SVERIGES RIKSBANK
WORKING PAPER SERIES
399

Modeling extreme events: time-varying extreme tail shape

Enzo D'Innocenzo, Bernd Schwaab, Xin Zhang and André Lucas

December 2020 (Updated June 2023)

WORKING PAPERS ARE OBTAINABLE FROM

www.riksbank.se/en/research
Sveriges Riksbank • SE-103 37 Stockholm
Fax international: +46 8 21 05 31
Telephone international: +46 8 787 00 00

The Working Paper series presents reports on matters in
the sphere of activities of the Riksbank that are considered
to be of interest to a wider public.
The papers are to be regarded as reports on ongoing studies
and the authors will be pleased to receive comments.

Modeling extreme events: time-varying extreme tail shape*

Enzo D'Innocenzo,[a] *André Lucas,*[b]
Bernd Schwaab,[c] *Xin Zhang*[d]

[a] University of Bologna,

[b] Vrije Universiteit Amsterdam and Tinbergen Institute,

[c] European Central Bank, [d] BIS Innovation Hub and Sveriges Riksbank

Sveriges Riksbank Working Paper Series
No. 399

Abstract

We propose a dynamic semi-parametric framework to study time variation in tail parameters. The framework builds on the Generalized Pareto Distribution (GPD) for modeling peaks over thresholds as in Extreme Value Theory, but casts the model in a conditional framework to allow for time-variation in the tail parameters. We establish parameter regions for stationarity and ergodicity and for the existence of (unconditional) moments and consider conditions for consistency and asymptotic normality of the maximum likelihood estimator for the deterministic parameters in the model. Two empirical datasets illustrate the usefulness of the approach: daily U.S. equity returns, and 15-minute euro area sovereign bond yield changes.

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

JEL classification: *C22, G11.*

*Author information: Enzo D'Innocenzo, University of Bologna, Bologna, Italy, Email: enzo.dinnocenzo2@unibo.it. 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, BIS Innovation Hub Nordic Centre and Sveriges Riksbank, Stockholm, Sweden, Email: xin.zhang@riksbank.se. The views expressed in this paper are those of the authors and do not necessarily reflect the views or policies of the BIS, the European Central Bank or Sveriges Riksbank.

1 Introduction

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

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

We illustrate our modeling framework using two applications: U.S. equity log-returns, and changes in euro area sovereign bond yields.[1] Each data set consists of two time series that we model separately. We first consider daily log-returns for a stock index (S&P500) and an individual stock (IBM). The S&P500 log-returns range from July, 3 1962 to December,

[1]The web appendix considers additional applications to exchange rates and commodity prices.

31 2020, while the IBM stock returns range from January 2, 1926 to December 31, 2020. Focusing on the left tail, and controlling for potential time-variation in the EVT thresholds, we find that both GPD parameters vary significantly over time. The tail shape varies between approximately 0.05 and 0.25 for the S&P 500, and between 0.05 and 0.35 for IBM. These values imply a maximum moment of order $1/0.25 = 4$ to $1/0.05 = 20$ for the S&P 500, and 3 to 20 for IBM. Confidence bands around the filtered time-varying tail shape parameters suggest that the tail shape parameter is almost always far from the thin-tailed setting.

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

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

Bund future returns. Galbraith and Zernov (2004) argues that certain regulatory changes in U.S. equity markets have altered the tail index dynamics of equities returns. de Haan and Zhou (2021) propose a non-parametric approach to estimating the extreme value index locally. Our paper adds to the literature by proposing a framework that allows us to study both the tail shape and tail scale dynamics directly in a semi-parametric way. Explanatory covariates can be included in the dynamics of both parameters, and likelihood ratio tests are available to test economically relevant hypotheses. Finally, unlike Patton et al. (2019), our tail VaR and ES dynamics explicitly account for fat tail shape beyond a threshold as emerging from EVT. The dynamics based on the score for the GPD contain weights for extreme observations. Such weights are absent in the elicitable score functions of Patton et al. (2019). The resulting dynamics in our model are more robust, particularly for the ES.

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

We proceed as follows. Section 2 presents our statistical model, the asymptotic statistical properties of the model and the maximum likelihood estimator. Section 3 discusses simulation results. Section 4 illustrates the model using U.S. equity data and euro area sovereign yields data. Section 5 concludes. Derivations and more results are in the web appendix.

2 Statistical model

2.1 Time-varying tail shape and tail scale

2.1.1 Conditional EVT framework

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

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

Fréchet, and Burr distribution with one or more time-varying shape parameters (for details and further discussion see e.g. Johnson et al. (1994), Embrechts et al. (1997), or McNeil et al. (2010, Ch. 7.3)). Rather than modeling the (dynamic) tail shape of an arbitrarily chosen parametric family of distributions, we appeal to well-known results from the EVT literature: the conditional cumulative distribution function (cdf) $G(y_t \mid \mathcal{F}_{t-1})$ of y_t can under very general conditions be approximated by $G(y_t \mid \mathcal{F}_{t-1}) \approx G(\tau_t \mid \mathcal{F}_{t-1}) + (1 - G(\tau_t \mid \mathcal{F}_{t-1}))P(x_t; \delta_t, \xi_t)$ with $x_t = y_t - \tau_t$ for sufficiently high thresholds $\tau_t \in \mathbb{R}_+$. More precisely, we have

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

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

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

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

The result in (1) is a limiting result. In any finite sample, the threshold τ_t has to be set to a specific, finite value, such that the GPD approximation will be inexact and the distribution is in that sense misspecified. This will also be the case in our setting. The

score-driven updates that we define later on for ξ_t and δ_t, however, still ensure that the expected Kullback-Leibler divergence between the approximate GPD model and the true, unknown conditional distribution $\mathbb{P}\left[Y_t \leq x_t + \tau_t \mid Y_t > \tau_t,\ \mathcal{F}_{t-1}\right]$ is improved on average at every step for sufficiently small steps, even if the GPD model is (partially) misspecified; see Blasques et al. (2015).

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

2.1.2 Time-varying parameters

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

$$f_{t+1} = \omega + \sum_{i=0}^{q-1} A_i s_{t-i} + \sum_{j=0}^{p-1} B_j f_{t-j}, \tag{3}$$
$$s_t = \mathcal{S}_t \nabla_t, \qquad \nabla_t = \partial \ln p(x_t \mid \mathcal{F}_{t-1}; f_t, \theta)/\partial f_t,$$

where vector $\omega = \omega(\theta)$ and matrices $A_i = A_i(\theta)$ and $B_j = B_j(\theta)$ depend on the 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 (3) updates f_t at every point in time via a scaled steepest ascent step to improve the expected fit to the GPD; see Creal et al. (2013); Blasques et al. (2015). The score of (2) required in (3) is given by

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

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

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

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

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

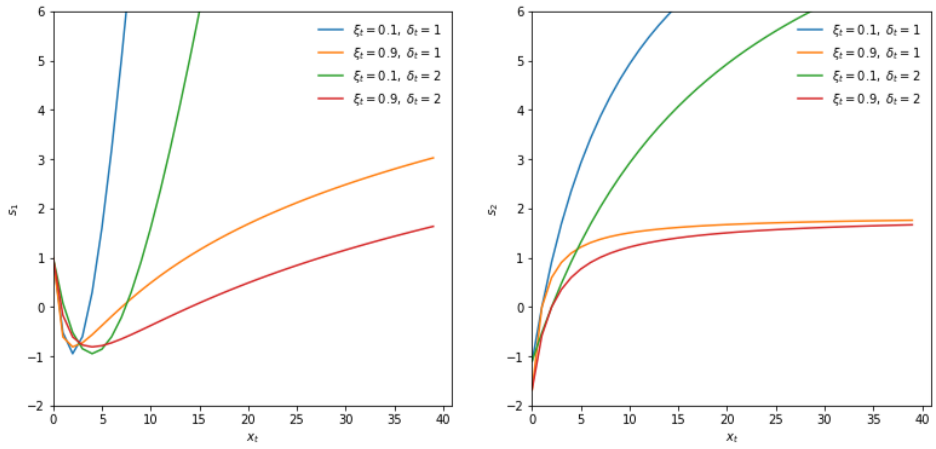
$$s_t = L_t' \nabla_t = \begin{bmatrix} \xi_t^{-2}(1 + \xi_t) \cdot \ln\left(1 + \xi_t\, \delta_t^{-1} x_t\right) + \dfrac{\delta_t - (\xi_t + 3 + \xi_t^{-1}) \cdot x_t}{\delta_t + \xi_t x_t} \\ \sqrt{1 + 2\xi_t} \dfrac{x_t - \delta_t}{\delta_t + \xi_t x_t} \end{bmatrix}. \tag{6}$$

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

Figure 1 plots the two elements of (6) as a function of x_t for different values of ξ_t and δ_t.

Figure 1: News impact curves

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



The behavior of the scaled score is intuitive: large x_ts imply that both ξ_t and δ_t are adjusted upwards and tails thus become fatter. The adjustments are largest when the current tail index ξ_t and tail scale δ_t are low. It is precisely in such cases that observing a large x_t is unlikely. If it occurs nonetheless, the parameters are (strongly) adjusted to account better for similar effects in the future.

The news impact curves become increasingly concave for lower values of ξ_t. For such parameters, large x_t values are already more likely due to the fat-tailed nature of the GPD itself. As a result, the parameters need to be adjusted less if a large x_t actually materializes. This resembles the well-known robustness properties of score-driven updates in the context of time-varying volatility modeling; see Creal et al. (2013) and Harvey (2013). It also distinguishes our current set-up sharply from an approach directly based on quantile functions; see Patton et al. (2019) and Catania and Luati (2023), in particular for risk measures such as ES. In Patton et al. (2019), ES reacts linearly rather than concavely to the VaR exceedance. This can result in noisy or unstable ES estimates. Figure 1 illustrates that the score-driven approach is less susceptible to such instabilities and can therefore result in more interpretable parameter paths.

We also note that small realizations of x_t imply downward adjustments of both ξ_t and δ_t, up to the point where x_t becomes very small. For very small $x_t > 0$, ξ_t is adjusted

upward: observations near the center of a fat-tailed distribution signal increased peakedness (=leptokurtosis) and thus higher ξ_t; compare Lucas and Zhang (2016) for a similar effect in the Student's t setting.

When there is no tail observation, i.e. $x_t = y_t - \tau_t \leq 0$, the new observation carries no information about ξ_t and δ_t; see McNeil et al. (2010, Chapter 7). In such cases we set the score to zero and continue to use (3) to update f_t. Long consecutive stretches of zero scores could potentially lead to mean-reverting paths for f_t, and thus (ξ_t, δ_t), with only discrete "jumps" when new observations $x_t > 0$ arrive following such stretches.[2] If so, smoothing the scaled score (6) can help to spread out the impact of the new information in x_t. Smoothing the scaled scores can also help when additional conditioning variables z_t are available at every $t = 1, \ldots, T$; see (8) and (10) below. Lagged values of the scaled score can be taken into account via an exponentially-weighted moving average specification

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

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

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

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

where all explanatory variables are stacked into the column vector z_t, and C is a conformable

[2] Alternatively, one could opt to not update f_t at all until a new $x_t > 0$ arrives. Empirically, both approaches seem to work equally well.

matrix of impact coefficients that needs to be estimated. We illustrate this in our second application in Section 4.

2.1.3 Time-varying thresholds

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

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

where $\omega^\tau \equiv (1-b^\tau)\cdot \hat{q}^\kappa$, $\hat{q}^\kappa$ is the (observed) unconditional κ-quantile of y_t, $a^\tau > 0$ and $0 < b^\tau < 1$ are two parameters to be estimated, and $(1-\kappa)$ is a sufficiently small tail probability corresponding to the dynamic quantile τ_t, such as, for example, 10% or 5%. We initialize τ_t at $\tau_1 = \hat{q}^\kappa$. The recursive specification (9) is driven by a zero mean innovation process since $\mathbb{E}[1\{y_t > \tau_t\} \mid \mathcal{F}_{t-1}, \theta^\tau] = (1-\kappa)$. The threshold τ_{t+1} responds to quantile exceedances in an intuitive way: the next quantile value τ_{t+1} receives a positive shock of $a^\tau\kappa$ if $y_t > \tau_t$, i.e., if the previous quantile was exceeded, and a negative shock of $-a^\tau(1-\kappa)$ otherwise. For $0 < b^\tau < 1$, the empirical unconditional quantile $\hat{q}^\kappa$ serves as a long-term attractor for (9). The transition equation for τ_t can also easily be extended to include exogenous variables z_t as in

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

for a suitable column vector of coefficients c^τ.

2.1.4 Interpretation of time-varying parameters

It is important to briefly comment on parameter interpretability. The tail shape parameter ξ_t can always be interpreted as observation y_t's inverse conditional tail index α_t^{-1}. By contrast, the estimated tail scale parameter δ_t need not have a straightforward interpretation in terms of y_t's conditional scale σ_t. For example, assume that y_t has a GPD conditional distribution with time-varying tail shape parameter α_t^{-1} and scale σ_t. Web Appendix B.1 demonstrates

that the conditional tail probability $\mathbb{P}[Y_t \leq y_t + \tau_t \mid Y_t > \tau_t, \mathcal{F}_{t-1}]$ then also has a GPD shape (exactly, not only approximately). The tail shape parameter is the same as that of the *center*: $\xi_t = \alpha_t^{-1}$. However, the tail scale parameter δ_t is very different from the scale parameter σ_t that applies to the center, in particular $\delta_t = \sigma_t + \alpha_t^{-1} \cdot \tau_t$. As a result, δ_t increases with the threshold τ_t, varies positively with the tail shape parameter ξ_t, and, importantly, should not be expected to provide a good estimate of the scale parameter σ_t that applies to the *center* of the distribution of y_t. A similar result can be derived if y_t were Student's t-distributed with scale σ_t and degrees of freedom parameter α_t when the tail probability $\mathbb{P}[Y_t \leq y_t + \tau_t \mid Y_t > \tau_t, \mathcal{F}_{t-1}]$ only has an approximate GPD shape; see Web Appendix B.2. We return to this issue in our simulation Section 3, where we consider pseudo-true values for ξ_t and δ_t to benchmark how well the model performs in terms of tracking an unknown data generating process.

2.2 Stationarity and moments

The score-driven dynamics for ξ_t and δ_t are highly non-linear. Still, the structure of the model allows us to obtain sufficient conditions for the stationarity and ergodicity (SE) of f_t and x_t, and for the existence of unconditional moments of f_t. To this end, we look at our statistical model in equations (3)–(7) as a data generating process (DGP).

Given the bivariate structure of our time-varying parameter model, the process $\{f_t\}_{t \in \mathbb{Z}}$ can be viewed as a stochastic recurrence equation (SRE) of the form

$$f_{t+1} = \Phi_t(f_t; \theta_0) = \Phi(f_t, x_t; \theta_0), \qquad t \in \mathbb{Z}, \tag{11}$$

where $\Phi : \mathbb{R}^2 \times \mathbb{R}^+ \times \Theta \mapsto \mathbb{R}^2$ is a Borel measurable function, and $\theta_0 \in \Theta$ is the true parameter vector contained in the parameter space $\Theta \subset \mathbb{R}^6$. We make the following two assumptions.

Assumption 1. *Assume that x_t is drawn from the GPD distribution defined in (2), such that the random variable $\epsilon_t := \xi_t^{-1} \ln(1 + \xi_t x_t / \delta_t)$ is an independent and identically distributed noise term with unit exponential distribution, $\epsilon_t \overset{iid}{\sim} \mathcal{E}xp(1)$.*

Assumption 2. *For some integer $r \geq 1$, let*

$$\mathbb{E}\left[\ln \sup_{f} \left\| \prod_{i=1}^{r} \dot{\Phi}_{t-i}\left(f_{t-i}; \theta_0\right) \right\|\right] < 0, \tag{12}$$

where $\|M\| = \sqrt{\text{trace}\left(M'M\right)}$ is the Frobenius norm of a real-valued matrix $M \in \mathbb{R}^{m \times n}$, and where $\dot{\Phi}_t\left(f_t; \theta_0\right) := B + A\frac{\partial \tilde{s}_t}{\partial s_t'} \cdot \frac{\partial s_t}{\partial f_t'}$, such that

$$\dot{\Phi}_t\left(f_t; \theta_0\right) := \begin{pmatrix} b^{\xi} + a^{\xi}\left(1-\lambda\right) \cdot \frac{\partial s_t^{\xi}}{\partial f_t^{\xi}} & 0 \\ a^{\delta}\left(1-\lambda\right) \cdot \frac{\partial s_t^{\delta}}{\partial f_t^{\xi}} & b^{\delta} \end{pmatrix}, \tag{13}$$

with

$$\begin{aligned} \frac{\partial s_t^{\xi}}{\partial f_t^{\xi}} &= -\xi_t^{-1}\epsilon_t - \xi_t\epsilon_t \exp(-\xi_t\epsilon_t) + (3\xi_t^{-1} + 2\xi_t^{-2})(1 - \exp(-\xi_t\epsilon_t)) \\ &\qquad - (\xi_t + 3 + \xi_t^{-1})\epsilon_t \exp(-\xi_t\epsilon_t) \\ \frac{\partial s_t^{\delta}}{\partial f_t^{\xi}} &= \frac{1}{\sqrt{1+2\xi_t}}\left(1 - \exp(-\xi_t\epsilon_t) - \xi_t \exp(-\xi_t\epsilon_t)\right) + \\ &\qquad \sqrt{1+2\xi_t} \cdot \left(-\xi_t^{-1}(1 - \exp(-\xi_t\epsilon_t)) + \epsilon_t \exp(-\xi_t\epsilon_t) + \xi_t\epsilon_t \exp(-\xi_t\epsilon_t)\right). \end{aligned}$$

Assumption A1 considers the model as the data generating process. By inverting the GPD cdf in (2), we obtain that $\epsilon_t = \xi_t^{-1}\ln\left(1 + \xi_t\, x_t/\delta_t\right) = -\ln(1-u_t) \sim \mathcal{E}xp(1)$ for a standard uniformly distributed u_t. Assumption A1 requires that these uniform random variables constitute an i.i.d. process.

