Score-driven tail shape, with application to bond yields at a high frequency*

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

[(a)] VU University Amsterdam, Tinbergen Institute

[(b)] European Central Bank

[(c)] Sveriges Riksbank, Research Division

Abstract

We propose an observation-driven modeling framework for estimating time series variation in the tail shape parameter of a Generalized Pareto Distribution. We discuss different ways of handling non-tail observations and relate tail shape variation to observed covariates. We then use the model to study the yield, volatility, and tail shape impact of bond purchases by the European Central Bank between 2010–2012. Bond purchases lowered the conditional mean of bond yields by about -0.2 to -3.5 bps per 1 bn of purchases. The announcement of the program had a significant impact on the tail shape.

Keywords: 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. We thank Olivier Vergote and the ECB's DG-M for comments and access to high-frequency data on bond purchases. We thank seminar participants at the 2016 CFE conference in Seville, the ECB, the EFA conference in Oslo, the IAAC in Milan, the 2017 EcoSta conference in Hong Kong, 10th SoFiE conference in New York, and Sveriges Riksbank. The views expressed in this paper are those of the author and they do not necessarily reflect the views or policies of the Sveriges Riksbank or European Central Bank.

1 Introduction

We propose a novel observation-driven modeling framework that introduces time series dynamics into the tail shape parameter of the Generalized Pareto Distribution (GPD). The GPD is of considerable interest to financial economists, as it is the only non-degenerate density that approximates the distribution of data exceedances beyond a given threshold; see, for example, Davidson and Smith (1990), Embrechts, Klüppelberg, and Mikosch (1997), and McNeil, Frey, and Embrechts (2010, Chapter 7). As a result, it plays a central role in the study of extremes, comparable to the role the normal distribution plays in the study of observations with finite variance. The proposed framework allows us to track the time variation in the tail index of observations from a wide class of fat-tailed distributions; see Rocco (2014) for a recent survey of extreme value theory (EVT) methods.

In our model, the tail shape dynamics are driven by the score of the predictive log-likelihood. Score-driven (GAS) models are developed in their full generality in Creal, Koopman, and Lucas (2013); see also Harvey (2013). In this setting, time-varying parameters are perfectly predictable one step ahead. This feature makes the new model observation-driven; see Cox (1981). The likelihood is known in closed-form through a standard prediction error decomposition, making parameter estimation straightforward via maximum likelihood procedures. We discuss different ways of treating non-tail observations, and relate score-driven tail shape to observed covariates.

The treatment of non-tail observations is an important issue in the modeling of a time-varying tail shape parameter. Most time series observations are, by definition, not extreme and thus fall into this category. We discuss three different approaches to the treatment of non-tail observations: i) deletion, ii) modeling as missing at random, and iii) as a draw from a mixture density with a point mass at zero. Each of these cases leads to different closed-form expressions for the score and the scaling function. We present these expressions, and discuss the implied news impact curves.

Monte Carlo experiments suggest that our statistical model reliably captures tail shape variation in a variety of settings. Modeling the tail shape dynamics based on a mixture

distribution performs well in general, and particularly so when mean reversion in the time-varying parameter is not pronounced. The deletion of missing values is not efficient, and only appropriate when a complete time series of tail shape estimates is not required. Modeling non-tail observations as missing at random, i.e., without information about the tail, performs well only if mean reversion in the time-varying tail shape parameter is strong.

We apply our score-driven modeling framework to study the yield, volatility, and tail shape impact of bond purchases undertaken by the European Central Bank (ECB) during the euro area sovereign debt crisis between 2010–2012. We focus on bond purchases within the ECB'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 fat-tails, we find that purchases lowered the conditional mean of bond yields by about -0.2 to -3.5 bps per 1 bn of purchases. In addition, the announcement of the program had a significant impact on the tail shape dynamics. This is relevant since elevated tail risks alone can force institutional investors and market makers to retreat from a given market, particularly if value-at-risk constraints are binding; see, for example, Vayanos and Vila (2009), and Adrian and Shin (2010)'. Once announced, however, bond purchases within the SMP did not have an impact on either volatility nor tail shape.

Our paper is related to at least two strands of literature. First, several studies investigate the dynamic behavior of the tail index. For example, Quintos, Fan, and Phillips (2001) derive formal tests for time-variation in the tail index. A number of subsequent studies subsequently applied these 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. We add to this strand of (dynamic) EVT literature by proposing a framework that allows us to estimate the tail shape dynamics directly. Explanatory covariates can be included in the score updating equation, and t-tests as well as likelihood ratio

tests can be used to test economically relevant hypotheses.

A second strand of literature assesses the impact of central bank asset purchases on bond yields and yield volatility. For example, Ghysels, Idier, Manganelli, and Vergote (2016) 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 may have introduced. The authors find that an intervention of €100 million had an immediate impact on bond yields of between -0.1 and -25 bps, depending on the debt market and timing. 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 basis points at the five-year maturity. Pooter, Martin, and Pruitt (2017) 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 this second strand of literature on impact identification by focusing on higher order moments. When doing so, we explicitly accommodate the fat-tails that characterize changes in bond yields during a severe sovereign debt crisis, particularly in stressed markets and at a high-frequency.

On the technical side, our modeling framework is most related to that of Massacci (2017). Massacci proposes a time series model for both the scale and tail shape parameters of the GPD distribution. Both parameters evolve in a bivariate system and are updated jointly. Our model is different in that we assume de-volatized (or 'prefiltered', 'de-GARCHed') data. This approach is common in the applied literature; see e.g. McNeil and Frey (2000), Poon, Rockinger, and Tawn (2004), Brownlees and Engle (2015), Lucas, Schwaab, and Zhang (2014, 2016), and others. If the data are de-volatized in a first step, the scale of the GPD can be treated as fixed at unity, leaving only one time-varying parameter to model. This leads to an obvious simplification of the model, as no bivariate updating of two parameters is required. Modeling volatility and tail shape parameters separately allows for more flexibility in the modeling of either parameter. Finally, our paper is different in that we study alternative ways

of treating non-tail observations, and relate tail shape dynamics to explanatory covariates in a relevant setting.

We proceed as follows. Section 2 introduces the statistical model. Section 3 discusses our Monte Carlo results. Section 4 applies the modeling framework to euro area sovereign bond yields and ECB asset purchases at a high frequency. Section 5 concludes.

2 Statistical model

2.1 Time-varying tail risk

This section introduces time variation into the tail shape parameter $\xi_t > 0$ of the Generalized Pareto Distribution (GPD). The probability density function (pdf) of a GPD distributed random variable $x_t > 0$ is given by

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

where $x_t = y_t - \tau > 0$ is the so-called peak-over-threshold (POT), or exceedance, of fat-tailed data y_t over a pre-determined threshold τ, $\delta > 0$ is an additional scale (variance) parameter, and $\xi_t > 0$ is the tail shape parameter; see, for example, McNeil, Frey, and Embrechts (2010).

Essentially all common continuous distributions of statistics and the actuarial sciences lie in a certain Maximum Domain of Attraction (MDA); see McNeil et al. (2010, Chapter 7.1). When considering $\xi_t > 0 \, \forall \, t$ we implicitly assume that a Fréchet limit (a power tail) applies. Examples of such distributions include the Student's t, inverse gamma, loggamma, F, Fréchet, Burr, and others.

If $y_t - \tau \leq 0$, we may consider x_t as missing, or as a draw from a mixture distribution, see Section 2.2 below. As a result, exceedances $x = (x_1, \ldots, x_T)$ and data $y = (y_1, \ldots, y_T)$, $t = 1, \ldots, T$, are univariate time series at the same frequency. The cumulative distribution function (cdf) and the log-likelihood of x_t is given by

$$P(x_t; \delta, \xi_t) = 1 - \left(1 + \xi_t \frac{x_t}{\delta}\right)^{-\frac{1}{\xi_t}}, \qquad l(x_t; \delta, \xi_t) = -\ln(\delta) - \left(1 + \frac{1}{\xi_t}\right) \ln\left(1 + \xi_t \frac{x_t}{\delta}\right), \tag{2}$$

respectively.

We assume that the pdf of y_t, $g(y_t)$, is a fat-tailed distribution with time-varying tail index $\alpha_t > 0$, such as, for example, a univariate Student's t distribution with $\nu_t = \alpha_t = 1/\xi_t$ degrees of freedom. In this case the cumulative distribution function $G(y_t)$ can be expressed as $G(y_t) = G(\tau) + (1 - G(\tau))P(x_t)$ for sufficiently high values of τ. As a result, all interesting tail behavior is captured by $p(x_t; \delta, \xi_t)$.

The choice of threshold τ is subject to a well-known bias-efficiency tradeoff; see, for instance, McNeil and Frey (2000). In theory, the limiting distribution of exceedances holds exactly only as the threshold $\tau \to +\infty$. A higher threshold, however, also implies a smaller number of exceedances, and consequently an increased sampling error in the estimation of the tail shape parameter. The 10%, 5%, and 1% empirical quantiles are common choices in the literature; see Chavez-Demoulin, Embrechts, and Sardy (2014).

Following Creal, Koopman, and Lucas (2013), and Harvey (2013), we endow ξ_t with score-driven (GAS) dynamics using the derivative of the log conditional observation density (1). We ensure positive values of ξ_t by specifying $\xi_t = \exp(f_t)$. The transition dynamics for f_t are given by

$$f_{t+1} = \omega + \sum_{i=0}^{p-1} a_i s_{t-i} + \sum_{j=0}^{q-1} b_j f_{t-j}, \tag{3}$$
$$s_t = \mathcal{S}_t \nabla_t, \quad \nabla_t = \partial \ln p(x_t | \mathcal{F}_{t-1}; f_t, \psi)/\partial f_t,$$

where $\omega = \omega(\psi)$ is a fixed intercept, $a_i = a_i(\psi)$ and $b_j = b_j(\psi)$ are fixed scalar parameters that depend on the vector ψ containing all time invariant parameters in the model, and $\mathcal{F}_{t-1} = \{x_1, \ldots, x_{t-1}\}$.

