$\begin{aligned}\nabla_{2t} &= \frac{\partial l(x_t;\delta_t,\xi_t)}{\partial f_{2t}}=\frac{\partial l(x_t;\delta_t,\xi_t)}{\partial \delta_t}\cdot\frac{\mathrm{d}\delta_t}{\mathrm{d}f_{2t}},\\ \frac{\partial l(x_t;\delta_t,\xi_t)}{\partial \delta_t} &= \frac{x_t-\delta_t}{\delta_t(\delta_t+\xi_t x_t)},\\ \frac{\mathrm{d}\delta_t}{\mathrm{d}f_{2t}} &= \exp(f_{2t})=\delta_t.\end{aligned}$$

Combining the two, the unscaled score vector is given by

$$\nabla_t=\begin{bmatrix}\frac{1}{\xi_t}\ln\left(1+\xi_t\frac{x_t}{\delta_t}\right)-\left(1+\frac{1}{\xi_t}\right)\frac{\xi_t x_t}{\delta_t+\xi_t x_t}\\ \frac{x_t-\delta_t}{\delta_t+\xi_t x_t}\end{bmatrix}.$$

A.2 The GPD scaling matrix

This section derives the scaled score (7). To this end we require the $[2 \times 2]$ conditional Fisher information matrix associated with (5),

$$\mathcal{I}_t = \mathrm{E}[\nabla_t \nabla_t' \mid \mathcal{F}_{t-1}; f_t, \theta] = \begin{bmatrix} \mathcal{I}_t^{(11)} & \mathcal{I}_t^{(12)} \\ \mathcal{I}_t^{(21)} & \mathcal{I}_t^{(22)} \end{bmatrix}. \tag{A.1}$$

We derive each element in turn.

Element $\mathcal{I}_t^{(11)}$

We recall that the score is zero in expectation if the model is well-specified; see Creal et al. (2013). This implies

$$\int_0^\infty \frac{1}{\xi_t^2} \ln\left(1 + \xi_t \frac{x_t}{\delta_t}\right) p(x_t; \delta_t, \xi_t) \mathrm{d}x_t = \int_0^\infty \left(1 + \frac{1}{\xi_t}\right) \frac{x_t}{\delta_t + \xi_t x_t} p(x_t; \delta_t, \xi_t) \mathrm{d}x_t. \tag{A.2}$$

The top left element of the conditional Fisher information matrix is

$$\mathcal{I}_t^{(11)} = \mathrm{E}\left[-\left(\frac{\partial l(x_t; \delta_t, \xi_t)}{\partial \xi_t}\right)^2 \left(\frac{\mathrm{d}\xi_t}{\mathrm{d}f_t}\right)^2 \mid \mathcal{F}_{t-1} \right] = \mathrm{E}\left[-\frac{\partial^2 l(x_t; \delta_t, \xi_t)}{\partial \xi_t^2} \mid \mathcal{F}_{t-1} \right] \exp(2f_t),$$

where the last equality uses the fact that f_t is fixed for given $\mathcal{F}_{t-1}$. The expected negative second derivative is given by

$$\begin{aligned}
& \mathrm{E}\left[-\frac{\partial^2 l(x_t; \delta_t, \xi_t)}{\partial \xi_t^2} \mid \mathcal{F}_{t-1} \right] \\
= & -\int_0^\infty \left[\left(1 + \frac{1}{\xi_t}\right) \frac{x_t^2}{(\delta_t + \xi_t x_t)^2} + \frac{2}{\xi_t^2} \frac{x_t}{\delta_t + \xi_t x_t} - \frac{2}{\xi_t^3} \ln\left(1 + \xi_t \frac{x_t}{\delta_t}\right) \right] p(x_t; \delta_t, \xi_t) \mathrm{d}x_t \\
= & -\int_0^\infty \left[\left(1 + \frac{1}{\xi_t}\right) \frac{x_t^2}{(\delta_t + \xi_t x_t)^2} + \frac{2}{\xi_t^2} \frac{x_t}{\delta_t + \xi_t x_t} - \frac{2}{\xi_t} \left(1 + \frac{1}{\xi_t}\right) \frac{x_t}{\delta_t + \xi_t x_t} \right] p(x_t; \delta_t, \xi_t) \mathrm{d}x_t \\
= & -\int_0^\infty \left[\left(1 + \frac{1}{\xi_t}\right) \frac{x_t^2/\delta_t^2}{(1 + \xi_t x_t/\delta_t)^2} - \frac{2}{\xi_t} \frac{x_t/\delta_t}{1 + \xi_t x_t/\delta_t} \right] \frac{1}{\delta_t} \left(1 + \xi_t \frac{x_t}{\delta_t}\right)^{-\frac{1}{\xi_t} - 1} \mathrm{d}x_t \\
= & -\int_0^\infty \left[\left(\frac{1 + \xi_t}{\xi_t^3}\right) \frac{\xi_t^2 x_t^2/\delta_t^2}{(1 + \xi_t x_t/\delta_t)^2} - \frac{2}{\xi_t^2} \frac{\xi_t x_t/\delta_t}{1 + \xi_t x_t/\delta_t} \right] \frac{1}{\delta_t} \left(1 + \xi_t \frac{x_t}{\delta_t}\right)^{-\frac{1}{\xi_t} - 1} \mathrm{d}x_t \\
= & -\frac{1 + \xi_t}{\xi_t^4} \int_1^\infty (u_t - 1)^2 u_t^{-1/\xi_t - 3} \mathrm{d}u_t + \frac{2}{\xi_t^3} \int_1^\infty (u_t - 1) u_t^{-1/\xi_t - 2} \mathrm{d}u_t,
\end{aligned} \tag{A.3}$$

where we used (A.2) in the second line, and where the last equality comes from a change of variable substituting $u_t = 1 + \xi_t x_t/\delta_t$.

It is straightforward to check that

$$\begin{aligned}
\int_1^\infty (u_t-1)^2 u_t^{-1/\xi_t-3}\mathrm{d}u_t &= \frac{2\xi_t^3}{(1+\xi_t)(1+2\xi_t)},\\
\int_1^\infty (u_t-1) u_t^{-1/\xi_t-3}\mathrm{d}u_t &= \frac{\xi_t^2}{(1+\xi_t)(1+2\xi_t)},\\
\int_1^\infty (u_t-1) u_t^{-1/\xi_t-2}\mathrm{d}u_t &= \frac{\xi_t^2}{1+\xi_t}.
\end{aligned}$$

Combining terms yields

$$\mathcal{I}_t^{(11)} = \frac{2}{(1+2\xi_t)(1+\xi_t)}\exp(2f_{1t}) = \frac{2\xi_t^2}{(1+\xi_t)(1+2\xi_t)}.$$

Element $\mathcal{I}_t^{(22)}$

The bottom right element of the conditional information matrix is given by

$$\mathcal{I}_t^{(22)} = \mathrm{E}\left[-\left(\frac{\partial l(x_t;\delta_t,\xi_t)}{\partial \delta_t}\right)^2\left(\frac{\mathrm{d}\delta_t}{\mathrm{d}f_{2t}}\right)^2 \mid \mathcal{F}_{t-1}\right] = \mathrm{E}\left[-\frac{\partial^2 l(x_t;\delta_t,\xi_t)}{\partial \delta_t^2}\mid \mathcal{F}_{t-1}\right]\exp(2f_{2t}).$$

The expectation term is given by

$$\begin{aligned}
&\mathrm{E}\left[-\frac{\partial^2 l(x_t;\delta_t,\xi_t)}{\partial \delta_t^2}\mid \mathcal{F}_{t-1}\right]\\
=\;& -\int_0^\infty \left[\frac{1/\delta_t^2 - 2x_t/\delta_t^3 - \xi_t x_t^2/\delta_t^4}{(1+\xi_t x_t/\delta_t)^2}\right]\frac{1}{\delta_t}\left(1+\xi_t\frac{x_t}{\delta_t}\right)^{-\frac{1}{\xi_t}-1}\mathrm{d}x_t\\
=\;& -\int_0^\infty \frac{1}{\delta_t^3}\left(1+\xi_t\frac{x_t}{\delta_t}\right)^{-\frac{1}{\xi_t}-3}\mathrm{d}x_t + \int_0^\infty \frac{2}{\delta_t^3}\frac{x_t}{\delta_t}\left(1+\xi_t\frac{x_t}{\delta_t}\right)^{-\frac{1}{\xi_t}-3}\mathrm{d}x_t + \int_0^\infty \frac{\xi_t}{\delta_t^3}\frac{x_t^2}{\delta_t^2}\left(1+\xi_t\frac{x_t}{\delta_t}\right)^{-\frac{1}{\xi_t}-3}\mathrm{d}x_t\\
=\;& -\frac{1}{\xi_t\delta_t^2}\int_1^\infty u_t^{-1/\xi_t-3}\mathrm{d}u_t + \frac{2}{\xi_t^2\delta_t^2}\int_1^\infty (u_t-1)u_t^{-1/\xi_t-3}\mathrm{d}u_t + \frac{1}{\xi_t^2\delta_t^2}\int_1^\infty (u_t-1)^2 u_t^{-1/\xi_t-3}\mathrm{d}u_t\\
=\;& -\frac{1}{\delta_t^2(1+2\xi_t)} + \frac{2}{\delta_t^2(1+\xi_t)(1+2\xi_t)} + \frac{2\xi_t}{\delta_t^2(1+\xi_t)(1+2\xi_t)}\\
=\;& \frac{1}{\delta_t^2(1+2\xi_t)},
\end{aligned} \tag{A.4}$$

such that

$$\mathcal{I}_t^{(22)} = \frac{1}{1+2\xi_t}.$$

Elements $\mathcal{I}_t^{(12)}$ and $\mathcal{I}_t^{(21)}$

The top right and bottom left elements of the conditional information matrix are given by

$$\mathcal{I}_t^{(12)} = \mathcal{I}_t^{(21)} = \mathrm{E}\left[-\frac{\partial^2 l(x_t;\delta_t,\xi_t)}{\partial\xi_t\partial\delta_t} \mid \mathcal{F}_{t-1}\right]\exp\left(f_{1t}+f_{2t}\right).$$