Assumption A2 requires contraction properties of the bivariate stochastic recurrence equation. We note that we use the general form of the r-fold contractions, i.e., r iterations of the transition equation of f_t. In univariate models, sharp contraction conditions (that ensure stationarity and ergodicity of the model) can usually be found by assuming that $r = 1$; see, for example, Blasques et al. (2022). In multivariate systems, however, the contraction condition with $r = 1$ is often violated, resulting in very small or uninteresting stationarity regions. Working with the more general condition is therefore important. We also show this in Figure 2 below. The idea behind A2 is that, when working with a multi-

variate system, one can still ensure the existence of an SE solution if the system becomes contractive eventually, i.e., after a sufficiently large number of r iterations. The contraction condition in A2 results in a meaningful (non-degenerate) SE region, because the supremum of $\dot{\Phi}_t(f_t;\theta_0)$ over f has no degeneracies.

Using A1 and A2, we can verify the conditions of Bougerol (1993, Theorem 3.1). This allows us to show that a unique SE solution exists for the bivariate score-driven process $\{f_t\}_{t\in\mathbb{Z}}$, and for the data $\{x_t\}_{t\in\mathbb{Z}}$ as generated by the DGP in Section 2.1. We summarize this in the following theorem.

Theorem 1. *Consider the model as defined by (2) and (3). Then under A1 and A2, the SRE in (11) admits a unique stationary and ergodic solution $\{\tilde{f}_t\}_{t\in\mathbb{Z}}$ such that, for each fixed initial condition $f_0 \in \mathbb{R}^2$,*

$$\left\| f_t - \tilde{f}_t \right\| \xrightarrow{e.a.s.} 0 \qquad as \qquad t \to \infty,$$

where $\xrightarrow{e.a.s.}$ denotes exponentially fast and almost sure convergence (Straumann and Mikosch (2006)). In addition, the process $\{x_t\}_{t\in\mathbb{Z}}$ generated by the model evaluated at θ_0 is stationary and ergodic.

Web Appendix C.1 presents the proof of Theorem 1. Next to the SE properties of the f_t process, we can also establish the existence of unconditional moments. This can be useful for proving the existence of moments of the log-likelihood function and its derivatives. We note that the contraction condition stated in A2 by its own is insufficient to ensure bounded unconditional moments of f_t in the DGP. Requiring the existence of moments typically makes the admissible parameter space smaller. We also note that the existence of unconditional moments of f_t does not imply the same (un)conditional moments for x_t. For instance, even if f_t has a finite 4th order moment, x_t may not if ξ_t can reach levels higher than $1/4$. We have the following result.

Theorem 2. *Consider the model as defined by (2) and (3), and let A1 and A2 be true. If*

in addition

$$\mathbb{E}\left[\sup_{f}\left\|\prod_{i=1}^{r}\dot{\Phi}_{t-i}\left(f_{t-i};\theta_0\right)\right\|^{p}\right]<1, \tag{14}$$

for some $p\geq 1$*, then the unique stationary and ergodic solution* $\{\tilde{f}_t\}_{t\in\mathbb{Z}}$ *to the SRE in* (11) *satisfies* $\mathbb{E}[\|\tilde{f}_t\|^p]<\infty$.

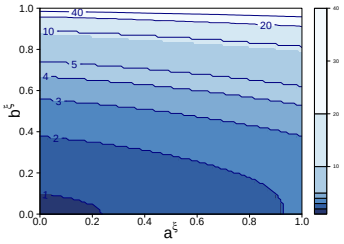
Web Appendix C.2 provides the proof of Theorem 2. Again, Theorem 2 makes use of r-fold contraction conditions rather than the standard $r=1$ case.

To give some insight into the size of the SE and moments regions, we compute them numerically in Figure 2. We let two (out of the six) parameters in Θ vary at a time. Computing the regions is far from trivial. It requires numerically solving a maximization problem embedded inside an integration problem. As a result, the maximization problem has to be solved for every value of the integration variable. Details are provided in Web Appendix D.

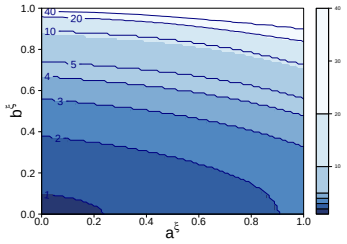
Figure 2 clearly illustrates the importance of multiple unfoldings of the SRE. If only $r=1$ iteration is used, the SE and finite-moments regions are typically small, or even empty; see for instance the small darkest blue region in panel 3a for $r=1$, or the fact that panel 3d only shows contracting behavior for $r\geq 17$ iterations given the empirical estimates. For $r=1$, empirical estimates typically lie outside the SE region. However, by iterating the SRE forward, the SE and finite-moments regions grow considerably to the extent that they also encompass the empirical estimates. For instance, the model may not be SE when evaluated at the empirical estimates for $r=1$. For $r=20$, 40, or even larger, however, the model's SE region at the empirical estimates increases such that even 4th order moments of f_t exist. Given the S&P500 data (on which Figure 2 is based) are at a daily frequency, we conclude that a time-varying parameter model may not have guaranteed contraction properties at the daily frequency, but may still have such contraction properties at a monthly ($r=22$) or lower frequency. The numerical results in the figure stress that analytical conditions for SE, as often encountered in the literature for $r=1$, may be quite uninteresting if not accompanied by a numerical check for the size of the resulting parameter space.

Figure 2: Stationarity and Moments regions

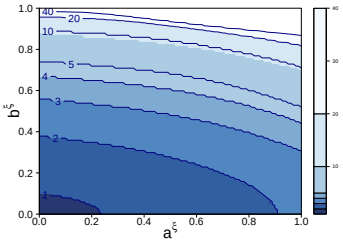
Top row: the combinations of a^ξ and b^ξ that satisfy A2 (left panel), and (14) for $p = 2$ (middle panel) and $p = 4$ (right panel), for an increasing number of iterations $r \geq 1$. Bottom row: similar, but for combinations of a^δ and b^δ instead. Other deterministic parameters for each panel are fixed at their empirical estimates for S&P500 returns; see Section 4.1.



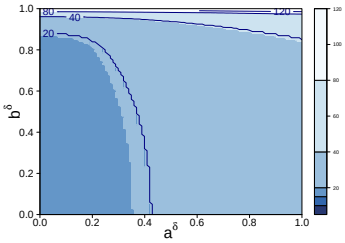
(a) SE region in (a^ξ, b^ξ)



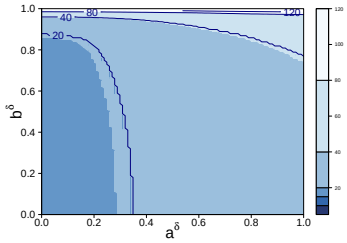
(b) 2nd moments region



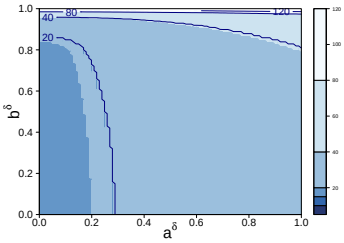
(c) 4th moments region



(d) SE region in (a^δ, b^δ)



(e) 2nd moments region



(f) 4th moments region

2.3 Parameter estimation

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

$$\begin{aligned} \mathcal{L}(\theta|\mathcal{F}_T) &= \frac{1}{T^*} \sum_{t=1}^{T} 1\{x_t > 0\} \cdot \ln p(x_t; \delta_t, \xi_t) \\ &= \frac{1}{T^*} \sum_{t=1}^{T} 1\{x_t > 0\} \cdot \left[-\ln(\delta_t) - \left(1 + \frac{1}{\xi_t}\right) \ln\left(1 + \xi_t \frac{x_t}{\delta_t}\right) \right], \end{aligned} \tag{15}$$

where $T^* = \sum_{t=1}^{T} 1\{x_t > 0\}$ is the number of POT values in the sample. Maximization of (15) can be carried out using a conveniently chosen quasi-Newton optimization method.

To establish consistency and asymptotic normality of the maximum likelihood estimator

(MLE), we define the MLE $\hat{\theta}$ as

$$\hat{\theta} := \arg\max_{\theta \in \Theta} \mathcal{L}\left(\theta | \mathcal{F}_T\right). \tag{16}$$

We establish these asymptotic properties by showing that the conditions for Lemma 1 in Jensen and Rahbek (2004) are satisfied. The main point to note is that the first three derivatives of the log-likelihood function (and of f_t) with respect to θ are stationary and ergodic stochastic recurrence equations that satisfy specific moment conditions. We make the following additional assumption on the parameter space Θ.

Assumption 3. *The parameter vector $\theta \in \Theta$ for a compact parameter space $\Theta \subset \mathbb{R}^6$, and the true parameter vector $\theta_0 \in int(\Theta)$. Additionally, for each $\theta \in \Theta$, the starting value $f_0(\theta) = (\ln(\xi_0), \ln(\delta_0))' \in \mathbb{R}^2$ is fixed to the true value.*

Finally, we impose conditions on the score vector ∇_t, as defined in (4), and its derivatives up to third-order. Let $\nabla_t(\theta)$, $\nabla_t^2(\theta)$ and $\nabla_t^3(\theta)$ denote the score vector and first two derivatives with respect to $f_t^{\xi}(\theta)$ and $f_t^{\delta}(\theta)$, evaluated at some $\theta \in V(\theta_0)$, where $V(\theta_0)$ denotes a small neighborhood of the true parameter vector θ_0. By imposing the conditions in Assumption 4 below, we can rule out explosive behavior of the first three derivatives of the log-likelihood function with respect to θ. Similar boundedness assumptions have been used in for instance Hetland et al. (2023) or Hafner and Wang (2023). Effectively, this further reduces the size of the parameter space (see also Blasques et al., 2022). Web Appendix E.1 provides the explicit analytic expressions for each of these derivatives.

Assumption 4. *The score vector $\nabla_t(\theta)$, and its derivatives $\nabla_t^2(\theta)$, and $\nabla_t^3(\theta)$ are p-dominated uniformly in $V(\theta_0)$ and t by dominating functions $D_{1,t}$, $D_{2,t}$ and $D_{3,t}$, respectively, such that*

$$\sup_{\theta \in V(\theta_0)} \|\nabla_t(\theta)\| \leq D_{1,t}, \qquad \sup_{\theta \in V(\theta_0)} \left\|\nabla_t^2(\theta)\right\| \leq D_{2,t}, \qquad \text{and} \qquad \sup_{\theta \in V(\theta_0)} \left\|\nabla_t^3(\theta)\right\| \leq D_{3,t},$$

where $D_{1,t}$, $D_{2,t}$, and $D_{3,t}$ are p-integrable uniformly in t for $p > 0$, that is $\|D_{1,t}\|^p \leq \Delta_1 < \infty$, $\|D_{2,t}\|^p \leq \Delta_2 < \infty$, and $\|D_{3,t}\|^p \leq \Delta_3 < \infty$.

The following theorem establishes the asymptotic properties of our MLE.

Theorem 3. *Consider the model as defined by (2) and (3). Let Assumptions 1–3 hold true. Then* $\hat{\theta} \xrightarrow{P} \theta_0$ *as* $T^* \to \infty$.

If, in addition, Assumption 4 holds true, as well as the contraction condition (14) in Theorem 2 for $p = 4$*, then* $\sqrt{T^*}\left(\hat{\theta} - \theta_0\right) \Rightarrow N\left(0, \Omega_I^{-1}\right)$ *as* $T^* \to \infty$*, where* Ω_I *denotes the Fisher information matrix evaluated at the true parameter vector* θ_0.

The proof of Theorem 3, together with detailed derivations of the derivatives up to the third-order of the log-likelihood function $\mathcal{L}(\theta|\mathcal{F}_T)$, and of the (bivariate) score-driven process $\{f_t(\theta)\}_{t\in\mathbb{Z}}$, for each $\theta \in \Theta$ can be found in Web Appendix E.2.

All our results are conditional on the time-varying thresholds τ_t. The parameters a^τ and b^τ for τ_t in (9) cannot be estimated using (15). 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

$$\begin{aligned} &\min_{\{a^\tau, b^\tau\}} \frac{1}{T}\sum_{t=1}^{T} \rho_\kappa\left(y_t - \tau_t\right) \iff \min_{\{a^\tau, b^\tau\}} \frac{1}{T}\sum_{t=1}^{T} (y_t - \tau_t)\left(\kappa - 1\{y_t < \tau_t\}\right) \\ \iff &\max_{\{a^\tau, b^\tau\}} \frac{1}{T}\sum_{t=1}^{T} (y_t - \tau_t)\left((1-\kappa) - 1\{y_t > \tau_t\}\right), \end{aligned} \tag{17}$$

where $\rho_\kappa(u_t) = u_t\left(\kappa - 1\{u_t < 0\}\right)$, and τ_t evolves as in (9). See also Engle and Manganelli (2004) and Catania and Luati (2023) for the use of this objective function in a different dynamic context. In practice, we estimate all thresholds τ_t via (17) before maximizing (15).

2.4 Confidence bands for tail shape and tail scale

Given the maximum likelihood estimate $\hat{\theta}$, confidence (or standard error) bands around $\hat{f}_t = f_t(\hat{\theta})$ allow us to visualize the impact of estimation uncertainty. Quantifying the uncertainty of the estimated parameter paths is important, as classical EVT estimators of time-invariant tail shape parameters can have sizeable standard errors; see e.g. Hill (1975) and Huisman et al. (2001). Web Appendix F explains how simulation-based and in-sample analytic confidence bands around $\hat{f}_t$ can be obtained. These bands are conditional on the estimated time-varying thresholds $\hat{\tau}_t$, and do not incorporate the associated estimation un-

certainty of the thresholds.

2.5 Market risk measures

Market risk measurement is a major application of EVT methods in practice; see e.g. McNeil et al. (2010). The GPD approximation (1)–(2) yields useful closed-form estimators of the VaR and ES for high upper quantiles $\gamma > G(\tau_t \mid \mathcal{F}_{t-1})$; see McNeil and Frey (2000) and Rocco (2014). We can estimate the $1-\gamma$ tail probability of y_t based on the GPD cdf for x_t, obtaining $\text{VaR}^{\gamma}(y_t \mid \mathcal{F}_{t-1}, \theta) = \tau_t + \delta_t \xi_t^{-1} \left[\left(\frac{1-\gamma}{t^*/t} \right)^{-\xi_t} - 1 \right]$, where t^* is the number of observations with $x_t > 0$ up to time t, i.e., the number of y_s for $s = 1, \ldots, t$ for $y_s > \tau_s$.

The conditional ES is the average conditional VaR in the tail across all quantiles γ (see McNeil et al., 2010, Chapter 2), provided $\xi_t < 1$. The closed-form expression is $\text{ES}^{\gamma}(y_t \mid \mathcal{F}_{t-1}, \theta) = \frac{\text{VaR}^{\gamma}(y_t|\mathcal{F}_{t-1},\theta)}{1-\xi_t} + \frac{\delta_t - \xi_t \tau_t}{1-\xi_t}$, see Web Appendix G for a derivation of the equations. Maximum likelihood estimators of the conditional VaR and conditional ES can be obtained by inserting filtered estimates of ξ_t and δ_t into the VaR and ES equations, respectively.

3 Simulation study

3.1 Simulation design

This section investigates the ability of our dynamic EVT model to simultaneously recover (i) the time variation in tail shape and tail scale parameters ξ_t and δ_t, (ii) EVT-based market risk measures $\text{VaR}^{\gamma}(y_t; \tau_t, \xi_t, \delta_t)$ and $\text{ES}^{\gamma}(y_t; \tau_t, \xi_t, \delta_t)$ at high confidence levels such as $\gamma = 99\%$, and (iii) parameter estimates and their standard errors for all deterministic parameters collected in θ. We do so using two sets of data generating processes (DGPs). The first set of DGPs is discussed in the main text, while the second set of DGPs is discussed in Web Appendix H. In our experiments, we track the performance of our score-driven modeling approach when putting it into a variety of challenging settings, such as when the conditional density is only approximately correct, the time-varying parameter process is misspecified or features (near)-unit roots, and/or when both tail parameters follow similar paths. We also

investigate whether the market risk estimates that follow from the model are reliable.

In our first set of DGPs, we consider $D = 2$ different densities (GPD and Student's t), $P = 4$ different parameter paths for tail shape and scale, and $H = 3$ different ways to obtain the appropriate thresholds τ_t. This yields $2 \times 4 \times 3 = 24$ simulation experiments. In each experiment, we draw $S = 100$ univariate simulation samples of length $T = 25,000$. We focus on the upper $1 - \kappa = 5\%$ tail. As a result, approximately $25,000 \cdot 0.05 = 1,250$ observations are available in each simulation to compute informative POTs $x_t > 0$.

We first simulate y_t from a GPD$(\alpha_t^{-1}, \sigma_t)$ distribution with time-varying tail shape α_t^{-1} and scale σ_t. In a second set-up, we consider a Student's t distribution with time-varying scale σ_t and degrees of freedom α_t. In the GPD case, our score-driven model uses the exact conditional density for x_t, while in the t case the GPD conditional density for x_t is only approximately correct (given finite thresholds τ_t in any given sample); see Web Appendix B.

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

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

Consequently, paths (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. Paths (2) allows the tail shape to vary considerably between 0.2 and 0.8, while keeping the scale (volatility) of the data constant. Paths (3) stipulates that both parameters vary over time. Finally, paths (4) considers the case of synchronized variation in both parameters.

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

recursive specification (9), initialized at the full-sample quantile $\tau_1 = Q^{0.95}_{1:T}$.

Our main evaluation metric for evaluating model performance is the root mean squared error RMSE $= \frac{1}{S}\sum_{s=1}^{S}\sqrt{\frac{1}{T}\sum_{t=1}^{T}(\hat{\xi}_t^s - \bar{\xi}_t^s)^2}$, where $\hat{\xi}_t^s$ is the estimated tail shape parameter in simulation s, $\bar{\xi}_t^s$ is the corresponding (pseudo-)true tail shape, $s = 1, \ldots, S$ denotes the simulation run, and $t = 1, \ldots, T$ is the number of observations in each draw. The RMSE for the tail scale parameter δ_t is obtained analogously. The pseudo-true values $\bar{\xi}_t^s$ and $\bar{\delta}_t^s$ are obtained by numerically minimizing the Kullback-Leibler divergence between the GPD and the data generating process beyond the true time-varying 95% quantile τ_t. As the true conditional density is known at all times in a simulation setting, these pseudo-true benchmarks are easily computed numerically for each s and t. Particularly the GPD tail scale parameter $\bar{\delta}_t$ may have very different dynamics from σ_t, as it combines dynamics in α_t and σ_t via the EVT limiting expression in (1); see also Section 2.1.4.