For the remainder of the paper, we make three empirical choices. First, we select the inverse conditional Fisher information of the observation density as our scaling function, $\mathcal{S}_t = \mathrm{E}[\nabla_t^2 | \mathcal{F}_{t-1}; f_t, \psi]^{-1} = \mathrm{E}[-\partial \nabla_t (x_t | \mathcal{F}_{t-1}; f_t, \psi)/\partial f_t]^{-1}$. Creal, Koopman, and Lucas (2013) and Creal, Schwaab, Koopman, and Lucas (2014) demonstrate that this choice of scaling function results in a stable model, and effectively yields a Gauss-Newton update of f_t over

time. Second, we consider a fixed $\delta = 1$ throughout the paper. This corresponds to the common practise of using volatility-filtered data before considering tail risk dynamics. If volatility clustering is not accounted for, movements in the tail may be confounded with movements in the conditional variance; see, for example, McNeil and Frey (2000). Finally, we assume $p = q = 1$ in (3), such that $a = a_0$ and $b = b_0$. Higher order terms are rarely necessary in practise; see Creal et al. (2013). To ensure stationarity of the factor process we require $\mid b \mid < 1$, and in addition restrict $a > 0$.

2.2 Three approaches to handling non-tail observations

This section discusses three approaches to accommodating non-tail observations $y_t \leq \tau$. For each approach, we derive the conditional score ∇_t and the scaling function $\mathcal{S}_t$ in (3).

First, we may simply want to delete the missing entries in the univariate time series. This yields a substantially shorter time series for x. In this case, closed-form expressions for the score s_t and the scaling function $\mathcal{S}_t$ can be derived as

$$\nabla_t = \frac{1}{\xi_t} \ln\left(1 + \xi_t \frac{x_t}{\delta}\right) - (\xi_t + 1)\frac{x_t}{\delta + \xi_t x_t}, \tag{4}$$

$$\mathcal{S}_t = \frac{(1 + 2\xi_t)(1 + \xi_t)}{2\xi_t^2}, \tag{5}$$

where, again, $\mathcal{S}_t = \mathrm{E}[-\partial \nabla_t(x_t|\mathcal{F}_{t-1}; f_t, \psi)/\partial f_t]^{-1}$ is the inverse conditional Fisher information quantity for x_t. Appendix A1 provides a derivation of (4) and (5). Deleting non-tail observations before applying (1) – (5) is straightforward, and in line with the spirit of EVT in the absence of time variation in the tail index; see McNeil, Frey, and Embrechts (2010, Chapter 7). One immediate downside of this approach, however, is that ξ_t is unavailable for all $y_t \leq \tau$. In such cases, ξ_t could remain constant at its last filtered value.

Second, we calculate the scaled score (4) – (5) only if x_t is observed, and assign a zero value to the scaled score if x_t is missing. This approach to missing variables is adopted in Creal, Schwaab, Koopman, and Lucas (2014). The updating equation becomes

$$f_{t+1} = \omega + a \cdot \mathrm{I}(x_t > 0)\mathcal{S}_t \nabla_t + b \cdot f_t, \tag{6}$$

where $x_t = y_t - \tau \in \mathbb{R}$. As a result, the score is zero if $y_t < \tau$, and $f_t = \ln(\xi_t)$ slowly reverts back to its unconditional mean $\omega/(1-b)$ in the absence of new information from the tail. The updating equation (6) takes into account that ξ_t evolves as a continuous process which governs the tail behavior of each y_t. An estimate of ξ_t is now available at any time t.

Finally, we use a mixture distribution to model the exceedance and to derive the score; see, for example, Davidson and Smith (1990) and Massacci (2017). In this case, observations $y_t \leq \tau$ generate the POT values $x_t = 0$. This approach allows us to incorporate information in the tail as well as from the center of the distribution. We consider the mixture

$$\phi(x_t; \delta, \xi_t) = \mathrm{I}(x_t > 0) \left[(1+\tau)^{-\frac{1}{\xi_t}} \frac{1}{\delta} \left(1 + \xi_t \frac{x_t}{\delta} \right)^{-\frac{1}{\xi_t}-1} \right] + \mathrm{I}(x_t = 0) \left(1 - (1+\tau)^{-\frac{1}{\xi_t}} \right), \quad (7)$$

where the GPD pdf applies with tail probability $\bar{G}(\tau) = (1+\tau)^{-\frac{1}{\xi_t}}$.[1] The respective log-likelihood is

$$\begin{aligned} \ell(x_t; \delta, \xi_t) &= \mathrm{I}(x_t > 0) \left[-\frac{1}{\xi_t} \ln(1+\tau) - \ln(\delta) - \left(1 + \frac{1}{\xi_t} \right) \ln \left(1 + \xi_t \frac{x_t}{\delta} \right) \right] \\ &+ \mathrm{I}(x_t = 0) \ln(1 - (1+\tau)^{-\frac{1}{\xi_t}}). \end{aligned} \quad (8)$$

The score-driven update of $f_t = \ln \xi_t$ becomes slightly more involved in this case. The score and scaling function are now given by

$$\begin{aligned} \nabla_t &= \mathrm{I}(x_t > 0) \left[\frac{1}{\xi_t} \ln(1+\tau) + \frac{1}{\xi_t} \ln \left(1 + \xi_t \frac{x_t}{\delta} \right) - (\xi_t + 1) \frac{x_t}{\delta + \xi_t x_t} \right] \\ &- \mathrm{I}(x_t = 0) \left[\frac{(1+\tau)^{-\frac{1}{\xi_t}}}{1 - (1+\tau)^{-\frac{1}{\xi_t}}} \frac{1}{\xi_t} \ln(1+\tau) \right], \end{aligned} \quad (9)$$

and

$$\mathcal{S}_t = \left[\frac{2\xi_t^2 (1+\tau)^{-\frac{1}{\xi_t}}}{(1+2\xi_t)(1+\xi_t)} + \frac{(1+\tau)^{-\frac{1}{\xi_t}}}{1 - (1+\tau)^{-\frac{1}{\xi_t}}} \frac{1}{\xi_t^2} \ln(1+\tau)^2 \right]^{-1}, \quad (10)$$

respectively. Appendix A2 derives (9) and (10). Since both expressions are available in

[1] The tail probability is common across fat-tailed distributions $g(y_t)$. If $G(y_t)$ lies in the MDA of the Fréchet limit, see McNeil, Frey, and Embrechts (2010, p. 268)), the survival function of y_t can be written as $\bar{G}(y_t) = y_t^{-1/\xi_t} \cdot L(y_t)$, where $L(y_t)$ is a slowly varying function. Adopting the choice $L(y_t) = (y_t/(1+y_t))^{1/\xi_t}$ yields the tail probability below (7).

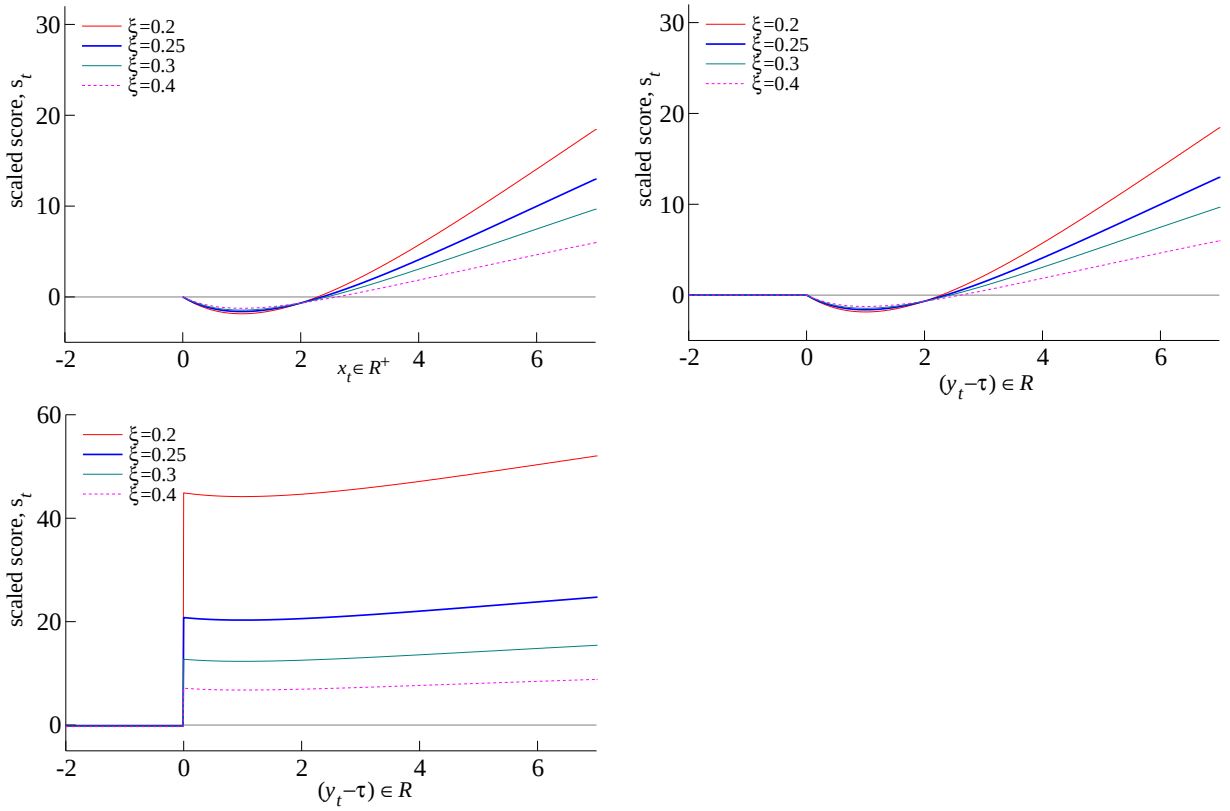

Figure 1: News impact curves

News impact curves for four different values of ξ_t. The curves refer to three different approaches to the modeling of non-tail observations. The top left panel plots the scaled score if missings are deleted; see (4) and (5). The top right panel refers to the case where the score is set to zero for values $y_t < \tau$; see (4) – (6). The bottom left panel refers to a mixture distribution with a point mass at zero; see (9) and (10). The scaled score takes a small negative value for $y_t < \tau$ in the bottom left panel.

closed form, the transition equation for f_t is straightforward.Modeling the tail index as the inverse tail shape $\alpha_t = 1/\xi_t$ in a score-driven way leads to equivalent expressions for the score and the same scaling function.