The derivation proceeds along similar lines as before,

$$\begin{aligned}
& \mathrm{E}\left[-\frac{\partial^2 l(x_t;\delta_t,\xi_t)}{\partial\xi_t\partial\delta_t} \mid \mathcal{F}_{t-1}\right] \\
= & -\int_0^\infty \left[\frac{-x_t/\xi_t}{\delta_t^2+\xi_t x_t\delta_t} + (1+\xi_t)\frac{x_t/\xi_t}{(\delta_t+\xi_t x_t)^2}\right]\frac{1}{\delta_t}\left(1+\xi_t\frac{x_t}{\delta_t}\right)^{-\frac{1}{\xi_t}-1}\mathrm{d}x_t \\
= & \frac{1}{\xi_t^3\delta_t}\int_1^\infty (u_t-1)u_t^{-1/\xi_t-2}\mathrm{d}u_t - \frac{1}{\xi_t^3\delta_t}\int_1^\infty (u_t-1)u_t^{-1/\xi_t-3}\mathrm{d}u_t - \frac{1}{\xi_t^2\delta_t}\int_1^\infty (u_t-1)u_t^{-1/\xi_t-3}\mathrm{d}u_t \\
= & \frac{1}{\xi_t\delta_t(1+\xi_t)} - \frac{1}{\xi_t\delta_t(1+2\xi_t)} \\
= & \frac{1}{\delta_t(1+\xi_t)(1+2\xi_t)}.
\end{aligned} \tag{A.5}$$

As a result,

$$\mathcal{I}_t^{(12)} = \mathcal{I}_t^{(21)} = \frac{\xi_t}{(1+\xi_t)(1+2\xi_t)}.$$

The scaling matrix

Collecting all elements $\mathcal{I}_t^{(11)}$ – $\mathcal{I}_t^{(22)}$ we obtain the conditional Fisher information matrix as

$$\mathcal{I}_t = \begin{bmatrix} \frac{2\xi_t^2}{(1+\xi_t)(1+2\xi_t)} & \frac{\xi_t}{(1+\xi_t)(1+2\xi_t)} \\ \frac{\xi_t}{(1+\xi_t)(1+2\xi_t)} & \frac{1}{1+2\xi_t} \end{bmatrix},$$

such that $\mathcal{I}_t^{-1} = L_t L_t'$ for

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

B Tail approximation for heavy-tailed random variables

B.1 Tail approximation for GPD random variables

Let $y_t \sim \text{GPD}(\alpha_t^{-1}, \sigma_t)$ be the data generating process (DGP) with $F(y_t) = 1 - (1 + y_t/(\alpha_t\sigma_t))^{-\alpha_t}$ as its cdf. Let τ be a threshold. We have

$$\begin{aligned}
\mathbb{P}\left[Y_t \leq y_t + \tau \mid Y_t > \tau\right] &= \frac{F(y_t + \tau) - F(\tau)}{1 - F(\tau)} \\
&= 1 - \frac{(1 + (y_t + \tau)/(\alpha_t\sigma_t))^{-\alpha_t}}{(1 + \tau/(\alpha_t\sigma_t))^{-\alpha_t}} \\
&= 1 - \frac{(1 + (y_t + \tau)/(\alpha_t\sigma_t))^{-\alpha_t}}{(1 + \tau/(\alpha_t\sigma_t))^{-\alpha_t}}
\end{aligned} \tag{B.1}$$

We put the slope at $y_t = 0$ for $\mathbb{P}\left[Y_t \leq y_t + \tau \mid Y_t > \tau\right]$ equal to that of $G_{\xi_t,\delta_{t,\tau}}(y_t)$, and also have the rate of decline the same. The latter implies $\xi_t = \alpha_t^{-1}$. The former implies

$$\delta_{t,\tau}^{-1} = \sigma_t^{-1}\frac{(1 + \tau/(\alpha_t\sigma_t))^{-\alpha_t - 1}}{(1 + \tau/(\alpha_t\sigma_t))^{-\alpha_t}} = \frac{\sigma_t^{-1}}{1 + \tau/(\alpha_t\sigma_t)} = \frac{\alpha_t}{\alpha_t\sigma_t + \tau} \quad \Leftrightarrow \quad \delta_{t,\tau} = \sigma_t + \alpha_t^{-1}\tau. \tag{B.2}$$

The EVT GPD fit now has a higher and higher scale. This is intuitive, as the DGP beyond the threshold has a flat tail. The GPD approximating tail, on the other hand, has a strong convexity near 0, where the convexity is larger for larger $\delta_{t,\tau}^{-1}$. As the DGP tail is fit, lower and lower values of $\delta_{t,\tau}^{-1}$ are needed as we go farther into the tails in order to match the fit of the GPD approximation to the flat tail. The resulting approximation is

$$\begin{aligned}
G_{\xi_t,\delta_{t,\tau}}(y_t) &= 1 - \left(1 + y_t/(\alpha_t(\sigma_t + \alpha_t^{-1}\tau))\right)^{-\alpha_t} \\
&= 1 - (1 + y_t/(\alpha_t\sigma_t + \tau))^{-\alpha_t} \\
&= 1 - \left(\frac{\alpha_t\sigma_t + \tau + y_t}{\alpha_t\sigma_t + \tau}\right)^{-\alpha_t} \\
&= 1 - \left(\frac{\alpha_t\sigma_t + \tau + y_t}{\alpha_t\sigma_t + \tau}\right)^{-\alpha_t} \\
&= 1 - \left(\frac{\alpha_t\sigma_t}{\alpha_t\sigma_t + \tau}\right)^{-\alpha_t}\left(1 + \frac{y_t + \tau}{\alpha_t\sigma_t}\right)^{-\alpha_t} \\
&= 1 - \frac{(1 + (y_t + \tau)/(\alpha_t\sigma_t))^{-\alpha_t}}{(1 + \tau/(\alpha_t\sigma_t))^{-\alpha_t}}.
\end{aligned} \tag{B.3}$$

Equation (B.3) coincides with (B.1). We can thus approximate the GPD DGP (B.1) by a GPD that has the *same* tail index, but a *different* scale parameter. The scale parameter $\delta_{t,\tau} = \sigma_t + \alpha_t^{-1}\tau$ increases with the threshold τ, varies positively with the tail shape parameter α^{-1}, and, importantly, should not be expected to provide a consistent estimate of σ_t. If σ_t were time-invariant, for example because pre-volatility-filtered data were modeled empirically, then the estimate $\delta_{t,\tau}$ may still vary over time to reflect time-variation in α_t.

B.2 Tail approximation for Student's t random variables

Let $y_t \sim \mathrm{t}(0, \sigma_t^2, \alpha_t)$ be the data generating process with $f(y_t)$ the pdf of a Student's t distribution with zero mean, scale σ_t^2, and α_t degrees of freedom. Let $\tau \in \mathbb{R}$ be a threshold.

Using the same method to match the parameters as in the previous example, we have $\xi_t^{-1} = \alpha_t$. For the scale, we equate the slopes and obtain

$$\begin{aligned}
\delta_{t,\tau}^{-1} &= \frac{f(\tau)}{1 - F(\tau)} \overset{\tau\to\infty}{\approx} \frac{-f'(\tau)}{f(\tau)} = \frac{\partial - \log f(\tau)}{\partial \tau} \\
&= \frac{(1 + \alpha_t^{-1})\tau/\sigma_t^2}{1 + \tau^2/(\alpha_t \sigma_t^2)} \\
&= \frac{(1 + \alpha_t)\tau}{\alpha_t \sigma_t^2 + \tau^2} \\
&\Leftrightarrow \\
\delta_{t,\tau} &\sim \frac{\alpha_t \sigma_t^2}{(1 + \alpha_t)\tau} + \frac{\tau}{1 + \alpha_t},
\end{aligned}$$

which again depends on α_t and increases in τ. For large τ, $\delta_{t,\tau}$ varies inversely with α_t, or positively with $\xi_t = \alpha_t^{-1}$. Again, we should not expect δ_t to be constant (or near one) even when working with volatility-filtered data.

C Analytic in-sample confidence bands for f_t

This section provides the expressions needed for the calculation of analytic in-sample confidence bands around the filtered time-varying parameters $\hat{f}_t(\theta)$. Such bands visualize the impact of estimation uncertainty associated with $\hat{\theta}$ on the filtered estimates $\hat{f}_t$. Delta-method-based bands are devised using a linear approximation of the non-linear transition function for f_t. As a by-product of our derivation we show how to extend Blasques et al. (2016, Section 3.2) to the case of a multivariate f_t with multiple lags.

If the linear approximation is not appropriate for a given dataset at hand, however, then delta-method-based bands can become unstable. This happens in our empirical application. In such cases we recommend using simulation-based bands; see Sections 2.2 and 4.