3.2 Simulation results

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

Table 1 presents the corresponding results. It reports RMSE statistics for tail shape $\hat{\xi}_t$ and tail scale $\hat{\delta}_t$. Figure 4 provides a representative example of the simulation outcomes where we compare median estimated parameter paths for $\hat{\xi}_t$, $\hat{\delta}_t$, $\widehat{\text{VaR}}^{0.99}$, and $\widehat{\text{ES}}^{0.99}$ to their (pseudo)-true values. Full results are found in Web Appendix H.1. The true parameters in Figure 4 follow Path 3 from Section 3.1, and time-varying thresholds are estimated based on the recursive specification (9) and objective function (17).

We focus on three main findings. First, all models seem to work well in recovering the true underlying ξ_t and δ_t dynamics. The median estimates in Figure 4 tend to be close to their (pseudo)-true values. The full results in in Web Appendix H.1 confirm this. Even the highly non-linear patters of δ_t are recovered well. The model also captures the peaks of ξ_t, which correspond to the episodes with the fattest tails. The model needs some time to

Table 1: RMSE outcomes for DGP1

Root mean squared error (RMSE) statistics for two different distributions (GPD and t, in columns) and for four different parameter paths for tail shape ξ_t and tail scale δ_t (paths (1) – (4), in rows). Thresholds τ_t, $\hat{\tau}_t$, and $\hat{\tau}_t^*$ denote, respectively, *i*) the infeasible true time-varying threshold, *ii*) the empirical quantile associated with an expanding window of observations $y_1, \ldots, y_t$, and *iii*) the estimated conditional quantile using (17) and a suitably calibrated $a^\tau = 0.25$ to speed up the computations. We consider 100 simulations for each DGP, and a time series of $25,000$ observations in each simulation. Model performance is measured by the RMSE from the true $\bar{\xi}_t$ and $\bar{\delta}_t$ in each draw.

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

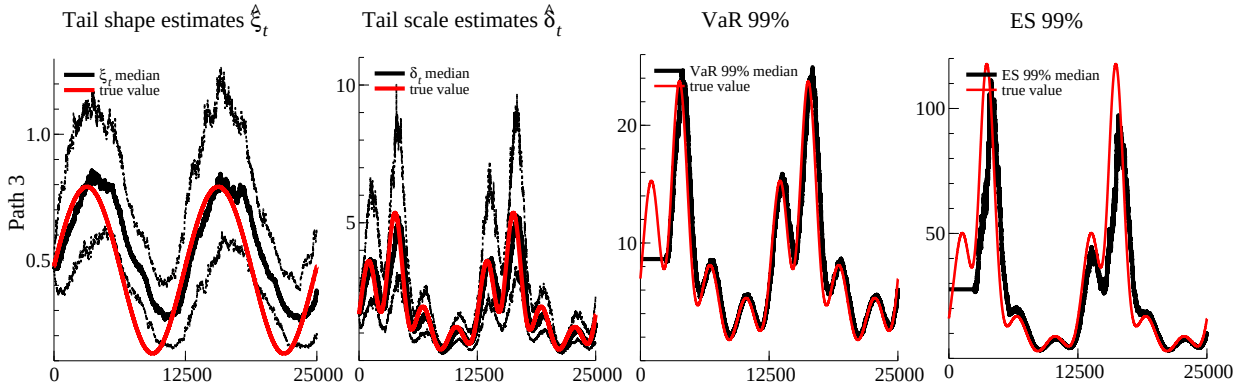
recognize that the extreme tail has become more benign, i.e., that ξ_t has gone down. The good fit is corroborated by Table 1. We also note that both estimation methods for τ_t only loose about 10% RMSE for ξ_t and δ_t compared to the use of the true (infeasible) τ_t.

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

Third, Figure 4 shows that EVT-based market risk measures such as high-confidence level ($\gamma = 0.99$) VaRs and ESs tend to be estimated sufficiently accurately with our dynamic EVT approach. Again, this is confirmed by the full results in Web Appendix H.1. Both

Figure 4: Simulation results: an example

Time series data is here generated as $y_t \sim \mathrm{t}(0, \sigma_t, \alpha_t)$, where $\alpha_t^{-1} = 0.5 + 0.3\sin(4\pi t/T)$ and $\sigma_t = 1 + 0.5\sin(16\pi t/T)$. This is Path 3 in Section 3.1. Pseudo-true parameter values are reported in solid red. The four panels report estimates of ξ_t, δ_t, VaR_t, and ES_t, respectively. Median filtered values are plotted in solid black. The first two panels also indicate the lower 5% and upper 95% quantiles of the estimates (black dots). The time-varying threshold $\hat{\tau}_t$ is estimated based on the recursive specification (9) in conjunction with the objective function (17).



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

We conclude this section by briefly summarizing the main simulation results from our second set of DGPs in Web Appendix H.2. Empirical estimates of the autoregressive parameters b^{ξ} and b^{δ} can be close to one; see Section 4. For this reason, we investigate the effect of (near-)unit root type dynamics and of additional covariates on the parameter estimates and their standard errors. We find that models with (near-)unit root type dynamics continue to work reliably. The estimated $\hat{\delta}$, $\hat{\xi}_t$, and $\hat{\delta}_t$ continue to be closely aligned to their true values. The usual asymptotic standard error estimates based on the inverse Hessian or sandwich estimates, however, are not always reliable then. In our set-up, these common estimates of the standard errors are typically too large, providing too conservative inference. A bootstrap procedure tailored to integrated processes could then be used to avoid this issue.[3]

[3]See for instance Boswijk et al. (2021).

4 Empirical illustrations

4.1 Equity log-returns

To illustrate our approach, we obtain end-of-day prices for the S&P500 index and for IBM stock as two easily and publicly available series from the CRSP database.[4] The S&P500 data range from July, 3 1962 to December, 31 2020, yielding 14,726 daily observations. The IBM stock data range from January 2, 1926 to December 31, 2020, yielding 25,028 daily observations. To model the adverse left tail of equity log-returns we consider *negative* log-returns $y_t = -100 \times (\ln p_t - \ln p_{t-1})$, with p_t the price level, before applying our methodology.

4.1.1 Deterministic parameter estimates

We rely on the time-variation in the thresholds τ_t to accommodate time-variation in any parameters describing the center of the distribution. The thresholds evolve over time according to (9), at $(1-\kappa) = 10\%$, and are initialized at $\tau_1 = \hat{q}^{0.9}$, the 90% empirical quantile of y_t. The factor process $f_t = (\ln \xi_t, \ln \delta_t)'$ is initialized at $f_1 = (\mathrm{I}_2 - B)^{-1}\omega$.

The first two columns in Table 2 present our estimates of the deterministic parameters of model (1) – (9). The estimates of $a^\tau > 0$ and $b^\tau < 1$ suggest that the thresholds are time-varying and mean-reverting. Parameters a^ξ and a^δ are statistically significant at any reasonable significance level, for both S&P500 and IBM. These parameters can be interpreted as the average size of the scores driving $\ln \xi_t$ and $\ln \delta_t$, respectively; see the statements above equation (5). Parameters b^ξ and b^δ are estimated to be close to one for both series, implying that shocks to each time-varying parameter die out only slowly. A numerical check reveals that the deterministic parameters a^ξ and b^ξ lie within the SE region implied by the sufficient conditions of Theorems 1 and 2. A diagnostic check in Web Appendix J suggests that the bivariate filter (3) is also invertible at these estimates.

[4] Web Appendix I provides two additional illustrations to other asset classes: exchange rates (GBP/USD) and commodities (Brent crude oil).

Table 2: Parameter estimates

Parameter estimates for the dynamic tail shape model. The second and third columns refer to the first application (equity log-returns of the S&P500 index and IBM stock). The estimation samples range from 3 July 1962 to 31 December 2020 for the S&P500 index, and from 2 January 1926 to 31 December 2020 for IBM stock. The remaining columns refer to the second application (changes in sovereign yields). Columns labeled IT 5y and PT 5y refer to yield changes for Italy and Portuguese five-year benchmark bonds sampled at the 15-minute frequency. The estimation samples range from 4 January 2010, 9AM, to 31 December 2012, 5PM. For the second application, b^ξ and b^δ are estimated indistinguishably different from, but still slightly below one. To be in line with theory, we fix the coefficients slightly below 1 at 0.9997^a for $a = 1/32$, such that at a daily (8 hours = 32 quarters) the coefficient is comparable to the highest IBM coefficient, b^ξ. Standard error estimates are in round brackets, p-values are in square brackets. Standard errors and p-values are based on a sandwich covariance matrix estimator for the first application, and on a bootstrap procedure for the second application.

	First illustration		Second illustration			
	S&P500	IBM	IT 5y yield		PT 5y yield	
a^ξ	0.025	0.018	0.018	0.023	0.028	0.028
	(0.01)	(0.00)	(0.003)	(0.002)	(0.003)	(0.004)
	[0.00]	[0.00]	[0.000]	[0.000]	[0.000]	[0.000]
a^δ	0.112	0.088	0.079	0.076	0.092	0.089
	(0.01)	(0.01)	(0.003)	(0.003)	(0.004)	(0.004)
	[0.00]	[0.00]	[0.000]	[0.000]	[0.000]	[0.000]
b^ξ	0.9993	0.9997	0.9997^a	0.9997^a	0.9997^a	0.9997^a
	(0.00)	(0.00)				
	[0.00]	[0.00]				
b^δ	0.995	0.994	0.9997^a	0.9997^a	0.9997^a	0.9997^a
	(0.001)	(0.001)				
	[0.00]	[0.00]				
c^ξ				-7.696		-47.855
				(2.225)		(22.555)
				[0.001]		[0.034]
c^δ				0.144		-10.782
				(0.737)		(4.323)
				[0.845]		[0.013]
λ	0	0	0.911	0.911	0.911	0.911
a^τ	0.241	0.260	0.221	0.221	0.773	0.773
b^τ	0.989	0.990	0.999	0.999	0.998	0.998
c^τ			0.017	0.017	-0.141	-0.141
T	14,726	25,028	24,416	24,416	24,576	24,576
T^*	1,495	2,532	2,448	2,448	2,490	2,490
loglik	-22,028.4	-62,727.4	-80,535.8	-80,513.9	-196,068.0	-196,031.0
AIC	44,068.7	125,466.9	161,079.7	161,039.8	392,143.9	392,074.0
BIC	44,114.3	125,515.7	161,112.1	161,088.4	392,176.3	392,122.7

4.1.2 Tail parameter estimates

Figure 5 presents the raw log-returns (top panels) along with filtered estimates of ξ_t and δ_t (middle and bottom panels). The filtered tail shape varies between approximately 0.05 and 0.25 for the S&P 500 index, and between approximately 0.05 and 0.35 for IBM. The filtered tail scales vary roughly between 0.2 and 2.0 for the S&P 500 and IBM. The confidence bands around each filtered parameter suggest that both are reasonably precisely estimated, and that the tail shape parameter is often far from zero. The reported confidence bands are conditional on the estimated thresholds $\hat{\tau}_t$.

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

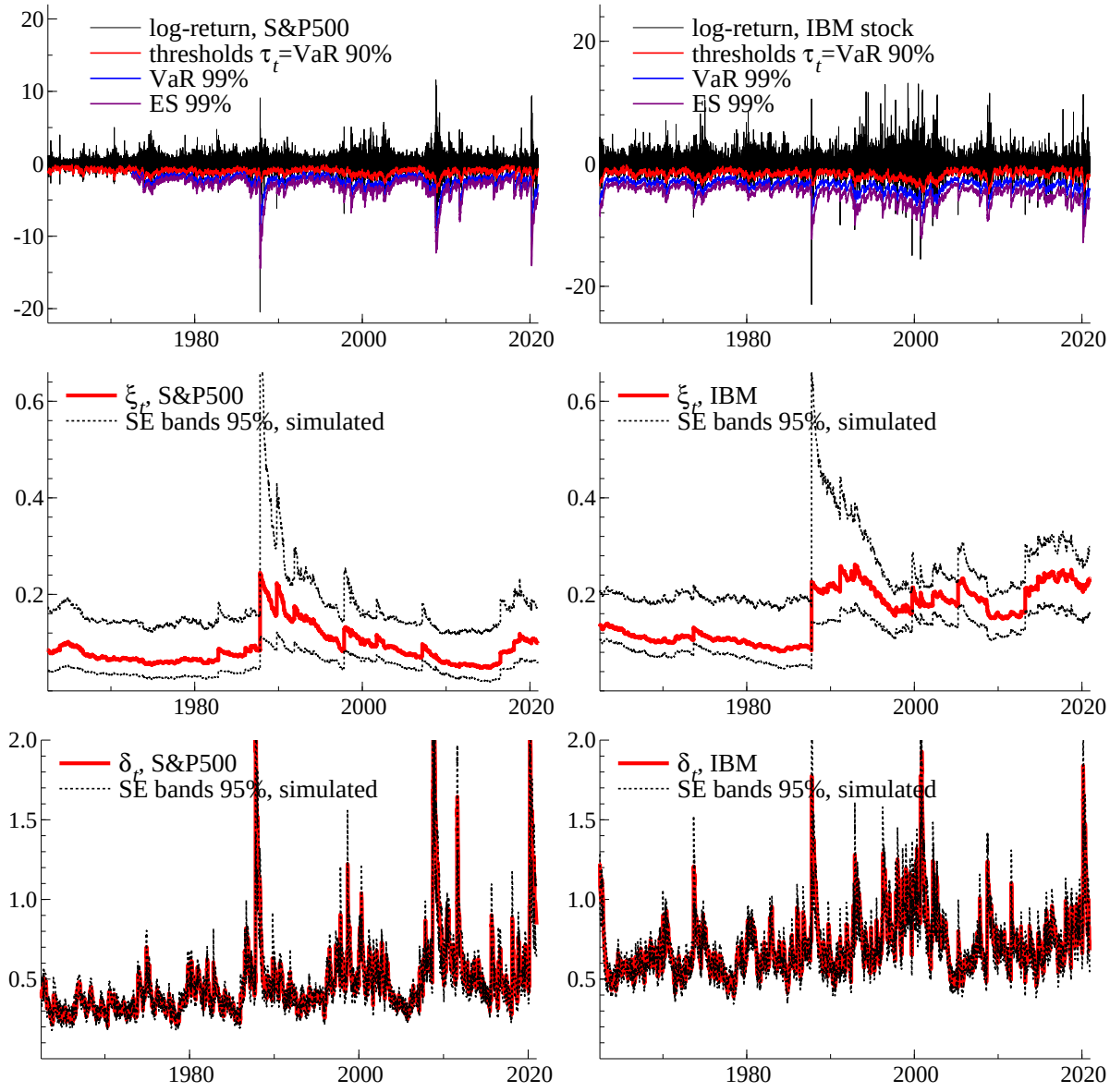
Figure 5 also plots estimates of VaR_t and ES_t at a 99% confidence level. We checked that indeed 1.0% of the log-returns lie beyond $\widehat{\text{VaR}}_t$ in the case of the S&P 500 index (0.9% for IBM stock). The average value of $-y_t$ conditional on it exceeding its VaR is -3.54% for the S&P 500, and -4.85% for IBM. These values are approximately in line with the time series average of $\widehat{\text{ES}}_t$ at -3.12% for the S&P 500, and -4.59% for IBM, respectively.

4.2 Changes in sovereign yields

In our second application we illustrate the inclusion of explanatory variables in the dynamics of ξ_t and δ_t as in equation (8). To do so, we study whether there was any tail risk impact of central bank asset purchases on changes in Italian (IT) and Portuguese (PT) five-year bond yields between 2010 and 2012. Both Italy and Portugal were in the "epicenter" of the existential euro area sovereign debt crisis at that time; see e.g. Eser and Schwaab (2016), and Ghysels et al. (2017). Italy is an example for a large euro area country that was

Figure 5: Filtered tail parameters for S&P500 (left) and IBM (right) log-returns

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



affected by the crisis relatively late (in mid-2011), and that benefited from Eurosystem bond purchases only during a relatively short period of time, between August 2011 and March 2012. Portugal, by contrast, is an example for a smaller euro area country that was affected relatively early (already in 2010), and that benefited from Eurosystem bond purchases more uniformly over time, between May 2010 and March 2012.

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

The bond yields y_t are sampled at the 15-minute frequency, between 9AM and 5PM, and are obtained from Bloomberg. We do not consider overnight changes in yield, such that the first 15-minute interval covers 9AM to 9:15AM. Our sample ranges from 04 January 2010 to 31 December 2012. This yields 32 intra-daily observations per trading day, with $T \approx 3 \times 260 \times 32 \approx 25,000$ observations per country.

Finally, we construct time series data z_t of country-specific bond purchases at the high (15-minute) frequency. 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. Disaggregated data on Eurosystem SMP purchases sampled at a high-frequency are still confidential at the time of writing. At the end of our sample, the Eurosystem held €99.0 bn in Italian sovereign bonds and €21.6 bn in Portuguese bonds; see the ECB Annual Report 2013. We including these as an additional conditioning variable in (8) to see whether they mitigated extreme tail behavior or not.