Figure 1 presents the news impact curves as implied by the three alternative approaches. The top left panel plots the scaled score based on (4) and (5). The top right panel plots the scaled score when the score is set to zero for values $y_t < \tau$; see (4) – (6). The bottom left panel refers to the mixture distribution case with a point mass at zero; see (9) and (10).

The scaled score takes on a global minimum at $\delta = 1$ in panels 1 and 2, and a local minimum at $\delta = 1$ in panel 3. High values of $y_t - \tau$ are associated with upward revisions of $f_t = \ln \xi_t$. Small or negative values are associated with a declining tail shape. This is

intuitive. The scaled score takes zero values for $y_t - \tau < 0$ in the top right panel, and a small negative value in the bottom left panel. The conditional expectation of the score is zero in each case. The fatter the currently prevailing tail, i.e., the higher xi_t, the lower the sensitivity of the scaled score to large values $y_t > \tau$.

2.3 Explanatory covariates

The score-driven dynamics for the tail shape parameter can be extended to include contemporaneous or lagged economic variables as additional conditioning variables. For example, central bank asset purchases and announcement effects may help explain the time-variation in the tail shape of changes in sovereign bond yields. We then consider extended factor dynamics

$$f_{t+1} = \omega + a \cdot \mathcal{S}_t \nabla_t + b \cdot f_t + c' \mathrm{X}_t, \tag{11}$$

where c is a vector of coefficients to be estimated, and all additional variables are collected in X_t.

3 Simulation study

This section presents our Monte Carlo simulation results. We are particularly interested in two questions: whether our statistical framework reliably recovers time variation in the tail shape, and which treatment of non-tail observations is appropriate.

We chose the Student's t distribution $y_t \sim \mathrm{t}(\nu)$ as our fat tailed distribution $g(y_t)$, and consider three values for the degrees of freedom parameter, $\nu = 3, 5, 7$. The threshold τ is the 90% percentile of the random sample generated from a $t(3)$ and $t(5)$ distribution, and the 95% percentile when the sample is generated from the $t(7)$ distribution (for which the tail is less fat). The exceedance sample contains $x_t \sim \mathrm{GPD}(0, \delta, \xi_t)$ if $y_t > \tau$, and zero (missing) otherwise. The GPD distributed random variables are drawn as $x_t = \delta \frac{u_t^{-\xi_t} - 1}{\xi_t}$, where $u_t \sim \mathrm{U}[0, 1]$ is uniform.

Each simulation uses a different path for the tail shape parameter $\xi_t = \exp(f_t)$. We con-

sider seven data generating processes (DGP). The first four processes are stylized dynamics in the literature; see, for example, Lucas and Zhang (2016). The final three processes consider random draws of f_t subject to different parameters ψ.

(1) Constant: $\xi_t = 0.9$,

(2) Sine: $\xi_t = 0.5 + 0.4\cos(2\pi t/200)$,

(3) Fast Sine: $\xi_t = 0.5 + 0.4\cos(2\pi t/20)$,

(4) Step: $\xi_t = 0.9 - 0.5(t > 500)$,

(5) sGPD^d : (delete missings) $f_{t+1} = \omega + a\mathcal{S}_t\nabla_t + bf_t$,

(6) sGPD^z : (assign zero) $f_{t+1} = \omega + a \cdot \mathrm{I}(x_t > 0)\mathcal{S}_t\nabla_t + bf_t$,

(7) sGPD^m : (mixture density): $f_{t+1} = \omega + a\mathcal{S}_t\nabla_t + bf_t$,

The parameters in simulation settings (5) – (7) are chosen in two different ways, as

$$\begin{aligned} \psi_1 \quad &: \quad \omega = -0.025, a = 0.01, b = 0.97, \delta = 1, \\ \text{and} \quad \psi_2 \quad &: \quad \omega = 0.01, a = 0.12, b = 0.85, \delta = 1. \end{aligned}$$

As a result, the tail risk dynamics under parameter vector ψ_1 are more persistent than in the model with ψ_2; note the higher value for b. In addition to exhibiting faster mean reversion, the tails are also fatter under ψ_2, as the unconditional mean of f_t is higher.

We end up with $4+2\times3 = 10$ stochastic GPDs, in three environments ($\nu = 3, 5, 7$). This yields 30 data generating processes. For each DGP, we draw 100 simulation samples of y with $10,000$ observations each. As a result, approximately 1,000 ($t(\nu = 3, 5)$) or 500 ($t(\nu = 7)$) tail observations are used to construct POT values. Our main metric for evaluating model performance is Mean Absolute Error, MAE $= \frac{1}{ST}\sum_{s=1}^{S}\sum_{t=1}^{T}|\hat{\xi}_{st} - \xi_{st}|$, where $\hat{\xi}_{st}$ is the estimated dynamic tail parameter in simulation s, ξ_{st} is the true tail shape, S is the number of simulations, and T is the number of observations in each draw.

Table 1 presents the main MAE simulation outcomes. We focus on three main findings. First, simply deleting missing values (our first approach) seems to be permissible as long as the tail fatness is sufficiently high. When $\nu = 3$, the respective factor estimates are close

Table 1: **Simulation results: Mean Absolute Error**

The table reports mean absolute error (MAE) statistics for 10 DGPs (columns) and three estimation approaches (rows), in three environments (top, middle, and bottom panels). The hit variable $\mathrm{I}(y_t > \tau)$ is simulated from a $t(3)$, $t(5)$, and $t(7)$. We consider 100 simulations for each DGP, and a time series y of $10,000$ observations in each simulation. Different approaches to the treatment of non-tail observations are indicated in the respective rows. sGPD^d deletes missing vales, sGPD^z assigns a zero value to the score at missing values, and sGPD^m uses a mixture density. Model performance is measured by the MAE from the true ξ_t in each draw. The number in bold indicates minimum MAE among the four approaches considered. For DGPs (5)–(8), two sets of parameters ψ_1 and ψ_2 apply.

Model	(1)	(2)	(3)	(4)	(5)-ψ_1	(5)-ψ_2	(6)-ψ_1	(6)-ψ_2	(7)-ψ_1	(7)-ψ_2
					t distributed, $\nu = 3$					
sGPD^d	0.049	**0.202**	0.259	**0.151**	**0.045**	**0.097**	**0.038**	**0.100**	0.044	0.136
	(0.037)	(0.016)	(0.006)	(0.044)	(0.022)	(0.064)	(0.021)	(0.031)	(0.018)	(0.017)
sGPD^z	**0.046**	0.293	0.260	0.274	0.054	0.534	0.040	0.123	**0.043**	**0.136**
	(0.037)	(0.042)	(0.008)	(0.040)	(0.028)	(0.139)	(0.029)	(0.052)	(0.024)	(0.020)
sGPD^m	0.359	0.255	**0.255**	0.249	0.092	0.619	0.090	0.528	0.061	0.300
	(0.008)	(0.001)	(0.001)	(0.002)	(0.013)	(0.065)	(0.008)	(0.010)	(0.006)	(0.006)
					t distributed, $\nu = 5$					
sGPD^d	0.048	**0.201**	0.258	**0.152**	0.043	**0.082**	0.036	**0.101**	0.039	0.108
	(0.034)	(0.015)	(0.006)	(0.046)	(0.022)	(0.037)	(0.021)	(0.032)	(0.019)	(0.015)
sGPD^z	**0.045**	0.297	0.259	0.280	0.057	0.566	0.036	0.117	0.036	**0.106**
	(0.035)	(0.040)	(0.012)	(0.046)	(0.026)	(0.154)	(0.022)	(0.047)	(0.018)	(0.016)
sGPD^m	0.433	0.256	**0.257**	0.249	**0.041**	0.695	**0.021**	0.603	**0.019**	0.313
	(0.007)	(0.001)	(0.001)	(0.002)	(0.005)	(0.064)	(0.005)	(0.008)	(0.003)	(0.005)
					t distributed, $\nu = 7$					
sGPD^d	0.069	**0.209**	0.260	0.066	0.058	**0.142**	0.050	0.097	0.052	0.096
	(0.046)	(0.026)	(0.009)	(0.049)	(0.031)	(0.076)	(0.034)	(0.046)	(0.032)	(0.033)
sGPD^z	**0.062**	0.314	**0.259**	**0.059**	0.060	0.504	0.043	**0.088**	0.045	**0.090**
	(0.044)	(0.113)	(0.007)	(0.046)	(0.030)	(0.221)	(0.032)	(0.047)	(0.029)	(0.029)
sGPD^m	0.477	0.260	0.261	0.477	**0.039**	0.740	**0.022**	0.646	**0.020**	0.344
	(0.006)	(0.001)	(0.001)	(0.006)	(0.010)	(0.090)	(0.005)	(0.010)	(0.002)	(0.005)

to the underlying true processes in many cases. Even if the true tail dynamics come from a mixture-likelihood model, the sGPD^d and sGPD^z approaches still deliver close estimates. Modeling non-tail observations as missing, without information about the tail, works well if mean reversion in the tail shape process is strong (ψ_2), but not necessarily otherwise (ψ_1).