Recall that $f_t = (f_t^\xi, f_t^\delta)'$, where $\xi_t = \exp(f_t^\xi)$, $\delta_t = \exp(f_t^\delta)$, and the transition equations as

$$\begin{aligned} f_{t+1} &= \omega + A\tilde{s}_t + Bf_t, \\ \tilde{s}_t &= (1-\lambda)s_t + \lambda\tilde{s}_t, \end{aligned} \tag{C.1}$$

where $\omega = (\omega^\xi, \omega^\delta)'$, $A = \text{diag}(a^\xi, a^\delta)$, $B = \text{diag}(b^\xi, b^\delta)$, and s_t is given in (7).

In practice, some parameters may need to be restricted. Vector $\bar{\theta} = (\omega^\xi, \omega^\delta, a^\xi, a^\delta, b^\xi, b^\delta, \lambda)' \in \mathbb{R}^{7\times 1}$ collects all deterministic parameters of the model, while $\theta = (\omega^\xi, \omega^\delta, \alpha^\xi, \alpha^\delta, \beta^\xi, \beta^\delta, \lambda^{uc})'$ collects all unconstrained parameters. The two are related, for example, through $a^\xi = \exp(\alpha^\xi)$, $a^\delta = \exp(\alpha^\delta)$, $b^\xi = \Lambda(\beta^\xi)$, $b^\delta = \Lambda(\beta^\delta)$, $\lambda = \Lambda(\lambda^{uc})$, and where $\Lambda(x) = (1+\exp(-x))^{-1}$ is the logistic function. In this way, $a^\xi, a^\delta > 0$ and $0 < b^\xi, b^\delta, \lambda < 1$. We proceed with these restrictions, keeping in mind that some derivatives below would need to be adjusted when other restrictions were chosen or some parameters were fixed (for example, $\omega^\xi = \omega^\delta = 0$ and $b^\xi = b^\delta = 1$).

Pre-multiplying the factor updating equation (C.1) by $(1-\lambda L)$ yields

$$(1-\lambda L)\, f_{t+1} = (1-\lambda L)\,\omega + (1-\lambda L)\, A\tilde{s}_t + (1-\lambda L)\, Bf_t,$$

which implies

$$\begin{aligned} f_{t+1} &= (1-\lambda)\omega + (\lambda I_2 + B)f_t - \lambda B f_{t-1} + (1-\lambda)As_t(x_t, f_t) \\ &= \varphi(f_t, f_{t-1}; \theta) \equiv \varphi_{t+1} \in \mathbb{R}^{2\times 1}. \end{aligned}$$

We assume that $\hat{\theta}-\theta_0 \tilde{\sim} \mathrm{N}(0, W)$, where W is the asymptotic covariance matrix associated with $\hat{\theta}$. A first-order Taylor series expansion around θ_0 yields

$$\begin{aligned}
\hat{f}_{t+1} - f_{t+1} &\approx \frac{\partial \varphi_{t+1}}{\partial \theta_0'} \times \left(\hat{\theta}-\theta_0\right) + \frac{\partial \varphi_{t+1}}{\partial f_t'} \cdot \frac{df_t}{d\theta_0'} \times \left(\hat{\theta}-\theta_0\right) + \frac{\partial \varphi_{t+1}}{\partial f_{t-1}'} \cdot \frac{df_{t-1}}{d\theta_0'} \times \left(\hat{\theta}-\theta_0\right) \\
&= \left[\frac{\partial \varphi_{t+1}}{\partial \theta_0'} + \frac{\partial \varphi_{t+1}}{\partial f_t'} \cdot \frac{df_t}{d\theta_0'} + \frac{\partial \varphi_{t+1}}{\partial f_{t-1}'} \cdot \frac{df_{t-1}}{d\theta_0'}\right] \times \left(\hat{\theta}-\theta_0\right) \\
&= G_{t+1} \times \left(\hat{\theta}-\theta_0\right) \qquad \text{(C.2)} \\
&\sim \mathrm{N}\left(0, G_{t+1} W G_{t+1}'\right),
\end{aligned}$$

where we defined

$$\frac{df_{t+1}}{d\theta'} = G_{t+1} = \frac{\partial \varphi_{t+1}}{\partial \theta'} + \frac{\partial \varphi_{t+1}}{\partial f_t'} \cdot \frac{df_t}{d\theta'} + \frac{\partial \varphi_{t+1}}{\partial f_{t-1}'} \cdot \frac{df_{t-1}}{d\theta'}. \qquad \text{(C.3)}$$

Interestingly, (C.3) is a recursion in G_{t+1} for given $\frac{\partial \varphi_{t+1}}{\partial \theta'}$, $\frac{\partial \varphi_{t+1}}{\partial f_t'}$, $\frac{\partial \varphi_{t+1}}{\partial f_{t-1}'}$. Put differently, (C.3) can be written as

$$G_{t+1} = \frac{\partial \varphi_{t+1}}{\partial \theta'} + \frac{\partial \varphi_{t+1}}{\partial f_t'} \cdot G_t + \frac{\partial \varphi_{t+1}}{\partial f_{t-1}'} \cdot G_{t-1}, \qquad \text{(C.4)}$$

which can be computed in parallel to the recursion for f_t itself. We set $G_1 = G_2 = 0 \in \mathbb{R}^{2\times 7}$ (or to other sensible values).

The derivative terms in recursion (C.4) can be derived as

$$\begin{aligned}
\frac{\partial \varphi_{t+1}}{\partial \theta'} &= (1-\lambda)\frac{\partial \omega}{\partial \theta'} - \omega \frac{\partial \lambda}{\partial \theta'} \\
&\quad + \begin{bmatrix} (1-\lambda) s_t^{\xi} \frac{\partial a^{\xi}}{\partial \theta'} \\ (1-\lambda) s_t^{\delta} \frac{\partial a^{\delta}}{\partial \theta'} \end{bmatrix} - A s_t \frac{\partial \lambda}{\partial \theta'} + (1-\lambda) A \frac{\partial s_t}{\partial \theta'} \\
&\quad + f_t \frac{\partial \lambda}{\partial \theta'} + \begin{bmatrix} f_t^{\xi} \cdot \frac{\partial b^{\xi}}{\partial \theta'} \\ f_t^{\delta} \cdot \frac{\partial b^{\delta}}{\partial \theta'} \end{bmatrix} - B f_{t-1} \frac{\partial \lambda}{\partial \theta'} - \lambda \begin{bmatrix} f_{t-1}^{\xi} \cdot \frac{\partial b^{\xi}}{\partial \theta'} \\ f_{t-1}^{\delta} \cdot \frac{\partial b^{\delta}}{\partial \theta'} \end{bmatrix}, \qquad \text{(C.5)} \\
\frac{\partial \varphi_{t+1}}{\partial f_t'} &= \lambda I_2 + B + (1-\lambda) A \frac{\partial s_t}{\partial f_t'}, \qquad \text{(C.6)} \\
\frac{\partial \varphi_{t+1}}{\partial f_{t-1}'} &= -\lambda B,
\end{aligned}$$

where $\frac{\partial s_t}{\partial \theta'} = 0$ (see (7)).

The derivative terms needed in (C.5) are

$$\begin{aligned}
\frac{\partial \omega^{\xi}}{\partial \theta'} &= \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \\
\frac{\partial \omega^{\delta}}{\partial \theta'} &= \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \\
\frac{\partial a^{\xi}}{\partial \theta'} &= \begin{bmatrix} 0 & 0 & \exp(\alpha^{\xi}) & 0 & 0 & 0 & 0 \end{bmatrix} \\
\frac{\partial a^{\delta}}{\partial \theta'} &= \begin{bmatrix} 0 & 0 & 0 & \exp(\alpha^{\delta}) & 0 & 0 & 0 \end{bmatrix} \\
\frac{\partial b^{\xi}}{\partial \theta'} &= \begin{bmatrix} 0 & 0 & 0 & 0 & \Lambda(\beta^{\xi})[1-\Lambda(\beta^{\xi})] & 0 & 0 \end{bmatrix} \\
\frac{\partial b^{\delta}}{\partial \theta'} &= \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & \Lambda(\beta^{\delta})[1-\Lambda(\beta^{\delta})] & 0 \end{bmatrix} \\
\frac{\partial \lambda}{\partial \theta'} &= \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & \Lambda(\lambda^{uc})[1-\Lambda(\lambda^{uc})] \end{bmatrix},
\end{aligned}$$

where $\Lambda(x) = (1+\exp(-x))^{-1}$ remains the logistic function. Finally, the expression $\frac{\partial s_t}{\partial f_t'}$ in (C.6) can be derived as $\frac{\partial s_t}{\partial f_t'} = \frac{\partial s_t}{\partial (\xi_t, \delta_t)'} \cdot \frac{\partial (\xi_t, \delta_t)}{\partial f_t'} = \frac{\partial s_t}{\partial (\xi_t, \delta_t)'} \cdot \operatorname{diag}(\xi_t, \delta_t)$, where

$$\begin{aligned}
\frac{\partial s_t^{\xi}}{\partial \xi_t} &= \frac{\ln\left(\frac{x_t \xi_t}{\delta_t}+1\right)}{\xi_t^2} + \frac{x_t\left(\frac{1}{\xi_t^2}-1\right)}{\delta_t + x_t\xi_t} - \frac{x_t\left(\delta_t - x_t\left(\xi_t + \frac{1}{\xi_t}+3\right)\right)}{(\delta_t + x_t\xi_t)^2} \\
&\quad - \frac{2\ln\left(\frac{x_t\xi_t}{\delta_t}+1\right)(\xi_t+1)}{\xi_t^3} + \frac{x_t(\xi_t+1)}{\delta_t \xi_t^2\left(\frac{x_t\xi_t}{\delta_t}+1\right)}, \\
\frac{\partial s_t^{\xi}}{\partial \delta_t} &= \frac{x_t(\xi_t+1)(2\delta_t - x_t)}{\delta_t(\delta_t + x_t\xi_t)^2}, \\
\frac{\partial s_t^{\delta}}{\partial \xi_t} &= \frac{(\delta_t - x_t)(x_t - \delta_t + x_t\xi_t)}{(\delta_t + x_t\xi_t)^2\sqrt{2\xi_t+1}}, \\
\frac{\partial s_t^{\delta}}{\partial \delta_t} &= -\frac{x_t\sqrt{2\xi_t+1}(\xi_t+1)}{(\delta_t + x_t\xi_t)^2}.
\end{aligned}$$