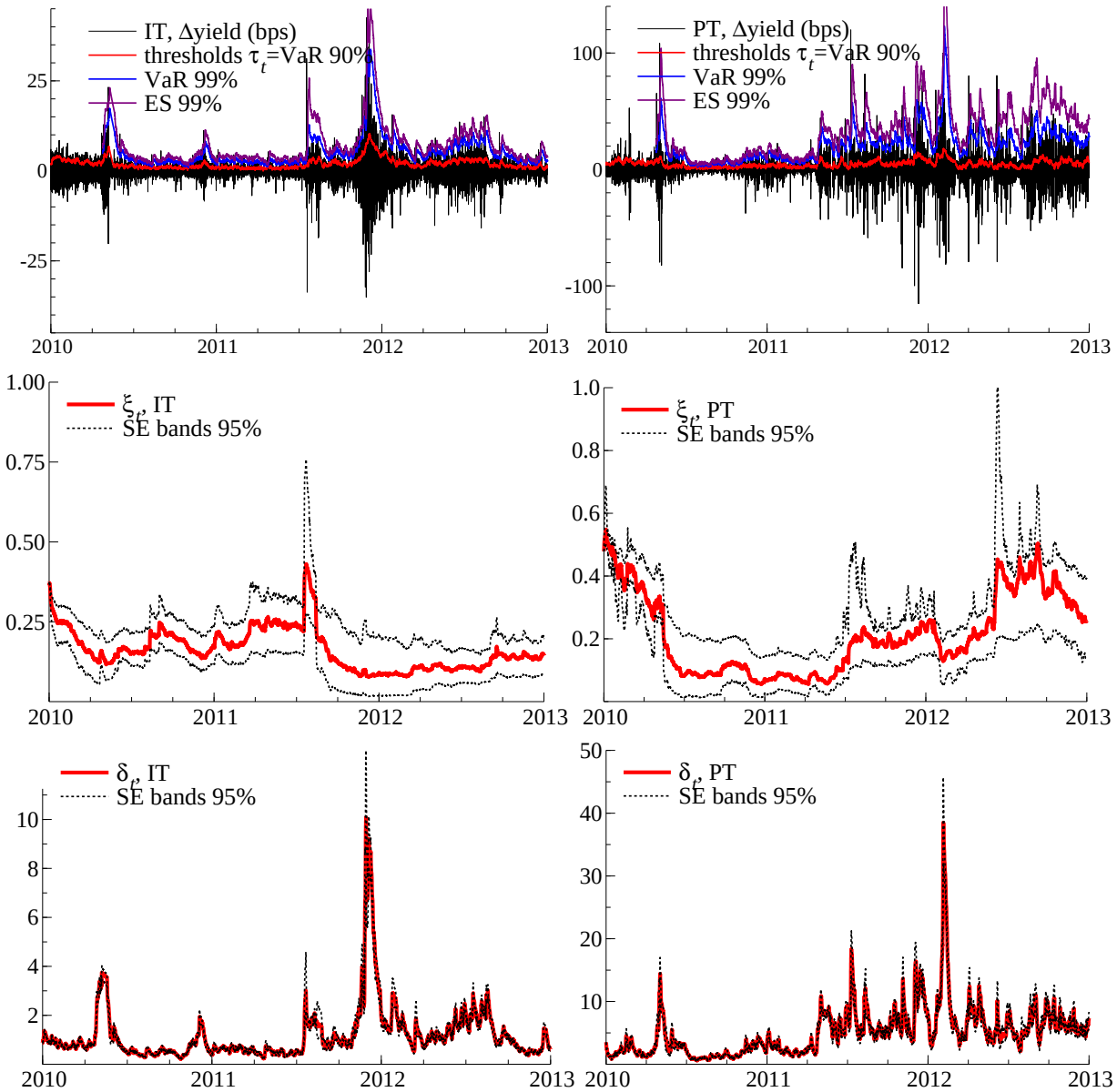
4.2.1 Deterministic parameter estimates

We continue to rely on the time-variation in the thresholds τ_t to control for time-variation in any parameters describing the center of the distribution. The thresholds now evolve according to (10), also taking account of the SMP purchases z_t. We choose $(1-\kappa) = 10\%$, and initialize $\tau_1 = \hat{q}^{0.9}$, the 90% empirical quantile of y_t.

Analyzing changes in the tail shape and tail scale parameters for high-frequency data

Figure 6: The tail shape and tail scale estimates

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



is challenging given the high persistence of parameters at such frequencies. We therefore introduce two simplifications. First, we fix the smoothing parameter λ at $0.05^{1/32} \approx 0.911$, such that 95% of the smoothing materializes within one day (i.e., 32 15-minute intervals). As there are only log-likelihood contribution for the GPD for $x_t = z_t - \tau_t > 0$, it is hard to identify λ empirically. Second, to keep the model in line with the theory from Section 2.2, we fix b^ξ and b^δ to a value close to the unit boundary, as otherwise they are estimated indistinguishably different from 1. We choose $(b^\xi)^{32} = (b^\delta)^{32} = 0.9997$, which implies a similar persistence level as (IBM) equity returns at a daily (= 8h = 32 quarters) frequency. Fixing λ, b^ξ, or b^δ at other reasonable values has little effect on the empirical findings.

Columns three to six of Table 2 present our estimates of the deterministic parameters for the model (1) – (10). Columns three and five refer to a baseline model without central bank purchases. Table 2 reports bootstrapped standard errors for the deterministic parameters using the procedure outlined in Web Appendix K.

The estimates of $a^\tau > 0$ and $b^\tau < 1$ suggest that the thresholds are time-varying and mean-reverting. Parameter c^τ is estimated to be negative for Portuguese bonds, and close to zero for Italian bonds. Parameters a^ξ and a^δ suggest pronounced and statistically significant time series variation in both the tail shape ξ_t and tail scale δ_t parameters, both of which are captured by our time-varying parameter model. The impact parameters c^ξ and c^δ of bond purchases on tail shape and scale are estimated to be negative in both cases. The log-likelihood increases by 21.9 points for IT, and 37.0 points for PT. A comparison of model selection criteria (AIC, BIC) further supports the inclusion of central bank asset purchases as a useful covariate to explain each time series' extreme tail dynamics.

4.2.2 Tail parameter estimates and VaR impact

Figure 6 plots the corresponding filtered estimates for time-varying tail shape ξ_t and tail scale δ_t. Time series variation is present and pronounced in both parameters. The filtered tail shape varies between approximately 0.1 and 0.4 for Italian yields, and between 0.05 and 0.6 for Portuguese yields. The filtered tail scale varies between approximately 1 and 10.0 for Italian yields, and between approximately 1 and 40 for Portuguese yields. The standard

error bands around each time-varying parameter suggest that both parameters are estimated reasonably precisely, and that the tail shape is often far from the Gumbel case of $\xi_t = 0$.

As was clear from Table 2, the coefficients c^{ξ} and c^{δ} measure the impact of bond purchases on the tail behavior of yield changes. As these parameters are difficult to interpret by themselves, we show in Web Appendix L how they can be translated into an impact on VaR via their link to τ_t, δ_t, and ξ_t. For Italy, we obtain a total VaR impact of 0.0 (tail threshold) $+0.8$ (tail scale) $-$ 5.9 (tail shape) $= -5.1$ bps for a €1 bn Eurosystem intervention. For Portugal, we obtain a larger impact of -0.1 (tail threshold) $-$ 172.5 (tail scale) $-$ 176.4 (tail shape) $= -349.0$ bps. These point estimates are of course subject to substantial estimation uncertainty; see Table 2. The 95% confidence intervals for VaR impact can be bootstrapped along with the parameters. They equal $[-17.6, 11.6]$ for IT and $[-664.0, -38.8]$ for PT. The stronger impact for Portugal than for Italy is likely due to a €1 bn intervention constituting a larger share of the overall market.

5 Conclusion

We introduced a semi-parametric modeling framework to study time variation in tail parameters for long univariate time series. To this end we modeled the time variation in the shape and scale parameters of the Generalized Pareto Distribution, which approximates the tail of most heavy-tailed densities used in econometrics and the actuarial sciences. We discussed the handling of non-tail time series observations, inference on deterministic and time-varying parameters, and how to relate tail variation to observed covariates if such variables are available. We also established conditions for stationarity and ergodicity of the model and conditions for consistency and asymptotic normality of the maximum likelihood estimator. The model therefore complements and extends recent work based on different methodologies, such as the non-parametric approach to tail index variation of de Haan and Zhou (2021), the time-varying quantile (and ES) approaches of Patton et al. (2019) and Catania and Luati (2023), and the parametric approach of Massacci (2017). We applied the model to study variation in the left tail of U.S. equity log-returns, and in the right tail of

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

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

References

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

Blasques, F., J. van Brummelen, S. J. Koopman, and A. Lucas (2022). Maximum likelihood estimation for Generalized Autoregressive Score models. *Journal of Econometrics 227*(2), 325–346.

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

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

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

Catania, L. and A. Luati (2023). Semiparametric modeling of multiple quantiles. *Journal of Econometrics*, (in press).

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

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

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

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

Creal, D., 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.

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

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

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

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

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

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

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.

Hafner, C. M. and L. Wang (2023). A dynamic conditional score model for the log correlation matrix. *Journal of Econometrics*, (in press).

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

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

Hetland, S., R. S. Pedersen, and A. Rahbek (2023). Dynamic conditional eigenvalue GARCH. *Journal of Econometrics*, (in 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 & Economic Statistics 19*(1), 208–216.

Jensen, S. T. and A. Rahbek (2004). Asymptotic Inference for nonstationary GARCH. *Econometric Theory 20*(6), 1203–1226.

Johnson, N. L., S. Kotz, and N. Balakrishnan (1994). *Continuous univariate distributions, Vol. 1, 2nd Edition*. Wiley.

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

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

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

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

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

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

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

Mhalla, L., M. de Carvalho, and V. ChavezDemoulin (2019). Regressiontype models for extremal dependence. *Scandinavian Journal of Statistics 46*(4), 1141–1167.

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

Patton, A. (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.

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.

Straumann, D. and T. Mikosch (2006). Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: A stochastic recurrence equations approach. *The Annals of Statistics 34*(5), 2449–2495.

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

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

Web Appendix to
Modeling extreme events:
time-varying extreme tail shape

Contents:

A Appendix: GPD score and scaling functions

A.1 The GPD score function

This section derives the score (4). 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 (6). To this end we require the $[2 \times 2]$ conditional Fisher information matrix associated with (4),

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

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

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

A.3 The GAS(2,1) dynamics for EWMA scheme

Specification (7) leads to a specification for f_t as in

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

where L is the lag operator, such that f_t follows the score-driven dynamics in (3) with $p = 2$ and $q = 1$, also known as GAS(2,1) dynamics. To see this, first rewrite the first equation in (7) as

$$(\mathrm{I}_2 - B\,L) f_{t+1} = \omega + A\tilde{s}_t,$$

and then multiply both sides by $(1 - \lambda\,L)/(1 - \lambda)$, using the second equation in (7)

$$(1 - \lambda\,L)\tilde{s}_t = (1 - \lambda)s_t.$$

to replace $(1 - \lambda\,L)\tilde{s}_t/(1 - \lambda)$ by s_t.

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_t be a threshold. In this case, the tail approximation is exact, as we have

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

for $\xi_t = \alpha_t^{-1}$ and $\delta_{t,\tau_t} = \sigma_t + \alpha_t^{-1}\tau_t$. The EVT GPD tail 'approximation' has the same tail as the original GPD, but a higher scale parameter $\sigma_t + \tau_t/\alpha_t$ rather than σ_t. This is intuitive, as the GPD beyond a high threshold has a flatter tail than at the origin. The scale parameter $\delta_{t,\tau_t} = \sigma_t + \alpha_t^{-1}\tau_t$ increases with the threshold τ_t, 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 δ_{t,τ_t} may still vary over time to reflect time-variation in α_t.

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

Let $y_t \sim \text{t}(0, \sigma_t^2, \alpha_t)$ 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 $\tau_t \in \mathbb{R}$ be a threshold.

In the simulations, we minimize the Kullback-Leibler divergence between the Student's t tail and the GPD tail approximation. Analytically, we can use the following approximate solution. The rate of tail decline in the extreme tail of the Student's t and the GPD should coincide, implying a tail shape of $\xi_t^{-1} = \alpha_t$. For the scale, we equate the slope at the origin of the GPD with that of

the Student's t at τ_t and obtain

$$\delta_{t,\tau_t}^{-1} = \frac{f(\tau_t)}{1-F(\tau_t)} \overset{\tau_t \to \infty}{\approx} \frac{-f'(\tau_t)}{f(\tau_t)} = \frac{\partial - \ln f(\tau_t)}{\partial \tau_t} = \frac{(1+\alpha_t^{-1})\tau_t/\sigma_t^2}{1+\tau_t^2/(\alpha_t\sigma_t^2)} = \frac{(1+\alpha_t)\tau_t}{\alpha_t\sigma_t^2+\tau_t^2} \Leftrightarrow$$
$$\delta_{t,\tau_t} \sim \frac{\alpha_t\sigma_t^2}{(1+\alpha_t)\tau_t} + \frac{\tau_t}{1+\alpha_t},$$

which again depends on α_t and increases in τ_t. For large τ_t, δ_{t,τ_t} varies inversely with α_t, or positively with $\xi_t = \alpha_t^{-1}$. As a result, we should not expect δ_t to coincide with σ_t.

C Proof of stationarity, ergodicity and moments of f_t and x_t

C.1 Proof of Theorem 1

From the cdf of the GPD, together with Assumption 1, we know that for a standard uniform u_t and for a standard exponentially-distributed ϵ_t, we have

$$\begin{aligned} u_t &= 1 - (1 + \xi_t x_t/\delta_t)^{-\xi_t^{-1}} \quad \Leftrightarrow \\ -\ln(1-u_t) &= \epsilon_t = \xi_t^{-1} \cdot \ln(1 + \xi_t x_t/\delta_t) \quad \Rightarrow \\ \exp(\xi_t \epsilon_t) &= 1 + \xi_t x_t/\delta_t. \end{aligned}$$

Filling this out into the scaled score expressions evaluated at $\theta_0 \in \Theta$, we get

$$\begin{aligned} s_t^{\xi} &= \xi_t^{-2}(1+\xi_t)\ln(1+\xi_t x_t/\delta_t) + \frac{1-(1+3\xi_t^{-1}+\xi_t^{-2})\xi_t x_t/\delta_t}{1+\xi_t x_t} \\ &= \xi_t^{-1}(1+\xi_t)\ \epsilon_t + \frac{1-(1+3\xi_t^{-1}+\xi_t^{-2})(\exp(\xi_t\epsilon_t)-1)}{\exp(\xi_t\epsilon_t)} \\ &= (1+\xi_t^{-1})\ \epsilon_t + \exp(-\xi_t\epsilon_t) - (1+3\xi_t^{-1}+\xi_t^{-2})(1-\exp(-\xi_t\epsilon_t)) \end{aligned}$$

and

$$\begin{aligned} s_t^{\delta} &= \sqrt{1+2\xi_t}\frac{x_t-\delta_t}{\delta_t+\xi_t x_t} \\ &= \sqrt{1+2\xi_t}\xi_t^{-1}\frac{\xi_t x_t/\delta_t}{1+\xi_t x_t/\delta_t} - \sqrt{1+2\xi_t}\frac{1}{1+\xi_t x_t/\delta_t} \\ &= \sqrt{1+2\xi_t}\xi_t^{-1}\frac{\exp(\xi_t\epsilon_t)-1}{\exp(\xi_t\epsilon_t)} - \sqrt{1+2\xi_t}\frac{1}{\exp(\xi_t\epsilon_t)} \\ &= \sqrt{1+2\xi_t}\xi_t^{-1}(1-\exp(-\xi_t\epsilon_t)) - \sqrt{1+2\xi_t}\exp(-\xi_t\epsilon_t). \end{aligned}$$

Now, in order to apply Theorem 3.1 of Bougerol (1993), we first verify the required log-moment conditions for our bivariate score-driven process $\{f_t\}_{t\in\mathbb{Z}}$. We recall that from (11), $f_{t+1} = \Phi_t(f_t;\theta_0)$,

where

$$\Phi_t\left(f_t;\theta_0\right)=\begin{pmatrix}\omega^{\xi}\\ \omega^{\delta}\end{pmatrix}+\begin{pmatrix}a^{\xi} & 0\\ 0 & a^{\delta}\end{pmatrix}\begin{pmatrix}s_t^{\xi}\\ s_t^{\delta}\end{pmatrix}+\begin{pmatrix}b^{\xi} & 0\\ 0 & b^{\delta}\end{pmatrix}\begin{pmatrix}f_t^{\xi}\\ f_t^{\delta}\end{pmatrix}, \tag{C.1}$$

such that, by using Lemma 2.2 of Straumann and Mikosch (2006) for any fixed $f_0\in\mathbb{R}^2$, we obtain

$$\begin{aligned}\mathbb{E}\left[\ln^{+}\left\|\Phi_t\left(f_0;\theta_0\right)-f_0\right\|\right]\leq & 2\ln 2+\ln^{+}\left\|\begin{matrix}\omega^{\xi}\\ \omega^{\delta}\end{matrix}\right\|+\ln^{+}\left\|\begin{matrix}a^{\xi} & 0\\ 0 & a^{\delta}\end{matrix}\right\|+\mathbb{E}\left[\ln^{+}\left\|\begin{matrix}s_t^{\xi}\\ s_t^{\delta}\end{matrix}\right\|\right]\\ &+\ln^{+}\left\|\begin{matrix}b^{\xi}-1 & 0\\ 0 & b^{\delta}-1\end{matrix}\right\|+\ln^{+}\left\|\begin{matrix}f_0^{\xi}\\ f_0^{\delta}\end{matrix}\right\|.\end{aligned}$$

It therefore suffices to verify that the scaled score vector s_t has a finite log-moment. We have that

$$\mathbb{E}\left[\ln^{+}\left\|\begin{matrix}s_t^{\xi}\\ s_t^{\delta}\end{matrix}\right\|\right]\leq\mathbb{E}\left[\ln^{+}\left\|\begin{matrix}\left(1+\xi_0^{-1}\right)\epsilon_t+\exp(-\xi_0\epsilon_t)-(1+3\xi_0^{-1}+\xi_0^{-2})(1-\exp(-\xi_0\epsilon_t))\\ \sqrt{1+2\xi_0}\xi_0^{-1}\left(1-\exp(-\xi_0\epsilon_t)\right)-\sqrt{1+2\xi_0}\exp(-\xi_0\epsilon_t)\end{matrix}\right\|\right]<\infty,$$

since $\xi_0\in\mathbb{R}^{+}$ is a fixed real-valued point and, by Assumption 1, we already know that the process ϵ_t is i.i.d. and follows an exponential distribution with unit scale. Hence, we conclude that the required log-moment conditions are satisfied.