As the degrees of freedom parameter increases and the sample size decreases ($\nu = 7$), the appropriateness of simply deleting missing values in x becomes less clear. In all DGPs under ψ_1, the mixture model sGPD^m now outperforms the other approaches; see columns 6, 8, and 10. This is not surprising. When the tail observations are less frequent and contain less information about the (thinner) tail, then the mixture-density model benefits from being able to take into account information from non-tail observations as well. If the tails are

fatter (ψ_2), the first two approaches are again appropriate.

Second, the standard deviations about the MAE statistics are typically larger in the first two rows than for the mixture-density approach. This difference in standard errors points towards a loss of efficiency that occurs when we discard observations (as in the first approach), or model them by a zero value for the score (the second approach). Again, the mixture-density approach uses non-tail observations to at least some degree, implying less variation across simulations for these estimates.

Finally, the results are overall not too sensitive to which exact model is employed for inference on ξ_t. For example, the mixture density approach is accurate even if the true model is formulated differently. In other words, even (slightly) misspecified score-driven models appear to work well. Blasques, Koopman, and Lucas (2015) prove that score-based parameter updates always reduce the local Kullback-Leibler divergence between the true conditional density and the (potentially misspecified) model-implied conditional density, and are in this sense optimal from an information theoretic perspective. Table 1 suggests that slightly misspecified models can still work well in practise.

4 Empirical application to bond yield changes during the euro area sovereign debt crisis

4.1 The debt crisis and the ECB's SMP

Exceptional times can require exceptional policy measures. Since the onset of the financial crisis in 2007, the major central banks have implemented both standard and non-standard monetary policy measures to contain financial instability and adverse economic outcomes; see Lenza, Pill, and Reichlin (2010) and Eser, Carmona Amaro, Iacobelli, and Rubens (2012). In particular, major central banks have undertaken asset purchase programs. These include the large-scale asset purchases programs (LSAPs) of the Federal Reserve, the Securities Markets Program (SMP) and Public Sector Purchase Programme (PSPP) of the ECB, the quantitative easing (QE) of the Bank of England, and the qualitative and quantitative

monetary easing (QQE) of the Bank of Japan.

Similar to other asset purchase programs, the effectiveness of the ECB Securities Markets Programme (SMP) has been subject to intense academic, public, and policy debate. We study SMP interventions in government debt securities markets between 2010–2011 in five euro area countries: Greece, Ireland, Italy, Portugal, and Spain. Approximately €214 billion (bn) of bonds were acquired between 2010 and early 2012, see ECB (2013).[2] We focus on the following questions: Have SMP asset purchases affected bond yields in secondary debt markets for the respective countries, and if so, to what extent? In addition, have purchases affected the volatility and extreme tail behavior of yield changes?

The SMP was announced on 10 May 2010 and 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. The SMP and the OMTs are related but different programs, see Cœuré (2013).

The SMP had the objective of helping to restore the monetary policy transmission mechanism by addressing the malfunctioning of certain government bond markets; see, for instance, González-Páramo (2011). 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, see Constâncio (2011). We focus on the impact of the actual bond purchases, and treat announcement effects as additional effects.

Compared to other central bank asset purchase programs, the SMP differs in several dimensions and, in particular, contains features that resemble foreign exchange interventions. First, SMP purchases were made during a severe sovereign debt crisis, when sovereign yields in several euro area countries were at a high, on the rise, and volatile. During this phase, the targeted securities met little private sector demand. The purchases were undertaken during the most intense phases of the debt crisis and in the markets most affected by the

[2] At the end of 2012, the ECB held €99.0bn in Italian sovereign bonds, €30.8bn in Greek debt, €43.7bn in Spanish debt, €21.6bn in Portuguese debt, and €13.6bn in Irish bonds, see the ECB (2013) Annual Report.

crisis. This contrasts with the setting of the Federal Reserve's LSAPs and the Bank of England's QE, where longer-term yields and yield volatilities were relatively low and default risk premia negligible. Second, key features of the SMP were not disclosed while the program was active, for example, the targeted securities, the amounts to be purchased, and the duration of the program. Apart from the initial announcements about the SMP,[3] market participants learned about the program as purchases were implemented in a non-anonymous dealer market. Finally, the introduction of the SMP was subject to significant controversy, both outside and within the Eurosystem, i.e., the ECB and all National Central Banks (NCBs). The extent of the controversy within the Eurosystem balance sheet became evident with the resignation of the Bundesbank President in February 2011 and an ECB Executive Board member in September 2011.

In general, the identification of the yield impact of bond market interventions is a challenging task. A cursory look at bond yields and bond purchases within the SMP can, as some suggested,[4] lead to the impression that the SMP was ineffective. In particular, yields were rising as purchases were taking place. This calls for a careful 'impact identification.' We follow Ghysels et al. (2016) who argue that reliable impact identification can be achieved by moving to a high frequency, such as data sampled in 15-minute intervals.

4.2 High-frequency data on bond yields and purchases

Our data on euro area bond yields are from Thomson Reuters. Bond yields are sampled at the 15-minute frequency between 8AM and 6PM. We consider 5-year benchmark bond yields, and focus on the midpoint between ask and bid prices. Bond prices are expressed in yields-to-maturity and are obtained from continuous dealer quotes.

Figure 2 plots the yield-to-maturity of five-year benchmark bonds for five euro area countries between 04 January 2010 and 31 December 2012 at the 15-minute frequency. During the debt crisis, some yields exhibited occasional large and sudden moves. These sudden

[3]See the ECB press release of 10 May 2010 "ECB decides on measures to address severe tensions in financial markets" and 7 August 2011 "Statement by the President of the ECB."

[4]See, for instance, Fulcrum Research Notes, January 2013 or Natixis Flash Economics Economic Research, 13 September 2013.

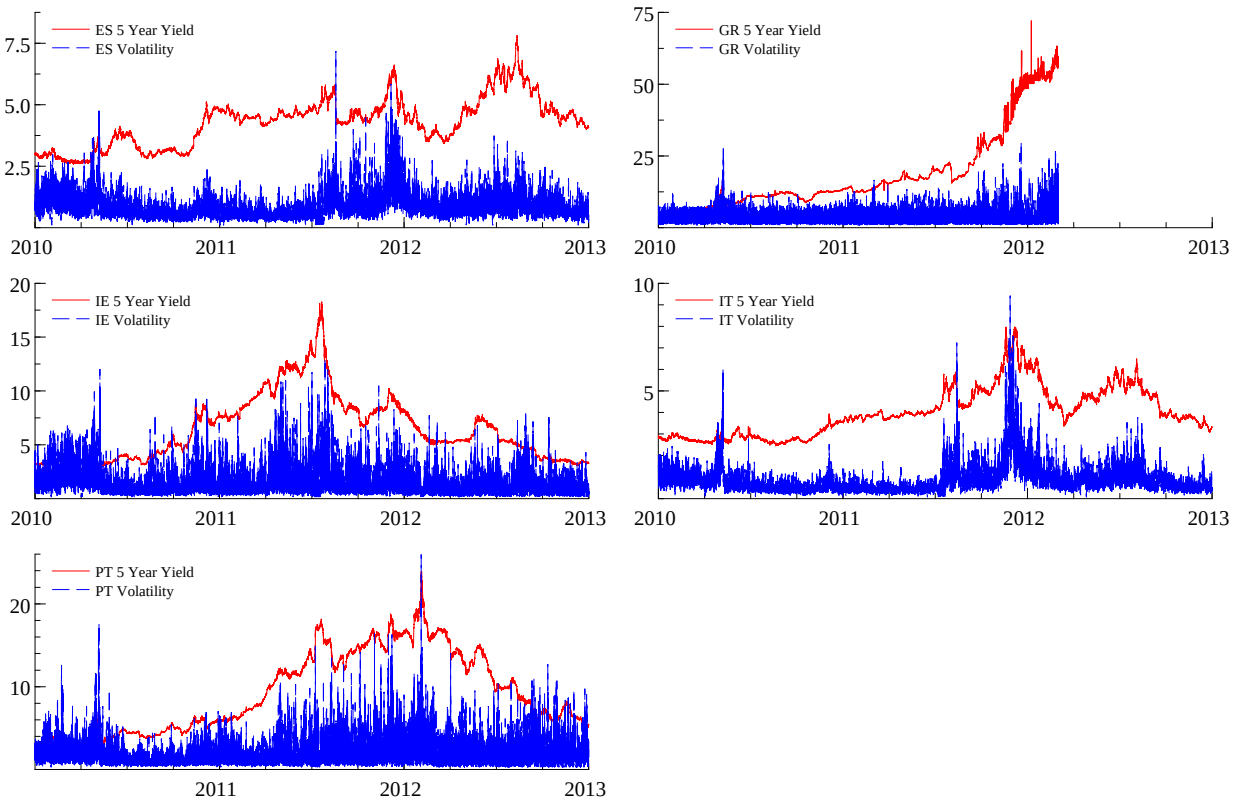

Figure 2: Five-year benchmark bond yields and volatility

Five-year benchmark bond yields (in levels) and volatility estimates σ_t (for changes in yields). The five panels refer to benchmark bond yields for Spain (ES), Greece (GR), Ireland (IE), Italy (IT), and Portugal (PT). Yields are in percentage points and are sampled at 15 minute intervals. Volatilities are in percentage points, and re-scaled to match yield levels to ensure visibility. Greek bonds discontinued trading after 02 March 2012, and experienced a credit event on 09 March 2012.

moves also led to strong volatility spikes.