The factor variance is given by $V_{t+1} = \operatorname{Var}\left(\hat{f}_{t+1}|x_t, f_t, \theta\right) = G_{t+1} W G_{t+1}'$, evaluated at $\theta = \hat{\theta}$. In a standard fashion we obtain a asymptotic 95% confidence interval for $\hat{f}_{k,t+1}$ as

$$\left[\hat{f}_{k,t+1} - 1.96\sqrt{V_{kk,t+1}}, \hat{f}_{k,t+1} + 1.96\sqrt{V_{kk,t+1}}\right],$$

where $k = 1, 2$ indexes the respective element of $\hat{f}_{t+1}$ and matrix V_{t+1}. Asymmetric confidence bands for $(\hat{\xi}_t, \hat{\delta}_t)' = \exp(\hat{f}_t)$ can be obtained from the confidence bands for $\hat{f}_t$ by exponentiation.

D Scaled scores for the location-scale-df model

This appendix derives the scaled scores of the univariate location–scale–df model presented in Section 2.4. We first recall the non-standardized Student's t log-density as

$$\ln g(y_t; \mu_t, \sigma_t, \nu_t) = c(\nu_t) - \ln \sigma_t - \frac{\nu_t + 1}{2} \ln \left[1 + \frac{1}{\nu_t \sigma_t^2} (y_t - \mu_t)^2 \right],$$

where $c(\nu_t) = \ln \Gamma\left(\frac{\nu_t+1}{2}\right) - \frac{1}{2} \ln (\nu_t \pi) - \ln \Gamma\left(\frac{\nu_t}{2}\right)$, $\sigma_t > 0$ and $\nu_t > 0$.

D.1 Scaled score for the location parameter

The score is obtained as

$$\begin{aligned} \nabla_t^\mu &= \frac{\partial \ln g(y_t; \mu_t, \sigma_t, \nu_t)}{\partial \mu_t} \\ &= \frac{(\nu_t + 1)(y_t - \mu_t)}{\nu_t \sigma_t^2 + (y_t - \mu_t)^2}. \end{aligned}$$

The associated Hessian is given by

$$\begin{aligned} H_t^\mu &= \frac{\partial^2 \ln g(y_t; \mu_t, \sigma_t, \nu_t)}{(\partial \mu_t)^2} \\ &= \frac{(\nu_t + 1)\left[(y_t - \mu_t)^2 - \nu_t \sigma_t^2\right]}{\left[(y_t - \mu_t)^2 + \nu_t \sigma_t^2\right]^2}. \end{aligned}$$

Taking the conditional expectation,

$$\begin{aligned} \mathbb{E}[-H_t^\mu \mid \mathcal{F}_{t-1}] &= \int_{-\infty}^{\infty} \frac{-(\nu_t + 1)\left[(y_t - \mu_t)^2 - \sigma_t^2 \nu_t\right]}{\left[(y_t - \mu_t)^2 + \sigma_t^2 \nu_t\right]^2} \cdot g(y_t \mid \mathcal{F}_{t-1}) \mathrm{d}y_t \\ &= \frac{\nu_t + 1}{(\nu_t + 3)\sigma_t^2}, \end{aligned}$$

see Lange et al. (1989, Appendix B). As a result,

$$s_t^\mu = S_t^\mu \nabla_t^\mu = \frac{(\nu_t + 3)(y_t - \mu_t)}{\nu_t + \sigma_t^{-2}(y_t - \mu_t)^2}.$$

D.2 Scaled score for the log-scale parameter

The score is obtained as

$$\begin{aligned}\nabla_t^\sigma &= \frac{\partial \ln g(y_t;\mu_t,\sigma_t,\nu_t)}{\partial f_t^\sigma} = \frac{\partial \ln g(y_t;\mu_t,\sigma_t,\nu_t)}{\partial \sigma_t}\cdot\frac{d\sigma_t}{d\ln\sigma_t}\\ &= \frac{(\nu_t+1)(y_t-\mu_t)^2}{\nu_t\sigma_t^2+(y_t-\mu_t)^2}-1.\end{aligned}$$

The associated Hessian is given by

$$\begin{aligned}H_t^\sigma &= \frac{\partial^2 \ln g(y_t;\mu_t,\sigma_t,\nu_t)}{(\partial f_t^\sigma)^2} = \frac{\partial \nabla_t^\sigma}{\partial \sigma_t}\cdot\frac{d\sigma_t}{d\ln\sigma_t}\\ &= \frac{-2\nu_t(\nu_t+1)\sigma_t^{-2}(y_t-\mu_t)^2}{\left[\nu_t+\sigma_t^{-2}(y_t-\mu_t)^2\right]^2}.\end{aligned}$$

Taking the conditional expectation with the help of the Beta prime distribution,

$$\begin{aligned}\mathbb{E}[-H_t^\sigma \mid \mathcal{F}_{t-1}] &= \int_{-\infty}^{\infty}\frac{2\nu_t(\nu_t+1)\sigma_t^{-2}(y_t-\mu_t)^2}{\left[\nu_t+\sigma_t^{-2}(y_t-\mu_t)^2\right]^2}\cdot g(y_t\mid\mathcal{F}_{t-1})\mathrm{d}y_t\\ &= \int_{-\infty}^{\infty}\frac{2\nu_t(\nu_t+1)\sigma_t^{-2}(y_t-\mu_t)^2}{\left[\nu_t+\sigma_t^{-2}(y_t-\mu_t)^2\right]^2}\cdot\frac{\Gamma\left(\frac{\nu_t+1}{2}\right)}{\Gamma\left(\frac{\nu_t}{2}\right)\sqrt{\nu_t\pi}\sigma_t}\left[1+\frac{(y_t-\mu_t)^2}{\nu_t\sigma_t^2}\right]^{-\frac{\nu_t+1}{2}}\mathrm{d}y_t\\ &= \int_{-\infty}^{\infty}2\nu_t^{-1}(\nu_t+1)\sigma_t^{-2}(y_t-\mu_t)^2\cdot\frac{\Gamma\left(\frac{\nu_t+1}{2}\right)}{\Gamma\left(\frac{\nu_t}{2}\right)\sqrt{\nu_t\pi}\sigma_t}\left[1+\frac{(y_t-\mu_t)^2}{\nu_t\sigma_t^2}\right]^{-\frac{\nu_t+5}{2}}\mathrm{d}y_t\\ &= 2\nu_t^{-1}(\nu_t+1)\sigma_t^{-2}\frac{\Gamma\left(\frac{\nu_t+1}{2}\right)}{\Gamma\left(\frac{\nu_t}{2}\right)\sqrt{\nu_t\pi}\sigma_t}\int_{-\infty}^{\infty}(y_t-\mu_t)^2\cdot\left[1+\frac{(y_t-\mu_t)^2}{\nu_t\sigma_t^2}\right]^{-\frac{\nu_t+5}{2}}\mathrm{d}y_t\\ &= 2\nu_t^{-1}(\nu_t+1)\sigma_t^{-2}\frac{\Gamma\left(\frac{\nu_t+1}{2}\right)}{\Gamma\left(\frac{\nu_t}{2}\right)\sqrt{\nu_t\pi}\sigma_t}\left(\nu_t\sigma_t^2\right)^{\frac{3}{2}}B\left(\frac{3}{2},\frac{\nu_t+2}{2}\right)\\ &= 2(\nu_t+1)B\left(\frac{3}{2},\frac{\nu_t+2}{2}\right)B\left(\frac{\nu_t}{2},\frac{1}{2}\right)^{-1}\\ &= 2(\nu_t+1)\frac{\nu_t}{(\nu_t+1)(\nu_t+3)}\\ &= \frac{2\nu_t}{\nu_t+3}.\end{aligned}$$

As a result,

$$s_t^\sigma = S_t^\sigma\nabla_t^\sigma = \frac{\nu_t+3}{2\nu_t}\cdot\left(\frac{(\nu_t+1)(y_t-\mu_t)^2}{\nu_t\sigma_t^2+(y_t-\mu_t)^2}-1\right).$$

D.3 Scaled score for the degrees-of-freedom parameter

The score is obtained as

$$\nabla_t^{\nu} = \frac{\partial \ln g(y_t; \mu_t, \sigma_t, \nu_t)}{\partial f_t^{\nu}} = \frac{\partial \ln g(y_t; \mu_t, \sigma_t, \nu_t)}{\partial \nu_t} \cdot \frac{d\nu_t}{d \ln \nu_t}$$

To simplify the derivation, we have

$$\begin{aligned}\frac{\partial \ln g(y_t; \mu_t, \sigma_t, \nu_t)}{\partial \nu_t} = & \frac{1}{2}\gamma'\left(\frac{\nu_t+1}{2}\right) - \frac{1}{2}\gamma'\left(\frac{\nu_t}{2}\right) - \frac{1}{2\nu_t} \\ & - \frac{1}{2}\ln\left[1 + \frac{(y_t-\mu_t)^2}{\nu_t\sigma_t^2}\right] + \frac{\nu_t+1}{2} \cdot \frac{\nu_t^{-2}\sigma_t^{-2}(y_t-\mu_t)^2}{1+\nu_t^{-1}\sigma_t^{-2}(y_t-\mu_t)^2},\end{aligned}$$

where $\gamma'(x) = \frac{\mathrm{d}}{\mathrm{d}x}\ln\Gamma(x) = \frac{\Gamma'(x)}{\Gamma(x)}$ is the polygamma function of order 0. Similarly, we will use $\gamma''(x)$ as the notation for the polygamma function of order 1.