Second, we show how the contraction condition in Assumption 2 is derived. For our bivariate process $\{f_t\}_{t\in\mathbb{Z}}$ it is easy to have the analytical form of the random matrix $\dot{\Phi}_t\left(f_t;\theta_0\right)$, which can be retrieved by directly taking the first partial derivatives of the mapping in equation (C.1) with respect to $\left(f_t^{\xi},f_t^{\delta}\right)'\in\mathbb{R}^2$. Moreover, to motivate the imposed contraction condition for this theorem, we repeatedly substitute the stationary and ergodic $\tilde{f}_t=\Phi_t\left(\tilde{f}_t;\theta_0\right)$ and get

$$\tilde{f}_t=\Phi_t\left(\tilde{f}_t;\theta_0\right)=\Phi\left(\epsilon_t,\tilde{f}_t;\theta_0\right)=\Phi\left(\epsilon_t,\Phi_{t-1}\left(\Phi_{t-2}\left(\ldots,\theta_0\right),\theta_0\right),\theta_0\right). \tag{C.2}$$

Using the chain rule, we then obtain

$$\frac{\partial\Phi_t\left(f_t;\theta_0\right)}{\partial f_{t-r}'}=\prod_{i=1}^{r-1}\dot{\Phi}_{t-i}\left(f_{t-i};\theta_0\right)\frac{\partial\Phi_{t-r}\left(f_{t-r},\theta_0\right)}{\partial f_{t-r}'}=\prod_{i=1}^{r}\dot{\Phi}_{t-i}\left(f_{t-i};\theta_0\right).$$

By Assumption 2, there exist some sufficiently large $r\geq 1$ number after which the bivariate process $\{f_t\}_{t\in\mathbb{Z}}$ is contracting. Therefore, by the mean value theorem together with Assumption 2, it follows

that

$$\begin{aligned}\left\|f_t - \tilde{f}_t\right\| =& \Big\| \Phi\left(\epsilon_t, \Phi_{t-1}\left(\Phi_{t-2}\left(\ldots, \Phi_{t-r}\left(f_{t-r}, \theta_0\right), \theta_0\right), \theta_0\right), \theta_0\right) \\ & - \Phi\left(\epsilon_t, \Phi_{t-1}\left(\Phi_{t-2}\left(\ldots, \Phi_{t-r}\left(\tilde{f}_{t-r}, \theta_0\right), \theta_0\right), \theta_0\right), \theta_0\right) \Big\| \\ \leq & \left\| \prod_{i=1}^{r} \dot{\Phi}_{t-i}\left(\bar{f}_{t-i}; \theta_0\right) \right\| \left\| f_{t-r} - \tilde{f}_{t-r} \right\|,\end{aligned}$$

where $\bar{f}_t$ is on the chord between f_t and $\tilde{f}_t$. Then, as shown in Theorem 3.1 of Bougerol (1993), using Assumption 2 we have

$$\left\| \prod_{i=1}^{r} \dot{\Phi}_{t-i}\left(\bar{f}_{t-i}; \theta_0\right) \right\| \xrightarrow{e.a.s.} 0, \qquad r \to \infty,$$

and since $\mathbb{E}\left[\ln^+ \left\|f_t - \tilde{f}_t\right\|\right] = \mathbb{E}\left[\ln^+ \left\|\Phi_{t-1}\left(f_{t-1}; \theta_0\right) - \tilde{f}_t\right\|\right] < \infty$ for any $f_t \in \mathbb{R}^2$ as proved above, it follows that $\left\|f_t - \tilde{f}_t\right\| \xrightarrow{e.a.s.} 0$ by a straightforward application of Lemma 2.1 of Straumann and Mikosch (2006). Finally, the stationarity and ergodicity of x_t then follows easily from Krengel (2011) and the fact that

$$x_t = \xi_t^{-1}\delta_t \cdot \left(\exp(\xi_t \epsilon_t) - 1\right).$$

This proves the theorem. ∎

C.2 Proof of Theorem 2

Consider the mapping in equation (C.1). By the Minkowsky's inequality and the mean value theorem, we have

$$\begin{aligned}\left\|\Phi_t\left(f_t, \theta_0\right)\right\|^p \leq & \left\|\Phi_t\left(f_t, \theta_0\right) - \Phi_t\left(\tilde{f}_t, \theta_0\right)\right\|^p + \left\|\Phi_t\left(\tilde{f}_t, \theta_0\right)\right\|^p \\ = & \left\|\dot{\Phi}_t\left(\bar{f}_t, \theta_0\right)\left(\Phi_{t-1}\left(f_{t-1}, \theta_0\right) - \Phi_{t-1}\left(\tilde{f}_{t-1}, \theta_0\right)\right)\right\|^p + \left\|\Phi_t\left(\tilde{f}_t, \theta_0\right)\right\|^p,\end{aligned}$$

where $\bar{f}_t$ is on the chord between f_t and $\tilde{f}_t$. Therefore, by similar arguments as the ones used in the proof of Theorem 1, we get

$$\|\Phi_t(f_t,\theta_0)\|^p \leq \left\|\prod_{i=1}^{r}\dot{\Phi}_{t-i}\left(\bar{f}_{t-i-1};\theta_0\right)\right\|^p \times \left\|\left(\Phi_{t-i-1}(f_{t-i-1},\theta_0)-\Phi_{t-i-1}\left(\tilde{f}_{t-i-1},\theta_0\right)\right)\right\|^p + \left\|\Phi_t\left(\tilde{f}_t,\theta_0\right)\right\|^p.$$

Using the condition stated in (14), we can then take expectations and write for all $r \geq 1$

$$\mathbb{E}\left[\|\Phi_t(f_t,\theta_0)\|^p\right] \leq \beta\mathbb{E}\left[\left\|\Phi_{t-r-1}(f_{t-r-1},\theta_0)-\Phi_{t-r-1}\left(\tilde{f}_{t-r-1},\theta_0\right)\right\|^p\right] + \mathbb{E}\left[\left\|\Phi_t\left(\tilde{f}_t,\theta_0\right)\right\|^p\right] \quad \text{(C.3)}$$

$$\leq \beta^{t-r}\mathbb{E}\left[\left\|\Phi_0(f_0,\theta_0)-\Phi_0\left(\tilde{f}_0,\theta_0\right)\right\|^p\right] + \sum_{j=1}^{t-r}\beta^{j-1}\mathbb{E}\left[\left\|\Phi_t\left(\tilde{f}_{t-j+1},\theta_0\right)\right\|^p\right]. \quad \text{(C.4)}$$

Since from (14) we have $\beta \in (0,1)$, the first term on the right-hand side of the inequality in (C.3) converges to 0 as $t \to \infty$. Moreover, we note that $\left\{\Phi_t\left(\tilde{f}_{t-j+1},\theta_0\right)\right\}_{t\in\mathbb{Z}}$ is stationary and ergodic, and hence, the right-hand side of the inequality in (C.3) will eventually converge as $t \to \infty$ if and only if $\mathbb{E}\left[\|\Phi_0(f_0,\theta_0)\|^p\right] < \infty$ for any fixed $f_0 \in \mathbb{R}^2$. It is also easy to see that, as $f_0 \in \mathbb{R}^2$ is a fixed point, and the moments bound $\mathbb{E}\left[\|\Phi_0(f_0,\theta_0)\|^p\right] < \infty$ is implied by $\mathbb{E}\left[\|s_0\|^p\right] < \infty$. We thus have

$$\mathbb{E}\left[\left\|\begin{matrix} s_0^{\xi} \\ s_0^{\delta}\end{matrix}\right\|^p\right] \leq \mathbb{E}\left[\left\|\begin{matrix} (1+\xi_0^{-1})\,\epsilon_0 + \exp(-\xi_0\epsilon_0) - (1+3\xi_0^{-1}+\xi_0^{-2})(1-\exp(-\xi_0\epsilon_0)) \\ \sqrt{1+2\xi_0}\xi_0^{-1}\left(1-\exp(-\xi_0\epsilon_0)\right) - \sqrt{1+2\xi_0}\exp(-\xi_0\epsilon_0)\end{matrix}\right\|^p\right] < \infty,$$

which directly follows from the fact that $\xi_0 \in \mathbb{R}^+$ is fixed, and that ϵ_0 is i.i.d. and exponentially distributed with unit scale, as implied by Assumption 1.

This proves the theorem. ∎

D Numerically computing SE and finite-moments regions

To compute the SE and finite-moments regions in Figure 2, we proceed as follows. First, we fix the deterministic parameters at $\hat{\theta}$, one of the empirical estimates, and vary only two of its elements over a finely grained mesh of points. Second, for each of these points on the mesh, we approximate the expectation by a Monte Carlo average. For this, we generate a long sequence of N i.i.d. unit exponential random variables and construct $f_{i,r} = f_{i,r}(f, \epsilon_i, \ldots, \epsilon_{i+r-1}, \hat{\theta})$ as the r-th forward iterate of

$$f_{i,r} = \omega + B_1 f_{i,r-1} + A_1 \cdot \begin{bmatrix} (1+\xi_{i,r}^{-1}) \cdot \epsilon_{i+r-1} + \frac{1-(1+3\xi_{i,r}^{-1}+\xi_{i,r}^{-2})(\exp(\xi_{i,r}\epsilon_{i+r-1})-1)}{\exp(\xi_{i,r}\epsilon_{i+r-1})} \\ \\ \sqrt{1+2\xi_{i,r}} \ \frac{\xi_{i,r}^{-1}(\exp(\xi_{i,r}\epsilon_{i+r-1})-1)-1}{\exp(\xi_{i,r}\epsilon_{i+r-1})} \end{bmatrix}$$

$$= \omega + B_1 f_{i,r-1} + A_1 \cdot \begin{bmatrix} (1+\xi_{i,r}^{-1}) \cdot \epsilon_{i+r-1} + e^{-\xi_{i,r}\epsilon_{i+r-1}} - (1+3\xi_{i,r}^{-1}+\xi_{i,r}^{-2}) \cdot \left(1-e^{-\xi_{i,r}\epsilon_{i+r-1}}\right) \\ \\ \sqrt{1+2\xi_{i,r}} \ \left(\xi_{i,r}^{-1} - (1+\xi_{i,r}^{-1}) \cdot e^{-\xi_{i,r}\epsilon_{i+r-1}}\right) \end{bmatrix},$$

with $f_{i,r} = (\ln \xi_{i,r}, \ln \delta_{i,r})'$, the deterministic parameters set to the values in $\hat{\theta}$, and every $f_{i,r}$ starting from the same initial $f_{i,0} \equiv f$. This SRE corresponds to the DGP and is optained by substituting $x_t = \delta \cdot \xi_t^{-1} \cdot (\exp(\xi_t \epsilon_t) - 1)$ into (7). Note that the score part in this DGP SRE expression no longer depends on $\delta_{i,r}$.

For every $i = 1, \ldots, N-r+1$ and every $r = 1, \ldots, r^{\max}$, we then compute the maximum of

$$\left\| \prod_{j=1}^{r} \dot{\Phi}_{i+j-1}\left(f_{i,j-1}; \hat{\theta}\right) \right\|$$

with respect to the initial value f. The optimization is carried out numerically using a bivariate grid of f values, yielding the maximizer $\hat{f}_{i,r}$. This is repeated for every mesh point for $\hat{\theta}$. Finally, we determine the smallest number $\hat{r}$ such that

$$\hat{r} = \min \left\{ r = 1, \ldots, r^{\max} \;\middle|\; \frac{1}{N-r+1} \sum_{i=1}^{N-r+1} \ln \left\| \prod_{j=1}^{r} \dot{\Phi}_{i+j-1}\left(\hat{f}_{i,r}; \hat{\theta}\right) \right\| < 0 \right\},$$

or

$$\hat{r} = \min \left\{ r = 1, \ldots, r^{\max} \;\middle|\; \frac{1}{N-r+1} \sum_{i=1}^{N-r+1} \left\| \prod_{j=1}^{r} \dot{\Phi}_{i+j-1} \left(\hat{f}_{i,r}; \hat{\theta} \right) \right\|^{p} < 1 \right\},$$

depending on whether we want to visualize the SE or the finite-moments region.

The plot is then the contour plot of $\hat{r}$.

The matrix products of $\dot{\Phi}$ can be numerically unstable for small values of ξ_t. To resolve these instabilities, we substitute the analytical expressions in these cases by Taylor-series expansions of $\exp(-\xi_t \varepsilon_t)$, combining terms and removing terms of the order ξ_t^a for $a < 0$ that cancel. Only afterwards, we then compute the numerical result. As a cutoff, we take $\xi_t < 10^{-5}$. Around this point, the numerical calculations with or without the approximation give the same results.

E Consistency and Asymptotic Normality of the MLE

In this appendix, we first derive the partial derivatives of the log-likelihood function and our score-driven process up to third-order. We then discuss the proof for consistency and asymptotic normality of the MLE.

E.1 Derivatives of the log-likelihood function

For a generic $\theta = \left((\theta^{\xi})', (\theta^{\delta})'\right)'$, where $\theta^{\xi} = \left(\omega^{\xi}, a^{\xi}, b^{\xi}\right)$ and $\theta^{\delta} = \left(\omega^{\delta}, a^{\delta}, b^{\delta}\right)$, denote with $I_N \in \mathbb{R}^{N\times N}$ the $N \times N$ identity matrix and $\mathcal{K}_{NN}$ the commutation matrix.

We first recall that the log-likelihood function evaluated at $\theta \in \Theta$ is defined as

$$\mathcal{L}(\theta|\mathcal{F}_T) = \frac{1}{T^*}\sum_{t=1}^{T}\left(-\ln\left(\delta_t(\theta)\right) - \left(1+\frac{1}{\xi_t(\theta)}\right)\ln\left(1+\xi_t(\theta)\frac{x_t}{\delta_t(\theta)}\right)\right).$$

Then, by direct calculations, we obtain the score vector

$$\frac{\partial\mathcal{L}(\theta|\mathcal{F}_T)}{\partial\theta} = \frac{1}{T^*}\sum_{t=1}^{T}\left(\frac{\partial f_t(\theta)}{(\partial\theta)'}\right)'\nabla_t = \frac{1}{T^*}\sum_{t=1}^{T}\begin{pmatrix}\nabla_t^{\xi}(\theta)\frac{\partial f_t(\theta)}{\partial\theta^{\xi}}\\ \nabla_t^{\delta}(\theta)\frac{\partial f_t(\theta)}{\partial\theta^{\delta}}\end{pmatrix}, \tag{E.1}$$

where

$$\nabla_t(\theta) = \begin{pmatrix}\nabla_t^{\xi}(\theta)\\ \nabla_t^{\delta}(\theta)\end{pmatrix} = \begin{pmatrix}\xi_t^{-1}(\theta)\ln\left(1+\xi_t(\theta)\,\delta_t^{-1}(\theta)x_t\right) - \left(1+\xi_t^{-1}(\theta)\right)\frac{\xi_t(\theta)x_t}{\delta_t(\theta)+\xi_t(\theta)x_t}\\ \frac{x_t-\delta_t(\theta)}{\delta_t(\theta)+\xi_t(\theta)x_t}\end{pmatrix}. \tag{E.2}$$

We also have the Hessian matrix

$$\begin{aligned}\frac{\partial^2\mathcal{L}(\theta|\mathcal{F}_T)}{(\partial\theta)(\partial\theta)'} &= \frac{1}{T^*}\sum_{t=1}^{T}\left(\left(\frac{\partial f_t(\theta)}{(\partial\theta)'}\right)'\nabla_t^2(\theta)\frac{\partial f_t(\theta)}{(\partial\theta)'} + \left(\nabla_t(\theta)'\otimes I_6\right)\frac{\partial\,\mathrm{vec}}{(\partial\theta)'}\left(\frac{\partial f_t(\theta)}{(\partial\theta)'}\right)\right) \qquad \text{(E.3)}\\ &= \frac{1}{T^*}\sum_{t=1}^{T}\begin{pmatrix}\nabla_t^{\xi\xi}(\theta)\frac{\partial f_t^{\xi}(\theta)}{(\partial\theta^{\xi})}\frac{\partial f_t^{\xi}(\theta)}{(\partial\theta^{\xi})'} + \nabla_t^{\xi}(\theta)\frac{\partial^2 f_t^{\xi}(\theta)}{(\partial\theta^{\xi})(\partial\theta^{\xi})'} & \nabla_t^{\xi\delta}(\theta)\frac{\partial f_t^{\xi}(\theta)}{(\partial\theta^{\xi})}\frac{\partial f_t^{\delta}(\theta)}{(\partial\theta^{\delta})'} + \nabla_t^{\xi}(\theta)\frac{\partial^2 f_t^{\xi}(\theta)}{(\partial\theta^{\xi})(\partial\theta^{\delta})'}\\ \nabla_t^{\delta\xi}(\theta)\frac{\partial f_t^{\delta}(\theta)}{(\partial\theta^{\delta})}\frac{\partial f_t^{\xi}(\theta)}{(\partial\theta^{\xi})'} + \nabla_t^{\delta}(\theta)\frac{\partial^2 f_t^{\delta}(\theta)}{(\partial\theta^{\delta})(\partial\theta^{\xi})'} & \nabla_t^{\delta\delta}(\theta)\frac{\partial f_t^{\delta}(\theta)}{(\partial\theta^{\delta})}\frac{\partial f_t^{\delta}(\theta)}{(\partial\theta^{\delta})'} + \nabla_t^{\delta}(\theta)\frac{\partial^2 f_t^{\delta}(\theta)}{(\partial\theta^{\delta})(\partial\theta^{\delta})'}\end{pmatrix},\end{aligned}$$

with

$$\nabla_t^2(\theta) = \begin{pmatrix} \nabla_t^{\xi\xi}(\theta) & \nabla_t^{\xi\delta}(\theta) \\ \nabla_t^{\delta\xi}(\theta) & \nabla_t^{\delta\delta}(\theta) \end{pmatrix}, \tag{E.4}$$

$$= \begin{pmatrix} \frac{(\delta_t(\theta)-\xi_t(\theta)\delta_t(\theta)+2\xi_t(\theta)x_t)x_t}{(\delta_t(\theta)+\xi_t(\theta)x_t)^2} - \xi_t^{-1}(\theta)\ln\left(1+\xi_t(\theta)\delta_t^{-1}(\theta)x_t\right) & \xi_t(\theta)x_t\frac{\delta_t(\theta)-x_t}{(\delta_t(\theta)+\xi_t(\theta)x_t)^2} \\ \xi_t(\theta)x_t\frac{\delta_t(\theta)-x_t}{(\delta_t(\theta)+\xi_t(\theta)x_t)^2} & -(1+\xi_t(\theta))\frac{\delta_t(\theta)x_t}{(\delta_t(\theta)+\xi_t(\theta)x_t)^2} \end{pmatrix}.$$