Pronounced announcement effects of the SMP are visible in the yield data. Five-year yields dropped by -829 bps in Greece, -95 bps in Ireland, and -187 bps in Portugal on 10 May 2010; and by -90 bps in Spain and -62 bps in Italy on 8 August 2011, measured as the difference in five-year benchmark bond yields between Monday 6pm and the respective preceding Friday at 6pm. In addition, the credit event for Greek bonds on 09 March 2012 appears to have led to pronounced temporary spikes in yield levels and volatility in the other four markets.

Our high-frequency data on SMP asset purchases is the same as employed in Ghysels et al. (2016). The SMP's daily cross-country breakdown of the purchase data is confidential at the time of writing. We use the confidential country-specific data for this study.

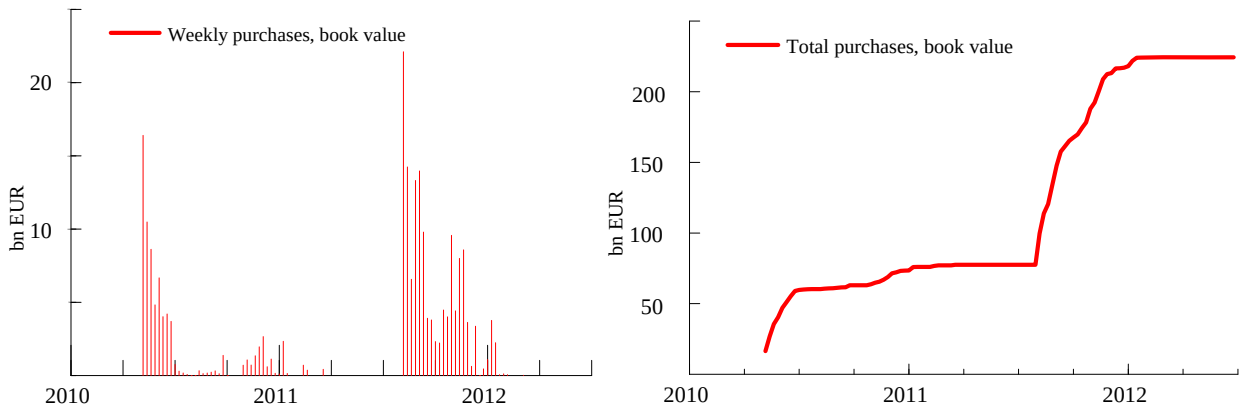

Figure 3: Weekly and total SMP purchase amounts.

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.

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. Visibly, the purchase data are unevenly spread over time. The SMP was announced on 10 May 2010 and focused on Greek, Irish, and Portuguese debt securities. The program was extended to include Italian and Spanish bonds on 8 August 2011. Between 10 May 2010 and Spring 2012 there are long periods during which the SMP was open but inactive.

4.3 Volatility modeling

We use a univariate parametric model to pre-filter our time series data before EVT estimation. The model is given by

$$\begin{aligned}
\tilde{y}_t &\sim \mathrm{t}(\tilde{y}_t; \mu_t, \sigma_t^2, \nu), && (12)\\
\mu_t &= \mu + \gamma_0 \cdot \tilde{y}_{t-1} + \gamma_1 \cdot \mathrm{SMP}_t + \gamma_2 \cdot \mathrm{D}_t^{\mathrm{10May}} + \gamma_3 \cdot \mathrm{D}_t^{\mathrm{08Aug}}, && (13)\\
\ln(\sigma_{t+1}) &= \omega_v + a_v \cdot s_{v,t} + b_v \cdot \ln(\sigma_t) + c_v \cdot \mathrm{I}(\tilde{y}_t > \mu_t) s_{v,t}\\
&+ d_v \cdot \mathrm{SMP}_{t+1} + e_v \cdot \mathrm{D}_{t+1}^{\mathrm{10May}} + f_v \cdot \mathrm{D}_{t+1}^{\mathrm{08Aug}},\\
\nu &= \bar{\nu}, && (14)
\end{aligned}$$

where $\tilde{y}_t$ is the change in the quoted yield of a certain bond after subtracting the intradaily seasonal pattern, ν is the degrees of freedom parameter, and $s_{v,t}$ is the scaled score from a t-distribution; see Creal et al. (2013) and Lucas et al. (2014) for details.

We extend the volatility factor specification with additional explanatory covariates along the lines of (11). In addition, we also model the impact of SMP purchases on the yield changes such that the mean μ_t is time varying. To complete the volatility model, we include a term $\mathrm{I}(\tilde{y}_t > \mu_t)s_{v,t}$ to capture the leverage effect between yield changes and volatility: the volatility is higher when the past yield change is positive. The key feature of the Student's t-GAS(1,1) model, which differentiates it from a conditionally Gaussian (GARCH) model, is a weighting term in its scaled score $s_{v,t}$ that lessens the impact of occasional extreme observations. Such extreme observations commonly occur during a sovereign debt crisis, and particularly at a high frequency.

The parametric volatility model is used to infer the impact of purchases on mean and volatility. We also use it to prefilter our data. EVT estimation is based on POT observations $x_t = \max(\tilde{y}_t/\hat{\sigma}_t - \tau, 0)$, where $\hat{\sigma}_t$ is the volatility estimate from (12).

4.4 Mean and volatility estimates

Table 2 presents the parameter estimates associated with specification (12).

We focus on three findings. First, bond purchases lowered the conditional mean of Bond yields by about -0.2 to -3.5 bps per 1 bn of purchases. The impact estimates are somewhat smaller than what is found in Eser and Schwaab (2016), and somewhat larger than what is found in Ghysels, Idier, Manganelli, and Vergote (2016). The highest impact per bn is observed for Greek bonds, which were the most illiquid at the time.

Second, there is no evidence that asset purchases lowered yields. The respective coefficients are insignificant in three out of five cases. When significant, the respective parameter estimates point towards a temporary *increase* in high-frequency volatility. This finding is sensitive to the model specification. If a Gaussian model is used instead of a fat-tailed t-distributed model, then all volatility coefficients are negative.

Finally, the announcement effects are not significant in our high-frequency specification.

Table 2: Parameter estimates

Parameter estimates for the univariate volatility model (12). ν converges from above to its lower bound at 2.05, and is therefore fixed at this value. Rows labeled ES, GR, IE, IT, and PT refer to Spanish, Greek, Irish, Italian, and Portuguese five-year bond yields. The estimation sample ranges from 04 January 2010 to 28 December 2012. For Greece, the sample ends on 02 March 2012. Standard errors are in round brackets and are constructed from the numerical second derivatives of the log-likelihood function. P-values are provided in square brackets. $D_{1,t}$ and $D_{2,t}$ are announcement dummy variables that capture announcement effects on 10 May 2010 and 08 August 2011, which takes value 1 for all time intervals during the announcement day.

	ES	GR	IE	IT	PT
	Parameters in the mean equation				
μ	0.018	0.004	0.014	0.003	0.018
	(0.005)	(0.018)	(0.004)	(0.005)	(0.007)
	0.000	0.820	0.002	0.536	0.005
γ_0	-0.053	-0.005	-0.048	-0.028	-0.063
	(0.005)	(0.002)	(0.004)	(0.006)	(0.005)
	0.000	0.002	0.000	0.000	0.000
SMP_t	-2.410	-3.516	-0.200	-1.588	-1.841
	(0.937)	(1.729)	(0.789)	(0.469)	(1.692)
	0.010	0.042	0.800	0.001	0.277
$\text{D}_t^{\text{10May}}$	-0.291	1.743	-1.179	-0.122	-1.657
	(0.239)	(5.644)	(0.866)	(0.366)	(0.907)
	0.224	0.757	0.174	0.739	0.068
$\text{D}_t^{\text{08Aug}}$	-0.220	-0.252	-0.531	0.302	-1.304
	(0.675)	(0.420)	(0.361)	(0.929)	(0.856)
	0.745	0.548	0.141	0.745	0.127
ν	2.554	2.050	2.050	2.668	2.050
	(0.046)	(0.000)	(0.000)	(0.048)	(0.000)
	0.000	0.000	0.000	0.000	0.000
	Parameters in the volatility equation				
ω^v	0.041	0.751	0.236	0.024	0.359
	(0.005)	(0.037)	(0.011)	(0.004)	(0.016)
	0.000	0.000	0.000	0.000	0.000
a^v	0.178	0.254	0.282	0.168	0.282
	(0.006)	(0.008)	(0.007)	(0.006)	(0.007)
	0.000	0.000	0.000	0.000	0.000
b^v	0.923	0.721	0.879	0.952	0.835
	(0.004)	(0.013)	(0.005)	(0.003)	(0.007)
	0.000	0.000	0.000	0.000	0.000
LEV	0.037	0.116	0.016	0.019	0.059
	(0.006)	(0.009)	(0.007)	(0.005)	(0.007)
	0.000	0.000	0.018	0.000	0.000
SMP_t	0.654	0.300	-0.283	0.241	-0.080
	(0.236)	(0.361)	(0.643)	(0.087)	(0.599)
	0.006	0.406	0.660	0.006	0.893
$\text{D}_t^{\text{10May}}$	0.057	0.645	0.230	0.006	0.244
	(0.043)	(0.110)	(0.076)	(0.037)	(0.091)
	0.186	0.000	0.003	0.879	0.008
$\text{D}_t^{\text{08Aug}}$	-0.018	-0.053	0.037	0.025	0.172
	(0.057)	(0.108)	(0.069)	(0.043)	(0.078)
	0.758	0.625	0.595	0.571	0.028
loglik	-56181.6	-69428.5	-70344.3	-56070.3	-80266.1

This means that the pronounced declines in bond yields that can be observed on these days do not happen in the first hour. Instead, they happen more gradually during the first day. The leverage terms are always positive and statistically significant. Consequently, increases in bond yields trigger larger increases in future volatilities than declining yields.