We take the further derivatives and obtain

$$\begin{aligned}\frac{\partial^2 \ln g(y_t; \mu_t, \sigma_t, \nu_t)}{(\partial \nu_t)^2} = & \frac{1}{4}\gamma''\left(\frac{\nu_t+1}{2}\right) - \frac{1}{4}\gamma''\left(\frac{\nu_t}{2}\right) + \frac{1}{2\nu_t^2} + \frac{\nu_t^{-1}(y_t-\mu_t)^2}{\nu_t\sigma_t^2+(y_t-\mu_t)^2} \\ & - \frac{\nu_t+1}{2} \cdot \frac{2\nu_t^{-3}\sigma_t^{-2}(y_t-\mu_t)^2 + \nu_t^{-4}\sigma_t^{-4}(y_t-\mu_t)^4}{\left[1+\nu_t^{-1}\sigma_t^{-2}(y_t-\mu_t)^2\right]^2}.\end{aligned}$$

From the equation $\mathbb{E}(\nabla_t^{\sigma}) = 0$ [1], we know

$$\begin{aligned}\mathbb{E}\left[\frac{(\nu_t+1)(y_t-\mu_t)^2}{\nu_t\sigma_t^2+(y_t-\mu_t)^2}\right] &= 1; \\ \mathbb{E}\left[\frac{\nu_t^{-1}(y_t-\mu_t)^2}{\nu_t\sigma_t^2+(y_t-\mu_t)^2}\right] &= \frac{1}{\nu_t(\nu_t+1)}.\end{aligned}$$

So we can compute the expectation

$$\begin{aligned}\mathbb{E}\left[\frac{\partial^2 \ln g(y_t; \mu_t, \sigma_t, \nu_t)}{(\partial \nu_t)^2}\right] = & \frac{1}{4}\gamma''\left(\frac{\nu_t+1}{2}\right) - \frac{1}{4}\gamma''\left(\frac{\nu_t}{2}\right) + \frac{1}{2\nu_t^2} + \frac{1}{\nu_t(\nu_t+1)} \\ & - \mathbb{E}\left[\frac{\nu_t+1}{2} \cdot \frac{2\nu_t^{-3}\sigma_t^{-2}(y_t-\mu_t)^2 + \nu_t^{-4}\sigma_t^{-4}(y_t-\mu_t)^4}{\left[1+\nu_t^{-1}\sigma_t^{-2}(y_t-\mu_t)^2\right]^2}\right].\end{aligned}$$

We employ a change of variable $\tilde{y}_t = \frac{y_t-\mu_t}{\sigma_t}$, and use the properties of the Beta prime distribution

[1] We simplify the expression $\mathbb{E}(\cdot \mid \mathcal{F}_{t-1})$ as $\mathbb{E}(\cdot)$ in this subsection.

in the following derivations.

$$
\begin{aligned}
&- \mathbb{E}\left[\frac{\nu_t+1}{2}\cdot\frac{2\nu_t^{-3}\sigma_t^{-2}(y_t-\mu_t)^2+\nu_t^{-4}\sigma_t^{-4}(y_t-\mu_t)^4}{\left[1+\nu_t^{-1}\sigma_t^{-2}(y_t-\mu_t)^2\right]^2}\right]\\
&= -\frac{\nu_t+1}{2}\int_{-\infty}^{\infty}\frac{2\nu_t^{-3}\sigma_t^{-2}(y_t-\mu_t)^2+\nu_t^{-4}\sigma_t^{-4}(y_t-\mu_t)^4}{\left[1+\nu_t^{-1}\sigma_t^{-2}(y_t-\mu_t)^2\right]^2}\cdot g(y_t;\mu_t,\sigma_t,\nu_t)\mathrm{d}y_t\\
&= -\frac{\nu_t+1}{2}\int_{-\infty}^{\infty}\left(2\frac{(y_t-\mu_t)^2}{\nu_t^3\sigma_t^2}+\frac{(y_t-\mu_t)^4}{\nu_t^4\sigma_t^4}\right)\cdot\frac{\Gamma\left(\frac{\nu_t+1}{2}\right)}{\Gamma\left(\frac{\nu_t}{2}\right)\sqrt{\nu_t\pi}\sigma_t}\left[1+\frac{(y_t-\mu_t)^2}{\nu_t\sigma_t^2}\right]^{-\frac{\nu_t+5}{2}}\mathrm{d}y_t\\
&= -\frac{\nu_t+1}{2}\int_{-\infty}^{\infty}\left(2\frac{\tilde{y}_t^2}{\nu_t^3}+\frac{\tilde{y}_t^4}{\nu_t^4}\right)\cdot\frac{\Gamma\left(\frac{\nu_t+1}{2}\right)}{\Gamma\left(\frac{\nu_t}{2}\right)\sqrt{\nu_t\pi}}\left[1+\frac{\tilde{y}_t^2}{\nu_t}\right]^{-\frac{\nu_t+5}{2}}\mathrm{d}\tilde{y}_t.\\
&= -\frac{\nu_t+1}{\sqrt{\nu_t}}B\left(\frac{1}{2},\frac{\nu_t}{2}\right)^{-1}\int_0^{\infty}\left(2\frac{\tilde{y}_t^2}{\nu_t^3}+\frac{\tilde{y}_t^4}{\nu_t^4}\right)\cdot\left[1+\frac{\tilde{y}_t^2}{\nu_t}\right]^{-\frac{\nu_t+5}{2}}\mathrm{d}\tilde{y}_t.\\
&= -\frac{\nu_t+1}{\sqrt{\nu}}B\left(\frac{1}{2},\frac{\nu_t}{2}\right)^{-1}\left[\frac{2}{\nu_t^3}\frac{\nu_t^{3/2}}{2}B\left(\frac{3}{2},\frac{\nu_t+2}{2}\right)+\frac{1}{\nu_t^4}\frac{\nu_t^{5/2}}{2}B\left(\frac{5}{2},\frac{\nu_t}{2}\right)\right]\\
&= -\frac{\nu_t+1}{\nu^2}\left[\frac{B\left(\frac{3}{2},\frac{\nu_t+2}{2}\right)}{B\left(\frac{1}{2},\frac{\nu_t}{2}\right)}+\frac{1}{2}\frac{B\left(\frac{5}{2},\frac{\nu_t}{2}\right)}{B\left(\frac{1}{2},\frac{\nu_t}{2}\right)}\right]\\
&= -\frac{2\nu_t+3}{2\nu_t^2(\nu_t+3)}
\end{aligned}
$$

As a result

$$
\begin{aligned}
\mathbb{E}\left[\frac{\partial^2\ln g(y_t;\mu_t,\sigma_t,\nu_t)}{(\partial\nu_t)^2}\right] &= \frac{1}{4}\gamma''\left(\frac{\nu_t+1}{2}\right)-\frac{1}{4}\gamma''\left(\frac{\nu_t}{2}\right)+\frac{1}{2\nu_t^2}+\frac{1}{\nu_t(\nu_t+1)}-\frac{2\nu_t+3}{2\nu_t^2(\nu_t+3)}\\
&= \frac{1}{4}\gamma''\left(\frac{\nu_t+1}{2}\right)-\frac{1}{4}\gamma''\left(\frac{\nu_t}{2}\right)+\frac{1}{2}\frac{\nu_t+5}{\nu_t(\nu_t+1)(\nu_t+3)}.
\end{aligned}
$$

The scaled score is now given by

$$
s_t^{\nu}=S_t^{\nu}\nabla_t^{\nu}=-\left(\frac{d\nu_t}{d\ln\nu_t}\right)^{-1}\cdot\left(\mathbb{E}\left[\frac{\partial^2\ln g(y_t;\mu_t,\sigma_t,\nu_t)}{(\partial\nu_t)^2}\right]\right)^{-1}\cdot\frac{\partial\ln g(y_t;\mu_t,\sigma_t,\nu_t)}{\partial\nu_t}.
$$

E Derivation of EVT-based market risk measures

This section derives the conditional risk measures (20) – (21).