Additionally, the third-order derivatives of the log-likelihood function can be expressed as

$$\frac{\partial \operatorname{vec}\partial^2\mathcal{L}(\theta|\mathcal{F}_T)}{(\partial\theta)'(\partial\theta)(\partial\theta)'} = \frac{\partial \operatorname{vec}}{(\partial\theta)'}\left(\frac{\partial^2\mathcal{L}(\theta|\mathcal{F}_T)}{(\partial\theta)(\partial\theta)'}\right) := Q_t(\theta), \tag{E.5}$$

where matrix $Q_t(\theta) \in \mathbb{R}^{12\times 6}$ collects the third-order derivatives of the log-likelihood function. A typical element of the matrix $Q_t(\theta)$ is given by $Q_t^{kkk}(\theta)$, for $k \in \{\xi, \delta\}$, and takes the form

$$\begin{aligned} Q_t^{kkk}(\theta) =& \nabla_t^{kkk}(\theta)\operatorname{vec}\left(\frac{\partial f_t^k(\theta)}{(\partial\theta^k)}\frac{\partial f_t^k(\theta)}{(\partial\theta^k)'}\right)\frac{\partial f_t^k(\theta)}{(\partial\theta^k)'} + \nabla_t^{kk}(\theta)\operatorname{vec}\left(\frac{\partial^2 f_t^k(\theta)}{(\partial\theta^k)(\partial\theta^k)'}\right)\frac{\partial f_t^k(\theta)}{(\partial\theta^k)'} \\ &+ \nabla_t^{kk}(\theta)\left(I_6+\mathcal{K}_{66}\right)\left(\frac{\partial f_t^k(\theta)}{(\partial\theta^k)}\otimes I_6\right)\frac{\partial^2 f_t^k(\theta)}{(\partial\theta^k)(\partial\theta^k)'} + \nabla_t^k(\theta)\frac{\partial \operatorname{vec}}{(\partial\theta^k)'}\left(\frac{\partial^2 f_t^k(\theta)}{(\partial\theta^k)(\partial\theta^k)'}\right), \end{aligned} \tag{E.6}$$

where

$$\nabla_t^{\xi\xi\xi}(\theta) = \xi_t^{-1}(\theta)\ln\left(1+\xi_t(\theta)\delta_t^{-1}(\theta)x_t\right) - \frac{\left(\delta_t^2(\theta)+\xi_t^2(\theta)x_t\left(3x_t-\delta\right)+\delta_t(\theta)\xi_t(\theta)\left(\delta_t(\theta)+2x_t\right)\right)x_t}{\left(\delta_t(\theta)+\xi_t(\theta)x_t\right)^3},$$

$$\nabla_t^{\delta\delta\delta}(\theta) = (1+\xi_t(\theta))\frac{\delta_t(\theta)x_t\left(\delta_t(\theta)-\xi_t(\theta)x_t\right)}{\left(\delta_t(\theta)+\xi_t(\theta)x_t\right)^3},$$

$$\nabla_t^{\xi\xi\delta}(\theta) = \nabla_t^{\xi\delta\xi}(\theta) = \xi_t(\theta)x_t\frac{\left(\delta_t(\theta)-x_t\right)\left(\delta_t(\theta)-x_t\xi_t(\theta)\right)}{\left(\delta_t(\theta)+\xi_t(\theta)x_t\right)^3},$$

$$\nabla_t^{\delta\xi\delta}(\theta) = \nabla_t^{\delta\delta\xi}(\theta) = -\xi_t(\theta)x_t\frac{\delta_t(\theta)\left(\delta_t(\theta)-x_t\left(2+\xi_t(\theta)\right)\right)}{\left(\delta_t(\theta)+\xi_t(\theta)x_t\right)^3}.$$

Derivatives of the score-driven recursions

In the log-likelihood derivatives above, we also need the derivative of the score-driven recursions. These are given by

$$\frac{\partial f_{t+1}(\theta)}{(\partial\theta)'} = \dot{\Phi}(f_t(\theta);\theta)\frac{\partial f_t(\theta)}{(\partial\theta)'} + U_t(\theta), \qquad U_t(\theta) = \Big(I_2 \quad ; \quad f_t(\theta)' \otimes I_2 \quad ; \quad s_t(\theta)' \otimes I_2\Big), \tag{E.7}$$

and

$$\frac{\partial^2 f_{t+1}(\theta)}{(\partial\theta)(\partial\theta)'} = \dot{\Phi}(f_t(\theta);\theta)\frac{\partial^2 f_t(\theta)}{(\partial\theta)(\partial\theta)'} + \frac{\partial f_t(\theta)}{(\partial\theta)}\frac{\partial\dot{\Phi}(f_t(\theta);\theta)}{(\partial\theta)'} + \frac{\partial U_t(\theta)}{(\partial\theta)'}. \tag{E.8}$$

Alternatively, by using the vectorization operator, we can also write the second derivative recursions as

$$\begin{aligned}\frac{\partial \operatorname{vec}}{(\partial\theta)'}\left(\frac{\partial f_{t+1}(\theta)}{(\partial\theta)'}\right) &= \Big(I_6 \otimes \dot{\Phi}(f_t(\theta);\theta)\Big)\frac{\partial \operatorname{vec}}{(\partial\theta)'}\left(\frac{\partial f_t(\theta)}{(\partial\theta)'}\right) \\ &\quad + \left(\left(\frac{\partial f_t(\theta)}{(\partial\theta)'}\right)' \otimes I_2\right)\frac{\partial \operatorname{vec}}{(\partial\theta)'}\Big(\dot{\Phi}(f_t(\theta);\theta)\Big) + \frac{\partial \operatorname{vec}}{(\partial\theta)'}\left(U_t(\theta)\right).\end{aligned}$$

Finally, we observe that the third-order derivatives of the score-driven process are

$$\frac{\partial \operatorname{vec}}{(\partial\theta)'}\left(\frac{\partial^2 f_{t+1}(\theta)}{(\partial\theta)(\partial\theta)'}\right) := S_t(\theta),$$

where the matrix $S_t(\theta)$ collects the third-order derivatives of the score-driven recursion $f_t(\theta)$. A typical element of this matrix is given by $S_t^{kkk}(\theta)$, for $k \in \{\xi, \delta\}$. Using i, j, l to denote the different elements in the parameter vector θ, it is easy to see that each term is given by $S_t^{kkk}(\theta) = \frac{\partial^3 f_t^k(\theta)}{\partial\theta_i^k\partial\theta_j^k\partial\theta_l^k}$, which takes the form of

$$\begin{aligned}\frac{\partial^3 f_{t+1}^k(\theta)}{\partial\theta_i^k\partial\theta_j^k\partial\theta_l^k} =& \dot{\Phi}(f_t^k(\theta);\theta)\frac{\partial^3 f_t^k(\theta)}{\partial\theta_i^k\partial\theta_j^k\partial\theta_l^k} + \frac{\partial\dot{\Phi}(f_t^k(\theta);\theta)}{\partial\theta_j^k}\frac{\partial^2 f_t^k(\theta)}{\partial\theta_i^k\partial\theta_l^k} + \frac{\partial\dot{\Phi}(f_t^k(\theta);\theta)}{\partial\theta_l^k}\frac{\partial^2 f_t^k(\theta)}{\partial\theta_i\partial\theta_j^k} \\ &+ \frac{\partial^2\dot{\Phi}(f_t^k(\theta);\theta)}{\partial\theta_j^k\partial\theta_l^k}\frac{\partial f_t^k(\theta)}{\partial\theta_i^k} + \frac{\partial^2 U_{i,t}^k(\theta)}{\partial\theta_j^k\partial\theta_l^k}\frac{\partial f_t^k(\theta)}{\partial\theta_i^k}.\end{aligned} \tag{E.9}$$

E.2 Proof of Theorem 3

From the derivative processes up to third-order of the (bivariate) score-driven process $\{f_t(\theta)\}_{t\in\mathbb{Z}}$ given by equations (E.7)–(E.9) above, it is easy to see that each of these follows an SRE of similar form as the one defined in (11) of the main paper, for each $\theta \in \Theta$. Hence, when evaluated at the true parameter vector θ_0, the contraction condition given in (12) of Assumption 2 ensures e.a.s. convergence to a strictly stationary and ergodic solution as a direct consequence of our Theorem 1. Under Assumption 3, we therefore directly obtain the consistency result of the MLE in (i) by a straightforward application of Lemma 1 of Jensen and Rahbek (2004). More formally, the consistency and asymptotic normality in Jensen and Rahbek (2004) follows under the following assumptions:

Assumption 1.

(i) *As* $T^* \to \infty$, $\sqrt{T^*}\frac{\partial \mathcal{L}(\theta_0|\mathcal{F}_T)}{\partial \theta} \xrightarrow{D} N(0, \Omega_S)$, *with* $\Omega_S > 0$ *and* $T^* = \sum_{t=1}^{T} 1\{x_t > 0\}$ *is the number of POT values in the sample.*

(ii) *As* $T^* \to \infty$, $\frac{\partial^2 \mathcal{L}(\theta_0|\mathcal{F}_T)}{\partial \theta \partial \theta'} \xrightarrow{P} \Omega_I$, *with* $\Omega_I > 0$.

(iii) $\max_{i,j,l=1,2,3} \sup_{\theta \in V(\theta_0)} \left| \frac{\partial^3 \mathcal{L}(\theta|\mathcal{F}_T)}{\partial \theta_i \partial \theta_j \partial \theta_l} \right| < c_T$, *where* $V(\theta_0)$ *denotes a neighbourhood of* θ_0 *and* c_T *is some stochastic sequence that satisfies* $0 \leq c_T < c$ *for* $0 < c < \infty$.

In order to prove the asymptotic normality of the MLE in (ii), we first note that by the contraction condition in (14) of Theorem 2, both the score vector in (E.1) and the first derivative process in (E.7) evaluated at θ_0 satisfy the required Lindeberg condition necessary to apply Brown (1971)'s CLT for martingales. Hence we get that

$$\sqrt{T^*}\frac{\partial \mathcal{L}(\theta_0|\mathcal{F}_T)}{\partial \theta} \Rightarrow N(0, \Omega_S), \qquad \Omega_S := \mathbb{E}\left[\frac{\partial \mathcal{L}(\theta_0|\mathcal{F}_T)}{\partial \theta}\frac{\partial \mathcal{L}(\theta_0|\mathcal{F}_T)}{\partial \theta'}\right],$$

as $T^* \to \infty$, such that $(A.1)$ in Lemma 1 of Jensen and Rahbek (2004) is satisfied.

Furthermore, it is also clear that, under the same assumptions, the Hessian matrix in (E.3) and the second derivative process in (E.8) evaluated at θ_0 are strictly stationary and ergodic with the appropriate number of bounded moments, such that a direct application of the ergodic theorem

implies that

$$-\frac{\partial^2 \mathcal{L}(\theta_0|\mathcal{F}_T)}{(\partial\theta)(\partial\theta')} \xrightarrow{P} -\mathbb{E}\left[\frac{\partial^2 \mathcal{L}(\theta_0|\mathcal{F}_T)}{(\partial\theta)(\partial\theta')}\right] := \Omega_I,$$

as $T^* \to \infty$, such that also $(A.2)$ in Lemma 1 of Jensen and Rahbek (2004) is satisfied. Here we remark that the Fisher information matrix equality $\Omega_S = \Omega_I$ follows because, under the maintained assumptions, our model is correctly specified. As a result, the conditional density in (2) evaluated at θ_0 and x_t satisfies

$$p(x_t; \delta_t, \xi_t; \theta_0) = \delta_t^{-1}(\theta_0) \cdot \left(1 + \xi_t(\theta_0)\frac{x_t}{\delta_t(\theta_0)}\right)^{-\xi_t^{-1}(\theta_0)-1},$$

and is the true density. The log-likelihood function $\mathcal{L}(\theta_0|\mathcal{F}_T)$ is twice continuously differentiable and has a bounded moment, which allows us to interchange integration with differentiation.

We are only left with the final condition $(A.3)$ in Lemma 1 of Jensen and Rahbek (2004), which essentially requires the boundedness of third-order derivatives of the log-likelihood function in a small neighbourhood of the true parameter θ_0. However, to check this condition, we note from (E.5) that it suffices to show that

$$\mathbb{E}\left[\sup_{\theta\in V(\theta_0)} \|Q_t(\theta)\|\right] = \mathbb{E}\left[\sup_{\theta\in V(\theta_0)} \left\|\frac{\partial \operatorname{vec} \partial^2 \mathcal{L}(\theta|\mathcal{F}_T)}{(\partial\theta)'(\partial\theta)(\partial\theta)'}\right\|\right] < \infty.$$

Taking into account that the general element of $Q_t(\theta)$ has the form given in (E.6), repeated application of the c_r-inequality yields for $k \in \{\xi, \delta\}$ that

$$\begin{aligned}
\mathbb{E}\left[\sup_{\theta\in V(\theta_0)} \left\|Q_t^{kkk}(\theta)\right\|\right] &\leq c_1 \mathbb{E}\left[\sup_{\theta\in V(\theta_0)} \left|\nabla_t^{kkk}(\theta)\right| \sup_{\theta\in V(\theta_0)} \left\|\frac{\partial f_t^k(\theta)}{(\partial\theta^k)}\right\|^3\right] \\
&+ 3c_2 \mathbb{E}\left[\sup_{\theta\in V(\theta_0)} \left|\nabla_t^{kk}(\theta)\right| \sup_{\theta\in V(\theta_0)} \left\|\frac{\partial^2 f_t^k(\theta)}{(\partial\theta^k)(\partial\theta^k)'}\right\| \sup_{\theta\in V(\theta_0)} \left\|\frac{\partial f_t^k(\theta)}{(\partial\theta^k)'}\right\|\right] \\
&+ c_3 \mathbb{E}\left[\sup_{\theta\in V(\theta_0)} \left|\nabla_t^{k}(\theta)\right| \sup_{\theta\in V(\theta_0)} \left\|\frac{\partial \operatorname{vec}}{(\partial\theta^k)'}\left(\frac{\partial^2 f_t^k(\theta)}{(\partial\theta^k)(\partial\theta^k)'}\right)\right\|\right].
\end{aligned} \tag{E.10}$$

Now, consider the first term in the right-hand side of inequality (E.10), and choose $r > 1$ such that

$3r \leq 3 + 3\gamma$ for some $\gamma > 0$. Then, using Hölder's inequality with $r^{-1} + s^{-1} = 1$, we get

$$\begin{aligned}
&\mathbb{E}\left[\sup_{\theta\in V(\theta_0)}\left|\nabla_t^{kkk}(\theta)\right|\sup_{\theta\in V(\theta_0)}\left\|\frac{\partial f_t^k(\theta)}{(\partial\theta^k)}\right\|^3\right] \\
&\leq \left\{\mathbb{E}\left[\sup_{\theta\in V(\theta_0)}\left|\nabla_t^{kkk}(\theta)\right|^s\right]\right\}^{1/s}\left\{\mathbb{E}\left[\sup_{\theta\in V(\theta_0)}\left\|\frac{\partial f_t^k(\theta)}{(\partial\theta^k)}\right\|^{3r}\right]\right\}^{1/r} \\
&\leq \left\{\mathbb{E}\left[\sup_{\theta\in V(\theta_0)}\left|\nabla_t^{kkk}(\theta)\right|^s\right]\right\}^{1/s}\left\{\mathbb{E}\left[\sup_{\theta\in V(\theta_0)}\left\|\frac{\partial f_t^k(\theta)}{(\partial\theta^k)}\right\|^{3+3\gamma}\right]\right\}^{3/(3+3\gamma)}.
\end{aligned}$$

Similarly, for the second term in the right-hand side of inequality (E.10), we first apply Hölder's inequality, and then the Cauchy-Schwartz inequality, to obtain

$$\begin{aligned}
&\mathbb{E}\left[\sup_{\theta\in V(\theta_0)}\left|\nabla_t^{kk}(\theta)\right|\sup_{\theta\in V(\theta_0)}\left\|\frac{\partial^2 f_t^k(\theta)}{(\partial\theta^k)(\partial\theta^k)'}\right\|\sup_{\theta\in V(\theta_0)}\left\|\frac{\partial f_t^k(\theta)}{(\partial\theta^k)'}\right\|\right] \\
&\leq \left\{\mathbb{E}\left[\sup_{\theta\in V(\theta_0)}\left|\nabla_t^{kk}(\theta)\right|^{1+\gamma}\right]\right\}^{1/(1+\gamma)} \\
&\quad\times\left\{\mathbb{E}\left[\sup_{\theta\in V(\theta_0)}\left\|\frac{\partial^2 f_t^k(\theta)}{(\partial\theta^k)(\partial\theta^k)'}\right\|^{(1+\gamma)/\gamma}\sup_{\theta\in V(\theta_0)}\left\|\frac{\partial f_t^k(\theta)}{(\partial\theta^k)'}\right\|^{(1+\gamma)/\gamma}\right]\right\}^{\gamma/(1+\gamma)} \\
&\leq \left\{\mathbb{E}\left[\sup_{\theta\in V(\theta_0)}\left|\nabla_t^{kk}(\theta)\right|^{1+\gamma}\right]\right\}^{1/(1+\gamma)} \\
&\quad\times\left\{\mathbb{E}\left[\sup_{\theta\in V(\theta_0)}\left\|\frac{\partial^2 f_t^k(\theta)}{(\partial\theta^k)(\partial\theta^k)'}\right\|^{2(1+\gamma)/\gamma}\right]\right\}^{\gamma/(2+2\gamma)}\left\{\mathbb{E}\left[\sup_{\theta\in V(\theta_0)}\left\|\frac{\partial f_t^k(\theta)}{(\partial\theta^k)'}\right\|^{2(1+1\gamma)/\gamma}\right]\right\}^{\gamma/(2+2\gamma)}.
\end{aligned}$$