4.5 Tail shape estimates

This section discusses our tail shape estimates. Table 3 presents the parameter estimates associated with our dynamic tail shape model. We use the mixture distribution approach with a point mass at zero; see (9) and (10). Parameter estimates are reported for the 95%

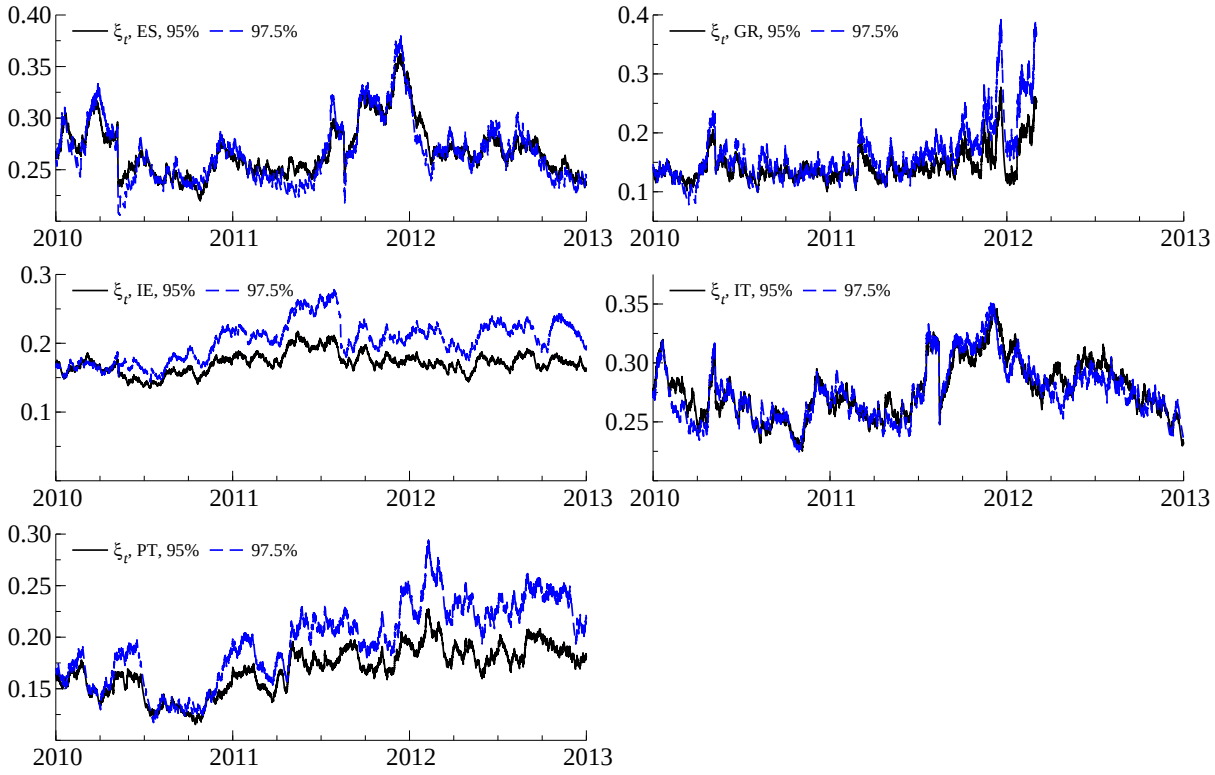

Figure 4: Tail risk dynamics in five euro area countries

Estimates of the time-varying tail shape parameter ξ_t at different levels of τ, corresponding to the 10%, 5%, and 2.5% quantile, for Spain (ES), Greece (GR), Ireland (IE), Italy (IT), and Portugal (PT). Daily estimates are obtained as the median value of the estimates at the 15 minute frequency. Shaded areas in 2010 and 2011 refer to frequent asset purchases within the SMP; see Figure 3. The shared area in late 2012 marks the period after the announcement of the technical details of the ECB's OMT program on 06 September 2012. Greek bonds discontinued trading after 02 March 2012, and experienced a credit event on 09 March 2012.

and 97.5% empirical quantile of yield changes at a 15-minute frequency.

The parameter estimates suggest that the announcement of the program had a significant impact on the tail shape in Italy and Spain. Once announced, there is no evidence that bond purchases lowered the tail shape parameter.

Figure 4 plots the filtered estimates of ξ_t. The tail shape is reported for different thresholds τ at the 5% and 2.5% quantiles of x_t, respectively. Grey shaded areas in 2010 and 2011 mark periods of focussed SMP purchases in the respective markets, cf. Figure 3. The grey shaded area in 2012 marks the time after the announcement of the technical details of the ECB's OMT on 06 September 2012.

Table 3: Parameter estimates for the tail shape model

Parameter estimates for the score-driven model for time-varying tail shape. We use the mixture distribution approach with a point mass at zero; see (9) and (10). Rows labeled ES, 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, except for Greece, for which the sample ends on 02 March 2012. Standard errors are in round brackets and are constructed from the numerical second derivatives of the log-likelihood function. P-values are provided in square brackets. The top and bottom panels refer to the 0.95 and 0.975 quantile. $D_{1,t}$ and $D_{2,t}$ are announcement dummy variables and take the value of one on 10 May 2010 and 08 August 2011.

Tail(0.95) model with GAS-t(2.05) filtering							
	ω	a	b	SMP_t	$\text{D}_{1,t}$	$\text{D}_{2,t}$	LogLik
ES	-0.001	0.002	0.999	0.022	-0.005	-0.007	-6877.2
	(0.000)	(0.000)	(0.000)	(0.008)	(0.003)	(0.003)	
	[0.013]	[0.000]	[0.000]	[0.010]	[0.090]	[0.036]	
GR	-0.003	0.007	0.998	0.001	-0.006	0.001	-6572.9
	(0.001)	(0.001)	(0.001)	(0.014)	(0.006)	(0.005)	
	[0.006]	[0.000]	[0.000]	[0.913]	[0.294]	[0.790]	
IE	-0.001	0.002	0.999	-0.021	-0.002	0.000	-7914.8
	(0.001)	(0.001)	(0.000)	(0.029)	(0.004)	(0.003)	
	[0.063]	[0.000]	[0.000]	[0.468]	[0.712]	[0.873]	
IT	-0.001	0.002	0.999	0.006	-0.004	-0.007	-7561.2
	(0.000)	(0.000)	(0.000)	(0.003)	(0.003)	(0.003)	
	[0.083]	[0.000]	[0.000]	[0.089]	[0.133]	[0.031]	
PT	-0.001	0.002	1.000	-0.008	0.001	0.002	-8739.1
	(0.000)	(0.001)	(0.000)	(0.017)	(0.004)	(0.003)	
	[0.095]	[0.000]	[0.000]	[0.613]	[0.713]	[0.419]	

Tail(0.975) model with GAS-t(2.05) filtering							
	ω	a	b	SMP_t	$\text{D}_{1,t}$	$\text{D}_{2,t}$	LogLik
ES	-0.001	0.002	0.999	0.013	-0.006	-0.006	-11672.1
	(0.000)	(0.000)	(0.000)	(0.006)	(0.003)	(0.003)	
	[0.000]	[0.000]	[0.000]	[0.022]	[0.036]	[0.035]	
GR	-0.002	0.006	0.999	-0.023	0.006	0.000	-8361.9
	(0.001)	(0.001)	(0.001)	(0.013)	(0.005)	(0.004)	
	[0.034]	[0.000]	[0.000]	[0.068]	[0.259]	[0.954]	
IE	-0.001	0.004	1.000	-0.031	-0.005	0.000	-11428.1
	(0.000)	(0.001)	(0.000)	(0.035)	(0.005)	(0.003)	
	[0.098]	[0.000]	[0.000]	[0.380]	[0.312]	[0.962]	
IT	-0.001	0.002	1.000	0.004	-0.009	-0.009	-11815.2
	(0.001)	(0.001)	(0.001)	(0.003)	(0.003)	(0.004)	
	[0.573]	[0.024]	[0.000]	[0.227]	[0.002]	[0.036]	
PT	-0.002	0.003	0.999	-0.006	-0.005	-0.001	-12124.2
	(0.001)	(0.001)	(0.001)	(0.025)	(0.005)	(0.003)	
	[0.148]	[0.001]	[0.000]	[0.818]	[0.339]	[0.855]	

5 Conclusion

This paper introduced time variation into the tail shape parameter of the Generalized Pareto Distribution, yielding a novel observation-driven model. Specifically, our modeling framework allows us to track the time variation in the tail index of time series observations from a wide class of fat-tailed distributions. In the empirical application, we demonstrated that two controversial unconventional monetary policies adopted by the ECB during the euro area sovereign debt crisis lowered the tail shape and mitigated the extreme market risks associated with holding certain sovereign bonds between 2010–2012.

References

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

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.

Brownlees, C. T. and R. Engle (2015). SRisk: A conditional capital shortfall index for systemic risk measurement. *Unpublished working paper*.

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.

Constâncio, V. (2011). Contagion and the European debt crisis. Keynote lecture at Bocconi University Milan on 10 October 2011.

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.

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.

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

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

Eser, F., M. Carmona Amaro, S. Iacobelli, and M. Rubens (2012). The use of the Eurosystem's monetary policy instruments and operational framework since 2009. ECB Occasional Paper 135, European Central Bank.