We recall that $\epsilon_t = (y_t - \mu_t)/\sigma_t$, $x_t = \epsilon_t - \tau_t$, and the notation used in Section 2. Note first that

$$\overline{G}(e_t) = 1 - G(e_t) = \mathbb{P}(\epsilon_t > e_t) = \mathbb{P}(\epsilon_t > \tau_t)\mathbb{P}(\epsilon_t > e_t | \epsilon_t > \tau_t) = \overline{G}(\tau_t)\overline{F}(x_t),$$

where the third equality sign uses a standard conditioning argument. We can use this result to derive $\text{VaR}^\gamma(\epsilon_t \mid \mathcal{F}_{t-1}, \theta) = q_t^\gamma(\epsilon_t)$, setting

$$\begin{aligned}
&\overline{G}(e_t) = \overline{G}(\tau_t)\overline{F}(x_t) = 1 - \gamma \\
\iff &\frac{t^*}{t}(1 + \xi_t \delta_t^{-1} x_t)^{-\frac{1}{\xi_t}} = 1 - \gamma \\
\iff &(1 + \xi_t \delta_t^{-1}(q_t^\gamma(\epsilon_t) - \tau_t)) = \left(\frac{1-\gamma}{t^*/t}\right)^{-\xi_t} \\
\iff &q_t^\gamma(\epsilon_t) = \tau_t + \delta_t \xi_t^{-1}\left[\left(\frac{1-\gamma}{t^*/t}\right)^{-\xi_t} - 1\right],
\end{aligned}$$

where t^*/t serves as an estimator of $\overline{G}(\tau_t)$. Since $y_t = \mu_t + \sigma_t \epsilon_t$,

$$\text{VaR}^\gamma(y_t) = \mu_t + \sigma_t \text{VaR}^\gamma(\epsilon_t) = \mu_t + \sigma_t \tau_t + \sigma_t \delta_t \xi_t^{-1}\left[\left(\frac{1-\gamma}{t*/t}\right)^{-\xi_t} - 1\right],$$

which coincides with (20).

The Expected Shortfall $\text{ES}^\gamma(\epsilon_t)$ is given by

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

and is derived by moving constant terms in front of the integral and noting that

$$\int_\gamma^1 (1-s)^{-\xi_t}\mathrm{d}s = \frac{(1-\gamma)^{1-\xi_t}}{1-\xi_t}$$

for $\xi_t < 1$.

The Expected Shortfall (21) is given by

$$\begin{aligned}
\mathrm{ES}^{\gamma}(y_t) &= \frac{1}{1-\gamma}\int_{\gamma}^{1} q_t^s(y_t)\mathrm{d}s \\
&= \frac{1}{1-\gamma}\int_{\gamma}^{1} \mu_t + \sigma_t q_t^s(\epsilon_t)\mathrm{d}s \\
&= \mu_t + \sigma_t \frac{1}{1-\gamma}\int_{\gamma}^{1} q_t^s(\epsilon_t)\mathrm{d}s \\
&= \mu_t + \sigma_t \mathrm{ES}^{\gamma}(\epsilon_t) \\
&= \mu_t + \sigma_t \frac{\mathrm{VaR}^{\gamma}(y_t \mid \mathcal{F}_{t-1}, \theta)}{1-\xi_t} + \sigma_t \frac{\delta_t - \xi_t \tau}{1-\xi_t}.
\end{aligned}$$

F Additional figures

F.1 Simulation outcomes

Figures F.1 and F.2 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 true values. The simulated data are obtained from the GPD and Student's t distributions, respectively.

Our EVT-based market risk measures, such as VaR and ES at a high confidence level $\gamma = 0.99$, tend to be estimated sufficiently accurately when used in combination with our modeling approach. The low and high frequency dynamics of the VaR and ES are both captured well. There only appears some under-estimation of the ES at its very peak where tails are extremely fat. Overall, the model captures well the dynamics of the tails, even if the model does not coincide with the data generating process.

Figure F.1: Simulation results for GPD parameters and risk measures

Rows refer to the different parameter paths (1) – (4) for GPD-distributed observations y_t. Columns refer to filtered estimates of ξ_t, δ_t, VaR_t, and ES_t, respectively. Pseudo-true values are reported in solid red. Median filtered values are reported in solid black. The first two columns also indicate the lower 5% and upper 95% quantiles of filtered tail shape and tail scale estimates. All panels use the time-varying empirical quantile $\hat{\tau}_t$ as the tail threshold.

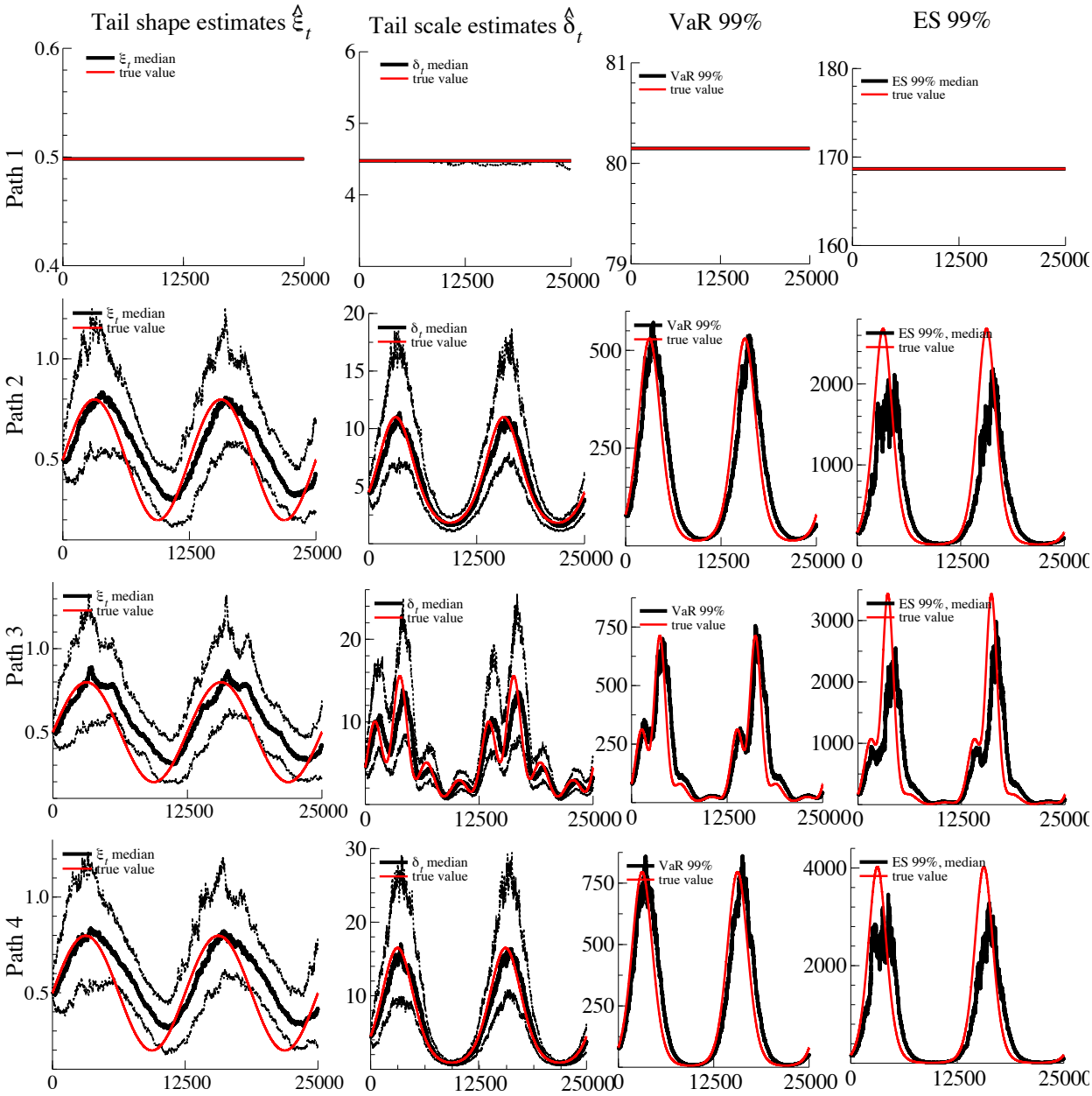
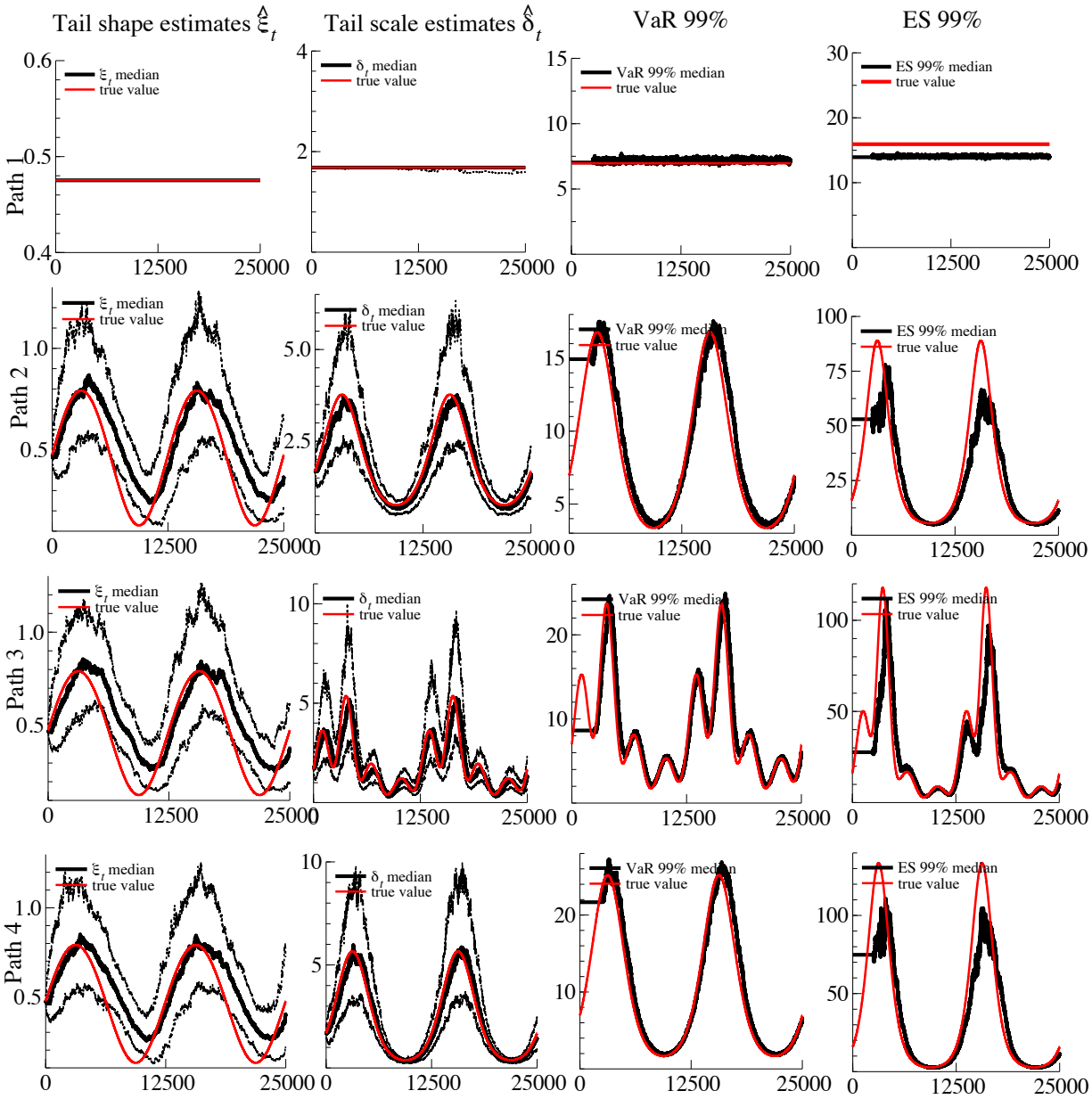