Finally, the third (and last) term in the right-hand side of inequality (E.10) can be bounded in a similar way, since

$$\begin{aligned}
&\mathbb{E}\left[\sup_{\theta\in V(\theta_0)}\left|\nabla_t^{k}(\theta)\right|\sup_{\theta\in V(\theta_0)}\left\|\frac{\partial\,\mathrm{vec}}{(\partial\theta^k)'}\left(\frac{\partial^2 f_t^k(\theta)}{(\partial\theta^k)(\partial\theta^k)'}\right)\right\|\right] \\
&\leq \left\{\mathbb{E}\left[\sup_{\theta\in V(\theta_0)}\left|\nabla_t^{k}(\theta)\right|^{1+\gamma}\right]\right\}^{1/(1+\gamma)}\left\{\mathbb{E}\left[\sup_{\theta\in V(\theta_0)}\left\|\frac{\partial\,\mathrm{vec}}{(\partial\theta^k)'}\left(\frac{\partial^2 f_t^k(\theta)}{(\partial\theta^k)(\partial\theta^k)'}\right)\right\|^{(1+\gamma)/\gamma}\right]\right\}^{\gamma/(1+\gamma)}.
\end{aligned}$$

To conclude the proof, we note that by using Assumption 4, it follows that there exist a universal constant $0 < c < \infty$ which we can use to upper-bound the inequality in (E.10), and thus verify

that

$$\mathbb{E}\left[\sup_{\theta\in V(\theta_0)}\|Q_t(\theta)\|\right] \leq \mathbb{E}\left[\max_{i,j,k=1,2,3}\sup_{\theta\in V(\theta_0)}\left|\frac{\partial^3\mathcal{L}(\theta|\mathcal{F}_T)}{(\partial\theta_i)'(\partial\theta_j)(\partial\theta_l)'}\right|\right] \leq c < \infty.$$

We thus establish that condition $(A.3)$ required for Lemma 1 of Jensen and Rahbek (2004) also holds true. Together, $(A.1)$ – $(A.3)$ imply the asymptotic normality of the MLE $\hat{\theta}$. ∎

F Confidence bands for tail shape and tail scale

F.1 Simulation-based confidence bands

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

Our confidence bands are based on the variance of $\hat{f}_t$, which we denote by $V_t = \text{Var}(\hat{f}_t)$. They are conditional on the estimated paths of the dynamic thresholds τ_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 F.2 below. In our empirical illustrations below, however, the linear approximations are typically insufficient to capture the uncertainty in the highly non-linear and persistent dynamics of $\hat{f}_t$; compare Figure 1. As a result, delta-method-based bands can become unstable. Therefore, we instead use simulation-based bands as in Blasques et al. (2016, Section 3.3).

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

F.2 Analytic confidence bands

For completeness, 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 2.4 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

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

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 (6).

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 (F.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 $\hat{\theta} - \theta_0 \dot{\sim} \text{N}(0, W)$, where W is the asymptotic covariance matrix associated with

$\hat{\theta}$. A first-order Taylor series expansion around θ_0 yields

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

where we defined

$$\frac{df_{t+1}}{d\theta'} = G_{t+1} = \frac{\partial \varphi_{t+1}}{\partial \theta'} + \frac{\partial \varphi_{t+1}}{\partial f_t'} \cdot \frac{df_t}{d\theta'} + \frac{\partial \varphi_{t+1}}{\partial f_{t-1}'} \cdot \frac{df_{t-1}}{d\theta'}. \tag{F.3}$$

Interestingly, (F.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, (F.3) can be written as

$$G_{t+1} = \frac{\partial \varphi_{t+1}}{\partial \theta'} + \frac{\partial \varphi_{t+1}}{\partial f_t'} \cdot G_t + \frac{\partial \varphi_{t+1}}{\partial f_{t-1}'} \cdot G_{t-1}, \tag{F.4}$$

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 (F.4) can be derived as

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

$$\frac{\partial \varphi_{t+1}}{\partial f_t'} = \lambda I_2 + B + (1-\lambda) A \frac{\partial s_t}{\partial f_t'}, \tag{F.6}$$

$$\frac{\partial \varphi_{t+1}}{\partial f_{t-1}'} = -\lambda B,$$

where $\frac{\partial s_t}{\partial \theta'} = 0$ (see (6)).

The derivative terms needed in (F.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 (F.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.

G Derivation of EVT-based market risk measures

This section derives the conditional market risk measures in Section 2.5.

To derive the one-step-ahead VaR, we note that

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

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

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

where t^*/t serves as an estimator of $\overline{G}(\tau_t)$. This expression coincides with the expression given in the main text.

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

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

which 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 $\xi_t < 1$.

H Simulation results

H.1 Additional figures for the first set of DGPs

This section presents two additional figures associated with our first simulation study in Section 3 (DGP1).

Figures H.1 and H.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 (pseudo-)true values. Figure H.1 refers to simulations from a GPD conditional density (Paths 1 – 4), for which the GPD conditional density is exact. Figure H.2 refers to simulations from a Student's t conditional density (Paths 1 – 4), for which the GPD conditional density is only approximate for any finite value of $\tau_t < \infty$. In the presence of misspecification, score updates continue to minimize the local Kullback-Leibler divergence between the true conditional density and the model-implied conditional density, and remain optimal in this sense; see Blasques et al. (2015). The time-varying thresholds τ_t evolve according to (9) at a $1 - \kappa = 5\%$ tail probability.

Figure H.1: Simulation results for GPD data

Simulation results for $y_t \sim \text{GPD}(\alpha_t^{-1}, \sigma_t)$ with time-varying tail shape α_t^{-1} and scale σ_t. Rows refer to different parameter paths (1) – (4); see Section 3.2. 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. The time-varying threshold $\hat{\tau}_t$ is estimated based on the recursive specification (9) in conjunction with the objective function (17).

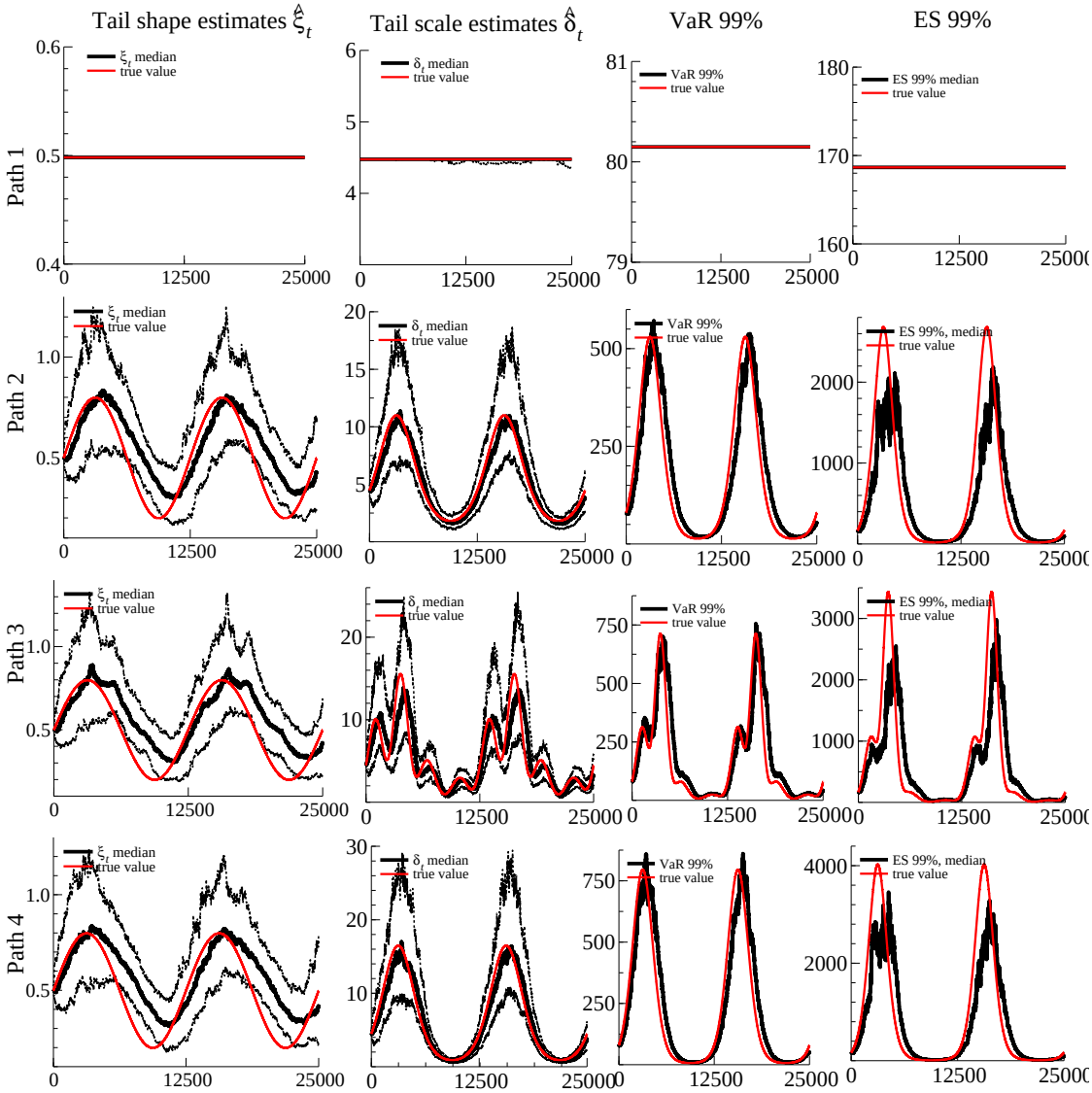
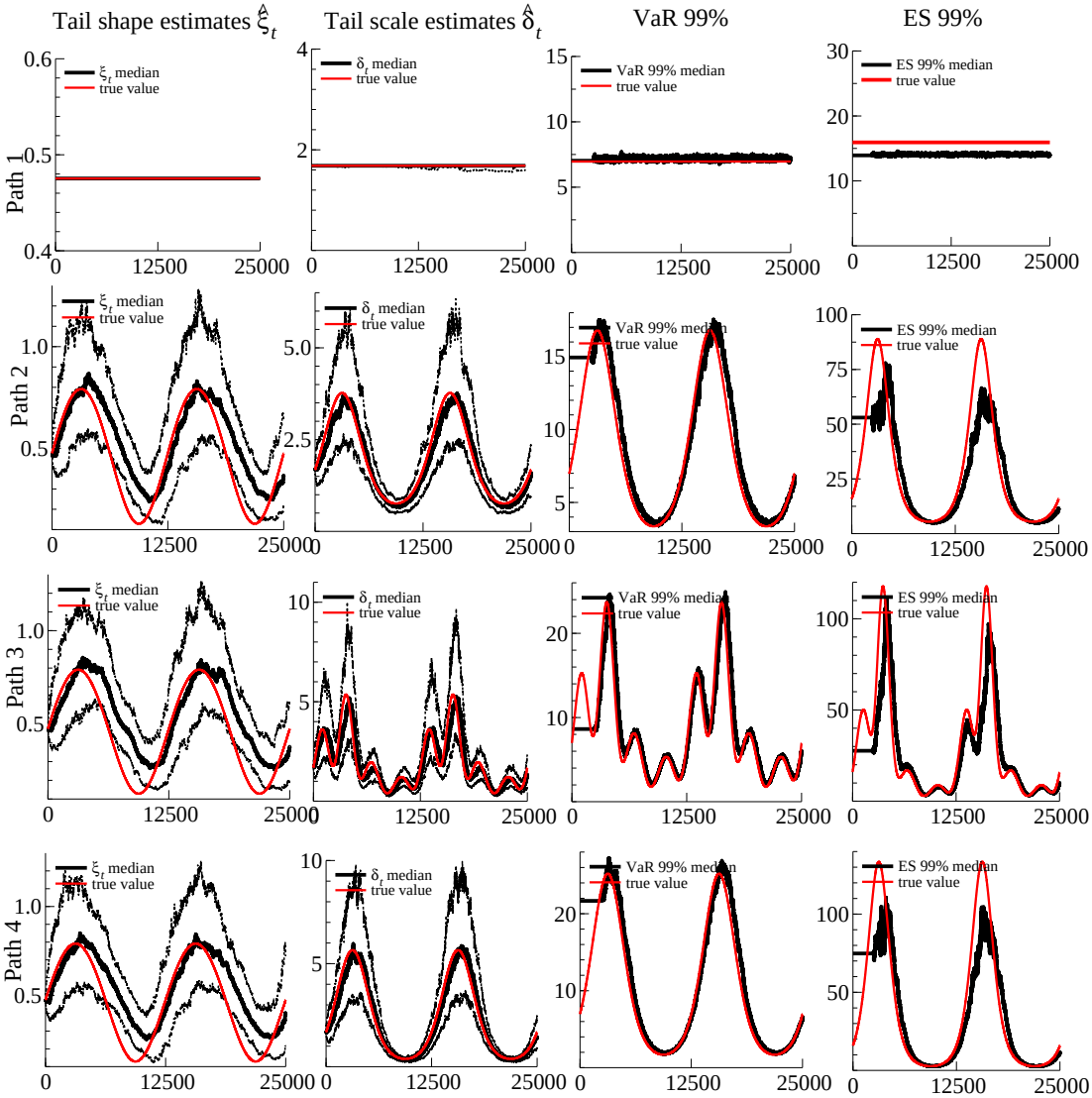


Figure H.2: Simulation results for Student's t data

Simulation results for $y_t \sim \mathrm{t}(0, , \sigma_t, \alpha_t^{-1})$ with time-varying scale σ_t and tail shape α_t^{-1}. Rows refer to different parameter paths (1) – (4); see Section 3.2. 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. The time-varying threshold $\hat{\tau}_t$ is estimated based on the recursive specification (9) in conjunction with the objective function (17).



H.2 The second set of DGPs

Simulation setup

Empirical estimates of the autoregressive parameters b^{ξ} and b^{δ} can be close to one; see Section 4. This section therefore investigates the effect of covariates and (near-)unit root type dynamics on the time-varying parameter paths and the deterministic parameter estimates and their standard errors in a second simulation design. Specifically, we simulate 100 samples from a GPD$(x_t; \xi_t, \delta_t)$ density, with T=25,000 observations each, thus abstracting from any misspecification effects. The factor $f_t = (\ln \xi_t, \ln \delta_t)'$ follows the transition equation

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

where matrices ω, A, B, and C take four different sets of values. As a first case, we consider a slowly mean-reverting factor process with $\omega = (0.50, 1.00)'$, $A = \text{diag}(0.03, 0.07)$, $B = \text{diag}(0.98, 0.98)$, and $C = 0$. The second case considers an integrated factor process: $B = \text{I}_2$, while ω, A, and C remain unchanged. A third and fourth case are identical to the first and second case, except that now $C = (-3, -1.5)'$ in (8) for an observed variable z_t. As our z_t we use the central bank purchases of Italian sovereign bonds as considered in our second application in Section 4.2.

Simulation results

We now turn to the simulation results for DGP2. Table H.1 presents RMSEs associated with the time-varying parameters ξ_t and δ_t and the deterministic parameters a^{ξ} and a^{δ}. We consider two settings: with (bottom panel) and without (top panel) a covariate. Within each of these, we consider a stationary (GAS) and unit root (iGAS) DGP (in columns), as well as the corresponding model specifications (in rows). Figure H.3 provides more results in the form of representative draws of $\hat{\xi}_t$ and $\hat{\delta}_t$ for each of the four cases, and results on standard error estimates for a^{ξ} and a^{δ}.

Table H.1 suggests that both the GAS and the iGAS models work well if they are correctly specified (iGAS row and iGAS column, or GAS row and GAS column, etc.). In particular, the estimated $\hat{\xi}_t$ and $\hat{\delta}_t$ tend to be closely aligned to their true values. Also the (slightly) misspecified cases of a GAS model for an iGAS DGP and vice versa continue to work reasonably well: ξ_t and δ_t remain close to their true paths.

Table H.1: RMSE outcomes for DGP2

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

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

When investigating the standard errors of the deterministic parameter estimates, Table H.2 suggests that, while parameter point estimates are close to their true values, the usual asymptotic standard error estimates based on the inverse Hessian or the sandwich estimates are not necessarily reliable in the two iGAS cases. In our set-up, these common estimates of the standard errors are typically too large, providing too conservative inference. A bootstrap procedure tailored to integrated processes could then be used to avoid this issue; see for instance Boswijk et al. (2021).

Figure H.3 plots example estimates of $\hat{\xi}_t$ and $\hat{\delta}_t$ that are associated with one particular draw for each of the four cases. Table H.2 reports the standard error estimates associated with the deterministic parameters.

Figure H.3: Tail shape and scale estimates from simulated data: one draw

Estimated tail shape $\hat{\xi}_t$ and tail shape $\hat{\delta}_t$ parameters for DGP2. First row: iGAS DGP and iGAS estimates (correctly specified). Second row: iGAS DGP and GAS estimates (misspecified). Third row: GAS DGP and iGAS estimates (misspecified). Fourth row: GAS DGP and GAS estimates (correctly specified). Factors of ξ_t and δ_t are initialized with a static GPD model using the first 250 observations.