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 (2016). A high frequency assessment of the ECB Securities Markets Programme. *Journal of European Economic Association, forthcoming*.

González-Páramo, J.-M. (2011). The ECB's monetary policy during the crisis. Closing speech at the Tenth Economic Policy Conference, Málaga, 21 October 2011.

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

Lenza, M., H. Pill, and L. Reichlin (2010). Monetary policy in exceptional times. *25*(4), 295–339.

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

Lucas, A., B. Schwaab, and X. Zhang (2016). Modeling financial sector joint tail risk in the euro area. *Journal of Applied Econometrics, forthcoming*.

Lucas, A. and X. Zhang (2016). Score driven exponentially weighted moving averages and Value-at-Risk forecasting. *International Journal of Forecasting, forthcoming*.

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

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.

Poon, S., M. Rockinger, and J. Tawn (2004). Extreme value dependence in financial markets. *Review of Financial Studies 17(2)*, 581–610.

Pooter, M. D., R. F. Martin, and S. Pruitt (2017). The liquidity effects of official bond market intervention. *Journal of Financial and Quantitative Analysis, forthcoming*.

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.

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.

Appendix A1: Score and scaling function for a GPD random variable

This section derives (4) – (5). Recall the GPD pdf

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

and that the respective log-likelihood is given by (2) as $l(x_t;\delta_t,\xi_t)=-\ln(\delta_t)-\left(1+\frac{1}{\xi_t}\right)\ln\left(1+\xi_t\frac{x_t}{\delta_t}\right)$, where $\delta_t>0$, $\xi_t>0$, $x_t>0$.

Firstly, we load unobserved factor into ξ_t so that $\xi_t=\exp(f_{1t})$. The score function (4) is straightforward, and is obtained as

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

Similarly, we have $\delta_t=\exp(f_{2t})$ and the derivation of the score follows

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

We are able to see that the score vector is

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

In the next few sections, we derive the 2 by 2 matrix which will be used for the scaling of the score. We know

$$\mathcal{S}_t=\begin{bmatrix}\mathcal{S}_t^{(11)} & \mathcal{S}_t^{(12)}\\ \mathcal{S}_t^{(21)} & \mathcal{S}_t^{(22)}\end{bmatrix}, \tag{A1}$$

where $\mathcal{S}_t^{(11)}$ and $\mathcal{S}_t^{(22)}$ are the scaling function for the tail index and scale parameter and the off-diagonal elements contain the cross derivatives of the two parameters.

A1.1: The scaling function for the tail index

The score is zero in expectation if the model is well-specified; see Creal et al. (2013), implying

$$\int_0^\infty \frac{1}{\xi_t^2} \ln\left(1+\xi_t \frac{x_t}{\delta_t}\right) p(x_t;\delta_t,\xi_t)\mathrm{d}x_t = \int_0^\infty \left(1+\frac{1}{\xi_t}\right) \frac{x_t}{\delta_t+\xi_t x_t} p(x_t;\delta_t,\xi_t)\mathrm{d}x_t. \tag{A2}$$

The scaling function is chosen as the inverse conditional Fisher information of the GPD,

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

where the last equality sign uses a standard result. The expected negative second derivative is

$$\begin{aligned}
& \mathrm{E}\left[-\frac{\partial^2 l(x_t;\delta_t,\xi_t)}{\partial \xi_t^2}\right] \\
= & -\int_0^\infty \left[\left(1+\frac{1}{\xi_t}\right)\frac{x_t^2}{(\delta_t+\xi_t x_t)^2} + \frac{2}{\xi_t^2}\frac{x_t}{\delta_t+\xi_t x_t} - \frac{2}{\xi_t^3}\ln\left(1+\xi_t\frac{x_t}{\delta_t}\right)\right] p(x_t;\delta_t,\xi_t)\mathrm{d}x_t \\
= & -\int_0^\infty \left[\left(1+\frac{1}{\xi_t}\right)\frac{x_t^2}{(\delta_t+\xi_t x_t)^2} + \frac{2}{\xi_t^2}\frac{x_t}{\delta_t+\xi_t x_t} - \frac{2}{\xi_t}\left(1+\frac{1}{\xi_t}\right)\frac{x_t}{\delta_t+\xi_t x_t}\right] p(x_t;\delta_t,\xi_t)\mathrm{d}x_t \\
= & -\int_0^\infty \left[\left(1+\frac{1}{\xi_t}\right)\frac{x_t^2/\delta_t^2}{(1+\xi_t x_t/\delta_t)^2} - \frac{2}{\xi_t}\frac{x_t/\delta_t}{1+\xi_t x_t/\delta_t}\right]\frac{1}{\delta_t}\left(1+\xi_t\frac{x_t}{\delta_t}\right)^{-\frac{1}{\xi_t}-1}\mathrm{d}x_t \\
= & -\int_0^\infty \left[\left(\frac{1+\xi_t}{\xi_t^3}\right)\frac{\xi_t^2 x_t^2/\delta_t^2}{(1+\xi_t x_t/\delta_t)^2} - \frac{2}{\xi_t^2}\frac{\xi_t x_t/\delta_t}{1+\xi_t x_t/\delta_t}\right]\frac{1}{\delta_t}\left(1+\xi_t\frac{x_t}{\delta_t}\right)^{-\frac{1}{\xi_t}-1}\mathrm{d}x_t \\
= & -\frac{1+\xi_t}{\xi_t^4}\int_1^\infty (u_t-1)^2 u_t^{-1/\xi_t-3}\mathrm{d}u_t + \frac{2}{\xi_t^3}\int_1^\infty (u_t-1)u_t^{-1/\xi_t-2}\mathrm{d}u_t,
\end{aligned} \tag{A3}$$

where we used (A2) in the second line, and where the last equality comes from a change of variable substituting $u_t = 1+\xi_t x_t/\delta_t$.

The two integrals in (A3) can be treated as scaled up first moments $\mathrm{E}(u_t)$. To see this, note that the new random variable u_t is Pareto (Type 1) distributed, with pdf

$$\tilde{p}(u_t;a_t,b) = \frac{a_t b^{a_t}}{{u_t}^{a_t+1}}, \qquad u_t > b, \tag{A4}$$

with nth un-centered moment available as

$$\mathrm{E}(u_t^n) = \frac{a_t b^n}{a_t - n}, \qquad \text{if } a_t > n \tag{A5}$$

The first integral in (A3) corresponds to $\tilde{p}(u_t; 1/\xi_t + 2, 1)$, while the second integral corresponds to $\tilde{p}(u_t; 1/\xi_t + 1, 1)$. This implies that

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

Adding terms and inverting, we obtain the scaling function (5) in closed form as

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

A1.2: The scaling function for the scale parameter

The scaling function is the inverse conditional Fisher information of the GPD with respect to δ_t,

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

The formula in the expectation is

$$\begin{aligned}
& \mathrm{E}\left[-\frac{\partial^2 l(x_t; \delta_t, \xi_t)}{\partial \delta_t^2}\right] \\
= & -\int_0^\infty \left[\frac{1/\delta_t^2 - 2x_t/\delta_t^3 - \xi_t x_t^2/\delta_t^4}{(1+\xi_t x_t/\delta_t)^2}\right] \frac{1}{\delta_t} \left(1 + \xi_t \frac{x_t}{\delta_t}\right)^{-\frac{1}{\xi_t} - 1} \mathrm{d}x_t \\
= & -\int_0^\infty \frac{1}{\delta_t^3} \left(1 + \xi_t \frac{x_t}{\delta_t}\right)^{-\frac{1}{\xi_t} - 3} \mathrm{d}x_t + \int_0^\infty \frac{2}{\delta_t^3} \frac{x_t}{\delta_t} \left(1 + \xi_t \frac{x_t}{\delta_t}\right)^{-\frac{1}{\xi_t} - 3} \mathrm{d}x_t + \int_0^\infty \frac{\xi_t}{\delta_t^3} \frac{x_t^2}{\delta_t^2} \left(1 + \xi_t \frac{x_t}{\delta_t}\right)^{-\frac{1}{\xi_t} - 3} \mathrm{d}x_t \\
= & -\frac{1}{\xi_t \delta_t^2} \int_1^\infty u_t^{-1/\xi_t - 3} \mathrm{d}u_t + \frac{2}{\xi_t^2 \delta_t^2} \int_1^\infty (u_t - 1) u_t^{-1/\xi_t - 3} \mathrm{d}u_t + \frac{1}{\xi_t^2 \delta_t^2} \int_1^\infty (u_t - 1)^2 u_t^{-1/\xi_t - 3} \mathrm{d}u_t \\
= & -\frac{1}{\delta_t^2 (1+2\xi_t)} + \frac{2}{\delta_t^2 (1+\xi_t)(1+2\xi_t)} + \frac{2\xi_t}{\delta_t^2 (1+\xi_t)(1+2\xi_t)} \\
= & \frac{1}{\delta_t^2 (1+2\xi_t)}.
\end{aligned} \tag{A6}$$

The derivation uses the property of aforementioned Pareto distribution properties. We obtain the second scaling function in closed form

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

A1.3: The scaling function of the cross derivatives

The scaling matrix shall contain the cross derivatives with respect the the two latent factors,