Figure F.2: Simulation results for Student's t distribution

Rows refer to the different parameter paths (1) – (4) for Student's t-distributed observations y_t. Columns report filtered estimates of ξ_t, δ_t, VaR_t, and ES_t, respectively. Pseudo-true values are reported in solid red. Median filtered values are reported in solid black. The first two columns also indicate the lower 5% and upper 95% quantiles of filtered estimates. All panels use the time-varying empirical quantile $\hat{\tau}_t$ as the tail threshold.

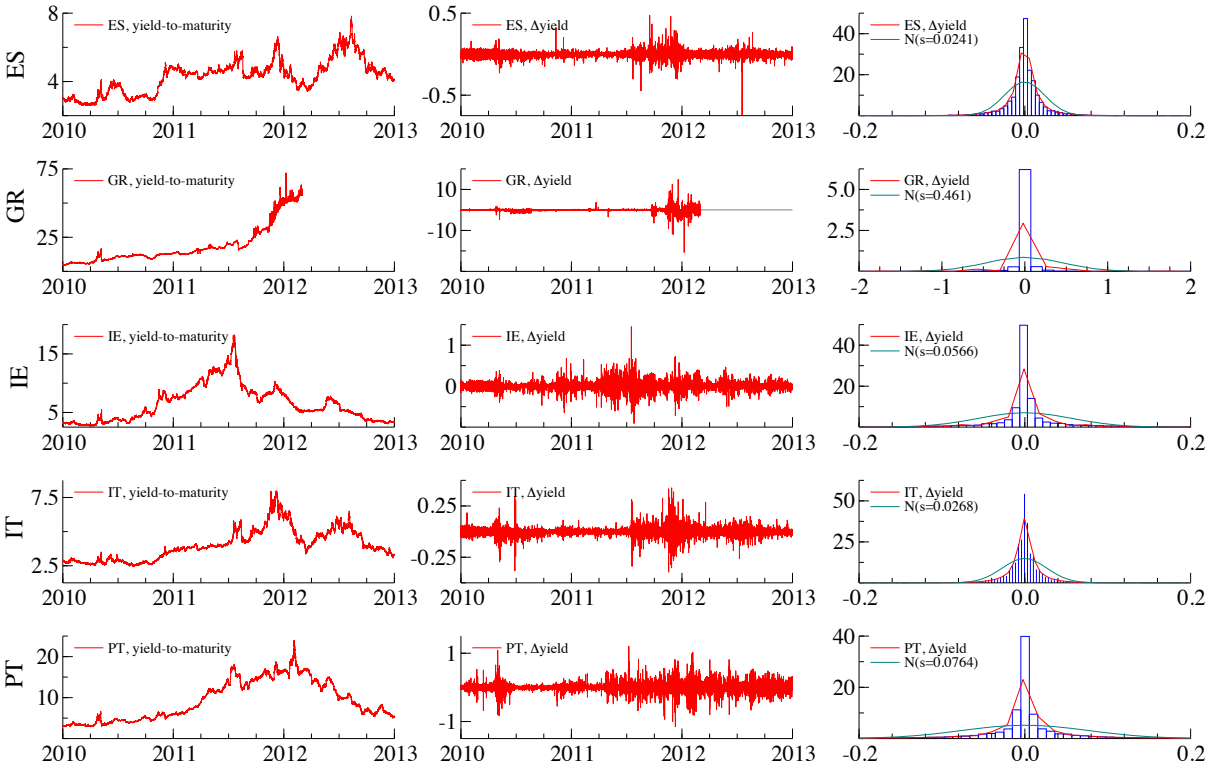


F.2 Plot of sovereign bond yields

Figure F.3 plots the yield-to-maturity of our five benchmark bond yields in levels (left panel) and in first differences (middle panels). A non-parametric Epanechnikov kernel and Gaussian density estimate are fitted to the histogram of yield changes over time (right panels). All five yields exhibited large and sudden moves, leading to volatility clustering and extreme realizations of yield changes during the euro area sovereign debt crisis. The Gaussian approximation in the right panels is a visibly bad fit.

Figure F.3: Five-year sovereign benchmark bond yields

Five-year sovereign benchmark bond yields in levels (left column), in first differences (middle column), and as a histogram with a non-parametric Epanechnikov kernel and Gaussian density estimate fitted to first differences (right column). Rows refer to Spanish (EN), Greek (GR), Irish (IE), Italian (IT), and Portuguese (PT) bond yields. Yields and yield changes are in percentage points and are sampled at 15-minute intervals. Greek benchmark bonds discontinued trading after 02 March 2012, and experienced a credit event on 09 March 2012.



G Tail shape and scale estimates from unfiltered data

This Web Appendix presents our tail shape and scale from unfiltered bond yield data. This means that data y_t are not pre-filtered to remove a time-varying location and scale. Instead, we rely on the time-variation in thresholds τ_t to accommodate these effects. The thresholds evolve over time according to (10), and are initialized at the 90% empirical quantile of the respective full sample of yield changes.

Table G.1 presents the parameter estimates associated with the tail shape model. Not surprisingly, both τ_t and δ_t need to adapt more vigorously to reflect more pronounced variation in y_t. Both a^τ and a^δ have increased compared to their values in Table 4. The SMP impact estimates c^δ are negative in all five cases, and significantly negative in two cases (EN, IT). This suggests that δ_t now proxies for σ_t to some extent, and that the threshold (10) does not entirely succeed in absorbing all variation from the center of the distribution for our data at hand. The c^ξ estimates are negative in two out of five cases, and not statistically significant according to their t-values. The log-likelihood values reported in Table G.1 are lower than those reported in Table 4 in three out of five cases. A meaningful comparison of these log-likelihood values is difficult, however, as each log-likelihood refers to a different set of POTs x_t.

Figure G.1 plots the filtered estimates of ξ_t and δ_t associated with deterministic parameter estimates as reported in Table G.1. Both tail shape and tail scale vary strongly over time, and more strongly than the estimates from the prefiltered (devolatized) data. The time-variation in δ_t is particularly pronounced as it needs to accommodate the pronounced time-variation in σ_t. The two estimates of ξ_t are visibly different.

Figure G.2 presents VaR estimates at a 97.5% (left column) and 99.5% (right column) confidence level. The VaR estimates are obtained from pre-filtered data using (20) evaluated at filtered estimates of μ_t, σ_t, ξ_t and δ_t (green lines), or alternatively obtained from raw data using (20), evaluated at $\mu_t = 0$ and $\sigma_t = 1$, in conjunction with (10) and (12) and filtered estimates of ξ_t and δ_t (red lines). Overall, both VaR estimates capture the same salient trends. The difference between the two VaR estimates is most pronounced for Greek bonds before the credit event in March 2012 when $\hat{\xi}_t >> 1$. The VaR estimate from pre-filtered data is less smooth, owing to pronounced variation in the estimated log-scale σ_t.

Table G.1: Parameter estimates for the tail shape model

Parameter estimates for the extended (with SMP purchases z_t) tail shape model. Columns labeled EN, GR, IE, IT, and PT refer to Spanish, Greek, Irish, Italian, and Portuguese five-year benchmark bond yields. The estimation sample ranges from 04 January 2010 to 28 December 2012 for all countries except Greece. Standard error estimates are in round brackets and are constructed from a sandwich covariance matrix. P-values are in square brackets. The estimation sample is not pre-filtered in terms of location or scale.

	EN	GR	IE	IT	PT
α^{ξ}	0.021	0.019	0.069	0.030	0.012
	(0.007)	(0.003)	(0.017)	(0.010)	(0.008)
	[0.005]	[0.000]	[0.000]	[0.003]	[0.151]
α^{δ}	0.040	0.125	0.092	0.067	0.083
	(0.008)	(0.030)	(0.015)	(0.012)	(0.013)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
c^{ξ}	-0.008	0.008	0.271	0.000	-0.022
	(0.021)	(0.011)	(0.212)	(0.007)	(0.022)
	[0.707]	[0.500]	[0.201]	[0.994]	[0.319]
c^{δ}	-0.009	-0.044	-0.277	-0.004	-0.124
	(0.021)	(0.033)	(0.116)	(0.012)	(0.061)
	[0.647]	[0.188]	[0.017]	[0.729]	[0.043]
a^{τ}	0.210	0.995	0.451	0.282	0.871
T	30279	21839	30799	30519	30719
T^*	3017	2230	3073	3044	3071
loglik	-110914.7	-327866.9	-212679.0	-114753.6	-274260.9
AIC	221837.4	655741.8	425366.1	229515.1	548529.8
BIC	221870.7	655773.7	425399.4	229548.5	548563.2

Figure G.1: Tail shape and tail scale estimates from raw data

Estimated tail shape ξ_t and tail shape δ_t parameters for Spanish (EN), Greek (GR), Irish (IE), Italian (IT), and Portuguese (PT) five-year benchmark bond yields between 2010 and 2012. The data were not pre-filtered based on the location–scale–df model introduced in Section 2.4 (red lines). The associated standard error bands are simulated and reported at a 95% confidence level (black dotted lines). The parameter estimates based on prefiltered data are plotted for comparison (green lines).

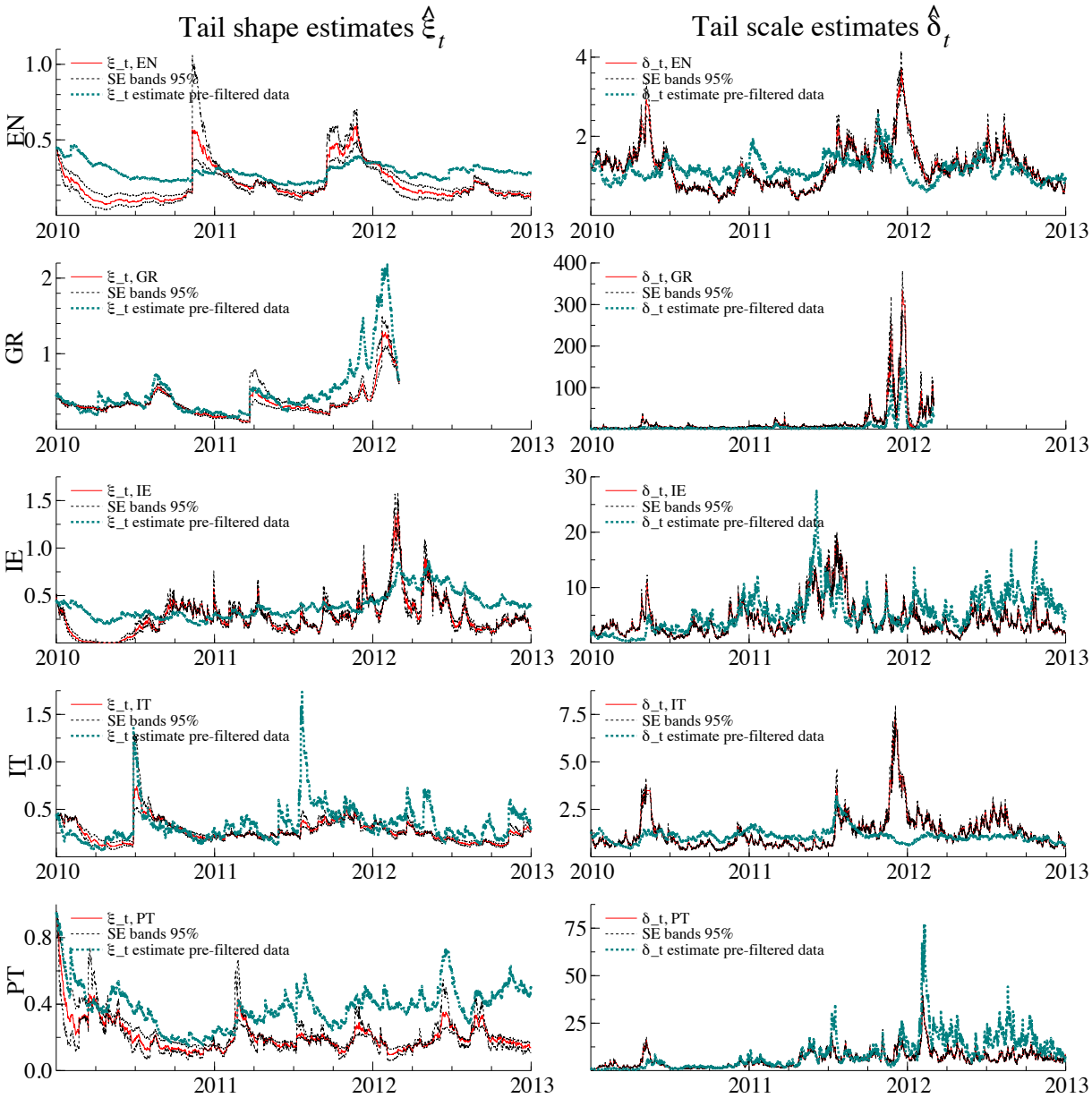
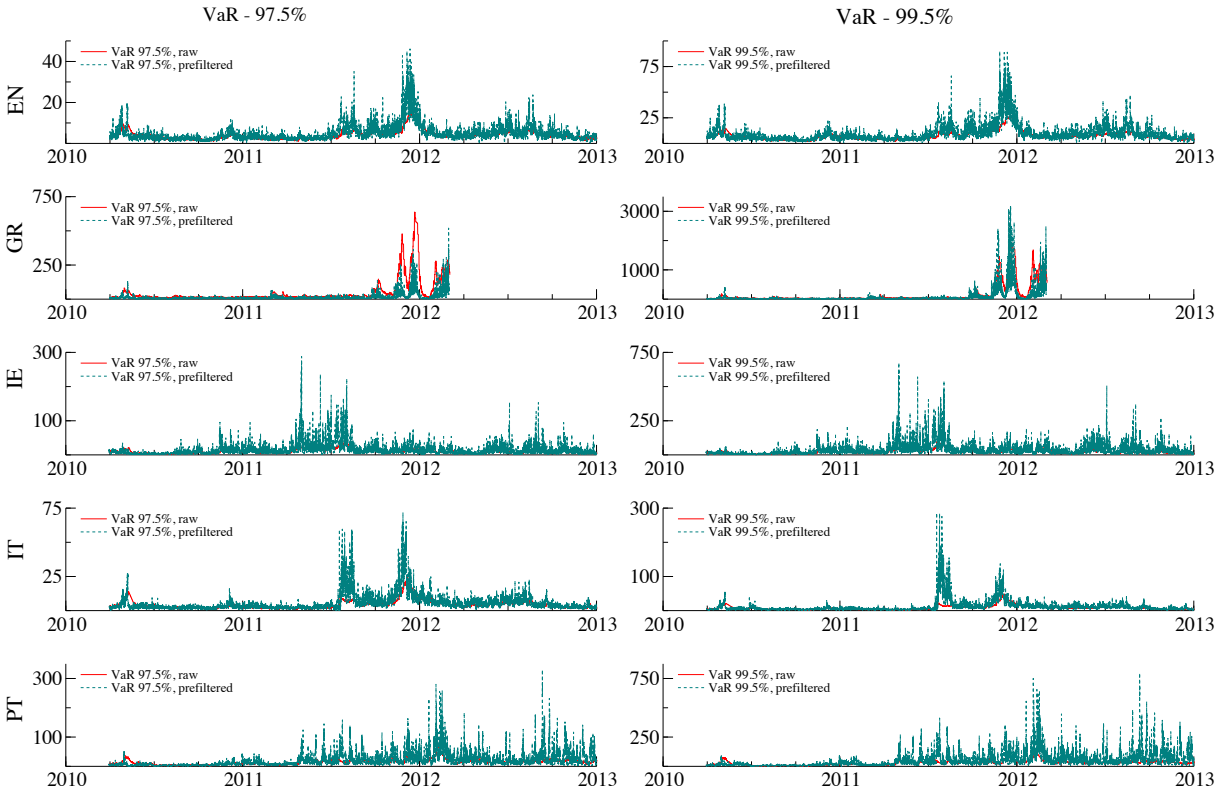


Figure G.2: Comparison of VaR estimates from prefiltered and raw data

Left panel: Value-at-Risk at a 97.5% confidence level for changes in Spanish (EN), Greek (GR), Irish (IE), Italian (IT), and Portuguese (PT) five-year benchmark bond yields between 2010 and 2012. Right panel: Value-at-Risk estimates at a 99.5% confidence level. The VaR estimates are obtained either from pre-filtered data using (20) evaluated at filtered estimates of μ_t, σ_t, ξ_t and δ_t (green lines), or alternatively from raw data using (20) evaluated at $\mu_t = 0$ and $\sigma_t = 1$, in conjunction with (10) and (12) and filtered estimates of ξ_t and δ_t (red lines). Greek bonds discontinued trading after 02 March 2012.



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.

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

Lange, K. L., R. J. A. Little, and J. M. G. Taylor (1989). Robust statistical modeling using the t distribution. *Journal of the American Statistical Association 84*(408), 881–896.