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

The derivations are similar as procedures before

$$\begin{aligned}
&\mathrm{E}\left[-\frac{\partial^2 l(x_t;\delta_t,\xi_t)}{\partial\xi_t\partial\delta_t}\right]\\
=\ & -\int_0^\infty\left[\frac{-x_t/\xi_t}{\delta_t^2+\xi_t x_t\delta_t}+(1+\xi_t)\frac{x_t/\xi_t}{(\delta_t+\xi_t x_t)^2}\right]\frac{1}{\delta_t}\left(1+\xi_t\frac{x_t}{\delta_t}\right)^{-\frac{1}{\xi_t}-1}\mathrm{d}x_t\\
=\ & \frac{1}{\xi_t^3\delta_t}\int_1^\infty(u_t-1)u_t^{-1/\xi_t-2}\mathrm{d}u_t-\frac{1}{\xi_t^3\delta_t}\int_1^\infty(u_t-1)u_t^{-1/\xi_t-3}\mathrm{d}u_t-\frac{1}{\xi_t^2\delta_t}\int_1^\infty(u_t-1)u_t^{-1/\xi_t-3}\mathrm{d}u_t\\
=\ & \frac{1}{\xi_t\delta_t(1+\xi_t)}-\frac{1}{\xi_t\delta_t(1+2\xi_t)}\\
=\ & \frac{1}{\delta_t(1+\xi_t)(1+2\xi_t)}.
\end{aligned}\tag{A7}$$

And the second term

$$\begin{aligned}
&\mathrm{E}\left[-\frac{\partial^2 l(x_t;\delta_t,\xi_t)}{\partial\delta_t\partial\xi_t}\right]\\
=\ & -\int_0^\infty\left[\frac{1}{\delta_t}\frac{x_t/\delta_t-x_t^2/\delta_t^2}{(1+\xi_t x_t/\delta_t)^2}\right]\frac{1}{\delta_t}\left(1+\xi_t\frac{x_t}{\delta_t}\right)^{-\frac{1}{\xi_t}-1}\mathrm{d}x_t\\
=\ & -\frac{1}{\xi_t^2\delta_t}\int_1^\infty(u_t-1)u_t^{-1/\xi_t-3}\mathrm{d}u_t+\frac{1}{\xi_t^3\delta_t}\int_1^\infty(u_t-1)^2u_t^{-1/\xi_t-3}\mathrm{d}u_t\\
=\ & -\frac{1}{\xi_t^2\delta_t}\frac{\xi_t^2}{(1+\xi_t)(1+2\xi_t)}+\frac{1}{\xi_t^3\delta_t}\frac{2\xi_t^3}{(1+\xi_t)(1+2\xi_t)}\\
=\ & \frac{1}{\delta_t(1+\xi_t)(1+2\xi_t)}.
\end{aligned}\tag{A8}$$

The derivation uses the property of aforementioned Pareto distribution properties. We obtain

the second scaling function in closed form

$$\mathcal{S}_t^{(12)} = \frac{(1+\xi_t)(1+2\xi_t)}{\xi_t},$$

and

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

Now we can gather all the elements in the scaling matrix as

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

Appendix A2: Score and scaling function for a mixture of a GPD with a point mass at zero

This section derives (9) – (10). Recall that the mixture GPD density (7) has a point mass at zero and is given by

$$\phi(x_t;\delta_t,\xi_t) = \mathrm{I}(x_t>0)\left[(1+\tau)^{-\frac{1}{\xi_t}}\frac{1}{\delta_t}\left(1+\xi_t\frac{x_t}{\delta_t}\right)^{-\frac{1}{\xi_t}-1}\right] + \mathrm{I}(x_t=0)\left(1-(1+\tau)^{-\frac{1}{\xi_t}}\right).$$

The respective log-likelihood (8) is given by

$$\begin{aligned}\ell(x_t;\delta_t,\xi_t) &= \mathrm{I}(x_t>0)\left[-\frac{1}{\xi_t}\ln(1+\tau)-\ln(\delta_t)-\left(1+\frac{1}{\xi_t}\right)\ln\left(1+\xi_t\frac{x_t}{\delta_t}\right)\right]\\ &+ \mathrm{I}(x_t=0)\ln(1-(1+\tau)^{-\frac{1}{\xi_t}}).\end{aligned}$$

The derivation of the score follows Appendix A1, with some adjustments. In particular,

$$\begin{aligned}\nabla_{1t} &= \frac{\partial\ell(x_t;\delta_t,\xi_t)}{\partial f_{1t}} = \frac{\partial\ell(x_t;\delta_t,\xi_t)}{\partial\xi_t}\cdot\frac{\mathrm{d}\xi_t}{\mathrm{d}f_{1t}},\\ \frac{\partial\ell(x_t;\delta_t,\xi_t)}{\partial\xi_t} &= \mathrm{I}(x_t>0)\left[\frac{1}{\xi_t^2}\ln\left(1+\xi_t\frac{x_t}{\delta_t}\right)-\frac{\xi_t+1}{\xi_t}\frac{x_t}{\delta_t+\xi_t x_t}\right] &\text{(A9)}\\ &+ \mathrm{I}(x_t>0)\left[\frac{1}{\xi_t^2}\ln(1+\tau)\right]+\mathrm{I}(x_t=0)\left[\frac{1}{\xi_t^2}\ln(1+\tau)\right] &\text{(A10)}\\ &- \mathrm{I}(x_t=0)\left[\frac{1}{1-(1+\tau)^{-\frac{1}{\xi_t}}}\frac{1}{\xi_t^2}\ln(1+\tau)\right] &\text{(A11)}\\ \frac{\mathrm{d}\xi_t}{\mathrm{d}f_{1t}} &= \exp(f_{1t}).\end{aligned}$$

The first part of the score (A9) is the same as in Appendix A1. (A9) is adjusted by (A10)–(A11) to take account of the point mass at zero. Combining (A9)–(A11) yields (9). Now we also provide the score with respect to $f_{2t} = ln(\delta_t)$.

$$\nabla_{2t} = \frac{\partial\ell(x_t;\delta_t,\xi_t)}{\partial f_{2t}} = \frac{\partial\ell(x_t;\delta_t,\xi_t)}{\partial\delta_t}\cdot\frac{\mathrm{d}\delta_t}{\mathrm{d}f_{2t}} = \mathrm{I}(x_t>0)\frac{x_t-\delta_t}{\delta_t+\xi_t x_t} \qquad \text{(A12)}$$

As before, the scaling function is the inverse conditional Fisher information of the mixture

density. We can also borrow from the derivation done in the Appendix A.1 in the derivation of

$$\mathcal{S}_t = \begin{bmatrix} \mathcal{S}_t^{(11)} & \mathcal{S}_t^{(12)} \\ \mathcal{S}_t^{(21)} & \mathcal{S}_t^{(22)} \end{bmatrix}, \tag{A13}$$

where $\mathcal{S}_t^{(11)}$, $\mathcal{S}_t^{(12)}$, and $\mathcal{S}_t^{(22)}$ are defined as before.

$$\mathcal{S}_t^{(11)} = \mathrm{E}[\nabla_t^2 | \mathcal{F}_{t-1}; f_t, \psi]^{-1} = \mathrm{E}\left[-\frac{\partial^2 \ell(x_t; \delta_t, \xi_t)}{\partial \xi_t^2} \right]^{-1} \exp(-2f_{1t}).$$

The expectation of the negative second derivative is given by the sum of three expected negative partial derivatives of (A9)–(A11). Specifically,

$$\begin{aligned}
\mathrm{E}\left[-\frac{\partial^2 \ell(x_t; \delta_t, \xi_t)}{\partial \xi_t^2} \right] &= \mathrm{E}\left[-\frac{\partial\ (\mathrm{A9})}{\partial \xi_t} \right] + \mathrm{E}\left[-\frac{\partial\ (\mathrm{A10})}{\partial \xi_t} \right] + \mathrm{E}\left[-\frac{\partial\ (\mathrm{A11})}{\partial \xi_t} \right], \\
\text{where} \quad \mathrm{E}\left[-\frac{\partial\ (\mathrm{A9})}{\partial \xi_t} \right] &= \frac{2(1+\tau)^{-\frac{1}{\xi_t}}}{(1+2\xi_t)(1+\xi_t)}, && \text{(A14)} \\
\mathrm{E}\left[-\frac{\partial\ (\mathrm{A10})}{\partial \xi_t} \right] &= \frac{2}{\xi_t^3} \ln(1+\tau), && \text{(A15)} \\
\mathrm{E}\left[-\frac{\partial\ (\mathrm{A11})}{\partial \xi_t} \right] &= \frac{(1+\tau)^{-\frac{1}{\xi_t}}}{1-(1+\tau)^{-\frac{1}{\xi_t}}} \frac{1}{\xi_t^4} \ln(1+\tau)^2 - \frac{2}{\xi_t^3} \ln(1+\tau), && \text{(A16)}
\end{aligned}$$

and where (A14) follows from the result in Appendix A1. The indicator functions in (A9)–(A11) disappear as they determine the domain of integration when taking expectations. Note that (A16) does not require integration, but evaluation at a point mass. Combining terms, we obtain (10) in closed form as

$$\mathcal{S}_t^{(11)} = \left[\frac{2\xi_t^2 (1+\tau)^{-\frac{1}{\xi_t}}}{(1+2\xi_t)(1+\xi_t)} + \frac{(1+\tau)^{-\frac{1}{\xi_t}}}{1-(1+\tau)^{-\frac{1}{\xi_t}}} \frac{1}{\xi_t^2} \ln(1+\tau)^2 \right]^{-1}.$$

We can combine the results from the last section with the new definition of mixture between the point mass and GPD distribution

$$\begin{aligned}
\mathcal{S}_t^{(22)} &= (1+\tau)^{\frac{1}{\xi_t}} (1+2\xi_t); && \text{(A17)} \\
\mathcal{S}_t^{(12)} = \mathcal{S}_t^{(21)} &= (1+\tau)^{\frac{1}{\xi_t}} \frac{(1+\xi_t)(1+2\xi_t)}{\xi_t}. && \text{(A18)}
\end{aligned}$$