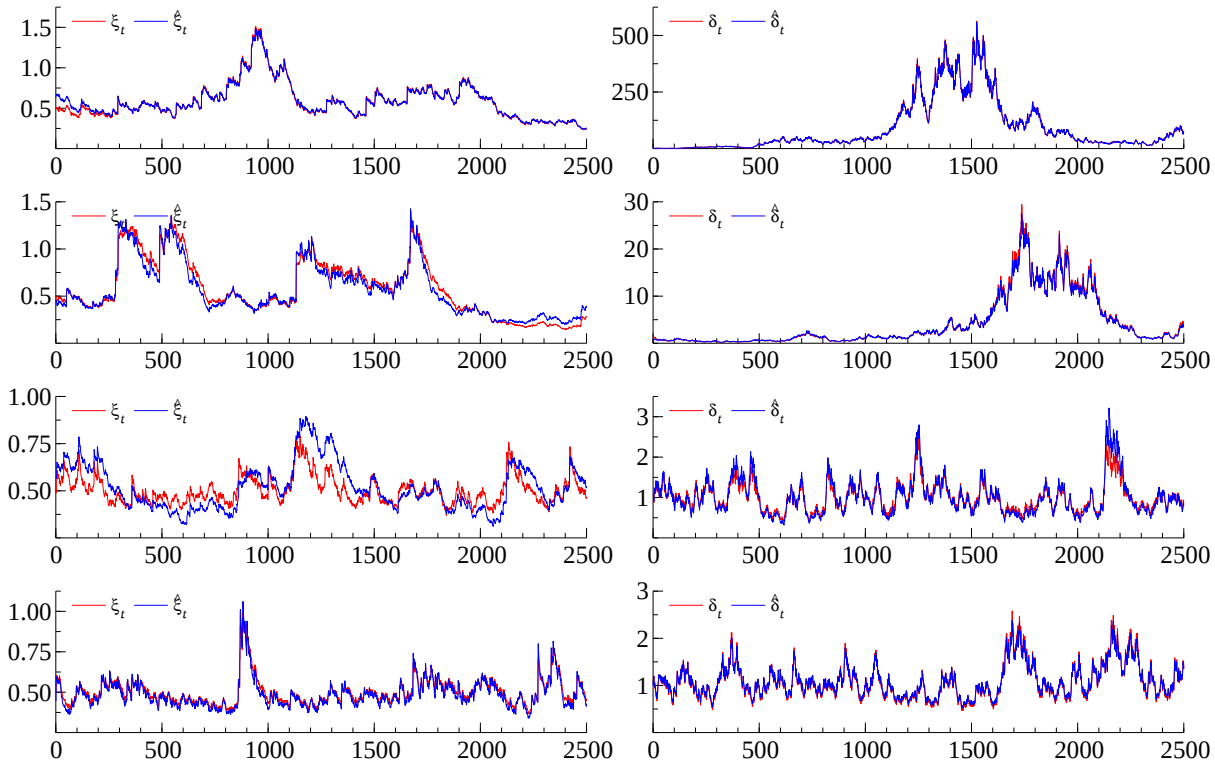


Table H.2: Simulation results: standard errors using different methods

Standard error estimates for the dynamic tail shape model's deterministic parameters. We first simulate from iGAS and GAS models, and then estimate back all parameters based on both iGAS and GAS specifications (top half). We then include an exogenous factor z_t, denoted iGAS-X or GAS-X in both the DGP and the statistical model (bottom half). Columns ParValue and EstValue report average (across simulations) constrained and unconstrained parameter estimates. StdErr^E denotes the standard deviation of the parameter estimates. Standard error estimates refer to the Empirical Hessian estimator (StdErr^H), the Outer Product of the Gradient estimator (StdErr^O), and a Sandwich covariance matrix estimator (StdErr^S).

	ParValue	EstValue	StdErr^E	StdErr^H	StdErr^O	StdErr^S
	DGP: iGAS, Estimate: iGAS					
α^ξ	0.030	-3.534	0.153	0.265	0.267	0.270
α^δ	0.069	-2.682	0.112	0.140	0.136	0.145
	DGP: iGAS, Estimate: GAS					
α^ξ	0.028	-3.606	0.309	0.353	0.369	0.376
α^δ	0.069	-2.690	0.167	0.149	0.150	0.150
β^ξ	0.996	5.986	0.913	0.907	1.028	0.875
β^δ	0.997	5.916	0.675	0.688	0.691	0.691
ω^ξ	1.468	1.468	1.667	0.955	1.205	0.833
ω^δ	2.054	2.054	3.641	1.280	1.238	1.334
	DGP: GAS, Estimate: iGAS					
α^ξ	0.015	-4.465	0.885	0.749	0.627	1.026
α^δ	0.064	-2.761	0.200	0.163	0.132	0.205
	DGP: GAS, Estimate: GAS					
α^ξ	0.033	-3.481	0.393	0.544	0.613	0.524
α^δ	0.068	-2.693	0.148	0.172	0.175	0.172
β^ξ	0.968	3.702	0.754	0.819	1.031	0.717
β^δ	0.978	3.839	0.313	0.370	0.375	0.369
ω^ξ	0.499	0.499	0.027	0.046	0.046	0.046
ω^δ	1.003	1.003	0.079	0.098	0.098	0.098
	DGP: iGAS-X, Estimate: iGAS-X					
α^ξ	0.029	-3.588	0.306	0.348	0.341	0.459
α^δ	0.068	-2.695	0.138	0.132	0.125	0.171
c^ξ	-2.562	-2.562	0.996	1.298	5.666	1.380
c^δ	-1.568	-1.568	0.242	0.257	0.257	0.313
	DGP: GAS-X, Estimate: iGAS-X					
α^ξ	0.029	-3.693	0.949	0.602	0.911	0.622
α^δ	0.071	-2.650	0.153	0.142	0.120	0.172
c^ξ	-0.221	-0.221	0.135	0.195	0.160	0.271
c^δ	-0.663	-0.663	0.236	0.282	0.298	0.276
	DGP: iGAS-X, Estimate: GAS-X					
α^ξ	0.023	-4.913	2.959	6.308	10.322	1.708
α^δ	0.067	-2.709	0.155	0.132	0.129	0.145
β^ξ	0.999	8.404	2.969	15.301	18.600	8.561
β^δ	1.000	10.411	2.675	20.271	33.219	15.809
ω^ξ	2.184	2.184	3.140	15.338	19.571	18.477
ω^δ	3.034	3.034	2.809	35.958	40.176	23.683
c^ξ	-2.394	-2.394	1.287	2.244	4.227	2.827
c^δ	-1.601	-1.601	0.274	0.271	0.274	0.306
	DGP: GAS-X, Estimate: GAS-X					
α^ξ	0.032	-3.613	0.903	1.032	1.226	0.842
α^δ	0.069	-2.685	0.160	0.163	0.164	0.167
β^ξ	0.977	3.802	0.372	0.556	0.642	0.625
β^δ	0.977	3.802	0.268	0.280	0.284	0.285
ω^ξ	0.505	0.505	0.033	0.051	0.051	0.053
ω^δ	0.990	0.990	0.091	0.101	0.102	0.102
c^ξ	-4.236	-4.236	2.579	3.137	3.794	3.180
c^δ	-1.585	-1.585	0.391	0.403	0.422	0.407

I Two additional empirical illustrations

This appendix reports empirical results for two additional asset classes: exchange rates and commodities. Specifically, we study daily GBP/USD log-returns, and daily Brent crude oil log-returns. The exchange rate sample ranges from 5 January 1971 to 30 December 2022. The Brent oil sample ranges from 20 May 1987 to 30 December 2022. We focus on the extreme left tail of each series.

Table I.1 presents the model's deterministic parameter estimates. A numerical check reveals that the deterministic parameters lie within the SE region implied by the sufficient conditions of Theorems 1 and 2. The table does not include parameters c^{τ}, c^{ξ}, and c^{δ} since the model does not include exogenous variables. Figure I.1 plots time-varying parameters, along with each series' VaR and ES over time. For both log-returns, we observe pronounced time variation in ξ_t and δ_t.

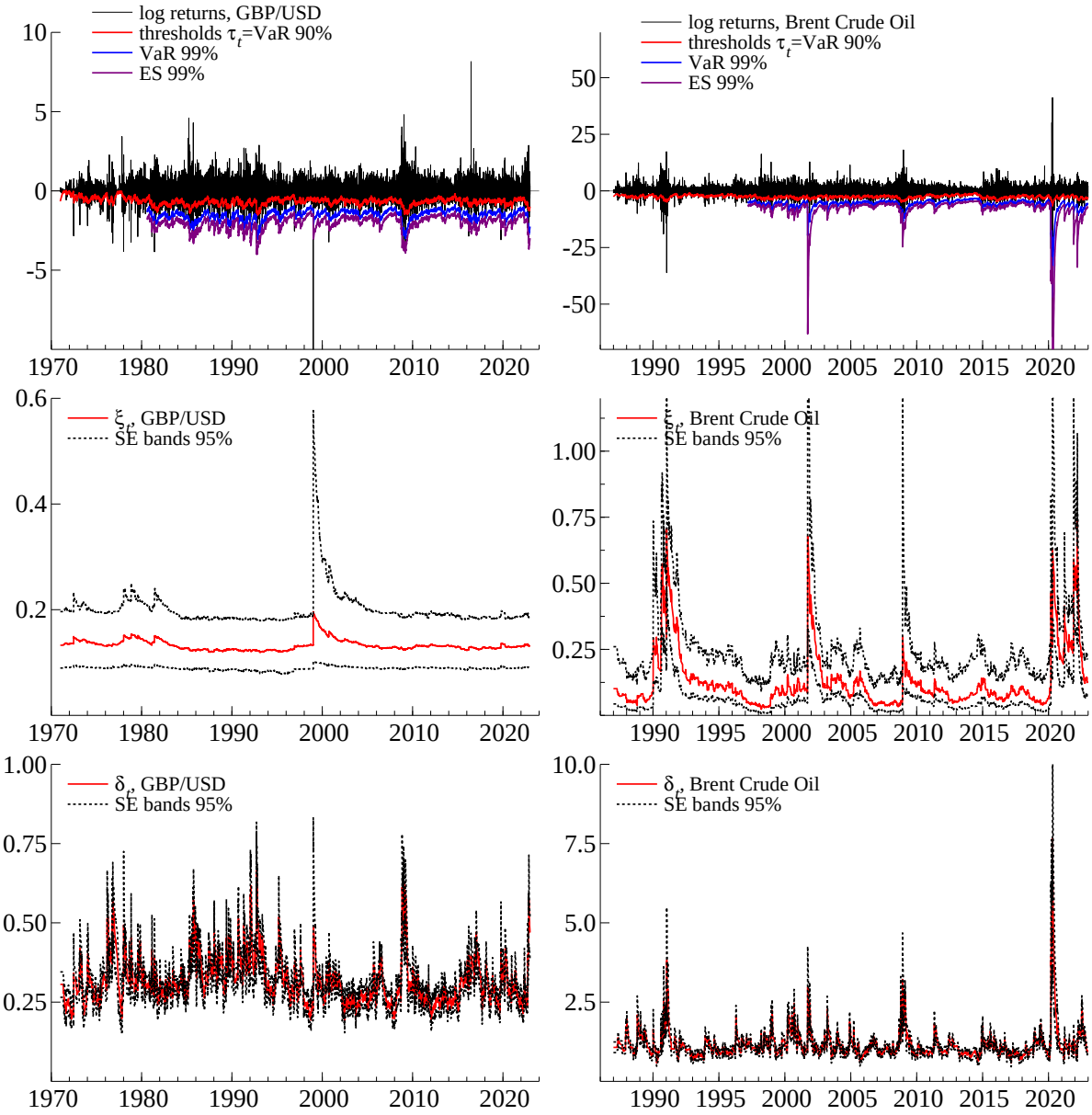
Table I.1: Parameter estimates

Parameter estimates for the dynamic tail shape model. The second and third columns refer to two additional illustrations: GBP/USD exchange rate log-returns, and Brent crude oil log-returns. The estimation samples ranges from 5 January 1971 to 30 December 2022, and from 20 May 1987 to 30 December 2022, respectively. Standard error estimates are in round brackets and are based on a sandwich covariance matrix estimator. p-values are in square brackets.

	Two additional illustrations	
	GBP/USD	Brent oil
ω^{ξ}	-2.019	-2.292
	(0.10)	(0.22)
	[0.00]	[0.00]
ω^{δ}	-1.184	0.070
	(0.03)	(0.03)
	[0.00]	[0.04]
a^{ξ}	0.008	0.148
	(0.01)	(0.03)
	[0.37]	[0.00]
a^{δ}	0.086	0.148
	(0.01)	(0.01)
	[0.00]	[0.00]
b^{ξ}	0.998	0.997
	(0.00)	(0.00)
	[0.00]	[0.00]
b^{δ}	0.992	0.985
	(0.00)	(0.00)
	[0.00]	[0.00]
λ	0	0
a^{τ}	0.096	0.420
b^{τ}	0.993	0.982
T	13,450	9,040
T^{*}	1,339	917
loglik	-9,011.8	-36,881.3
AIC	18,035.7	73,774.6
BIC	18,080.7	73,817.3

Figure I.1: Filtered tail parameters for GBP/USD and crude oil log-returns

Top panels: daily log-returns for the GBP/USD exchange rate (left) and Brent crude oil (right). Middle and bottom panels: filtered tail shape (ξ_t, middle) and tail scale (δ_t, bottom) parameters. The thresholds τ_t are reported at a 90% confidence level. Value-at-Risk (VaR) and Expected Shortfall (ES) are plotted at an extreme 99% confidence level (top panels). The estimation samples range from 5 January 1971 to 30 December 2022 for the exchange rate, and from 20 May 1987 to 30 December 2022 for Brent crude oil.

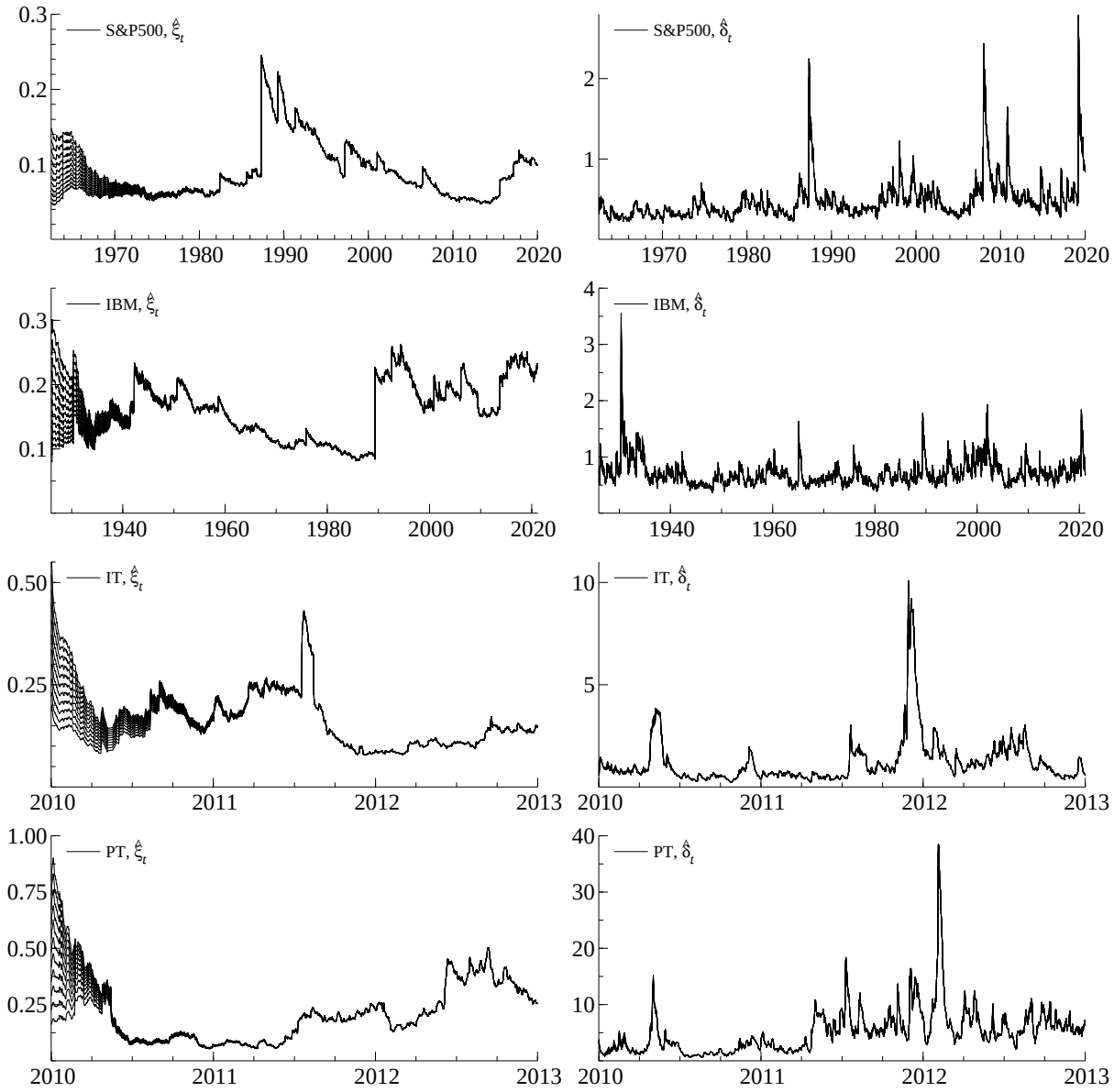


J Diagnostic checks for filter invertibility

Figures 4 and 5 in Section 4 plot filtered estimates of ξ_t and δ_t implied by maximum-likelihood estimates $\hat{\theta}$ of the deterministic parameters reported in Table 2. Figure J.1 plots the time-varying parameters for different initial values of f_1. Both ξ_t and δ_t all converge to the same path, suggesting that the bivariate filter is invertible at the empirical estimates $\hat{\theta}$.

Figure J.1: Feasible invertibility conditions for the model

The plots show the filtered paths of tail index $\hat{\xi}_t$ and tail shape $\hat{\delta}_t$ parameters when the factors are initialized with different starting values. The panels on the left are results for ξ_t while the right side panels are results for δ_t. The lines in the plots correspond to initiate the factors at $\hat{f}_0 + c \cdot \sigma(\hat{f})$, with $c = (-0.5, -0.4, \cdots, 0, \cdots, +0.4, +0.5)$ and $\sigma(\hat{f})$ is the standard deviations of the filtered factors.



K Bootstrapping standard errors of deterministic parameters

Section 4.2 reports bootstrapped standard errors; compare also Boswijk et al. (2021). This section explains how such standard error estimates can be obtained.

The bootstrap proceeds along the following steps. For completeness, the null hypothesis (H_0) states that observed covariate z_t has no impact on the tail parameters ξ_t and δ_t. The alternative hypothesis (H_1) states that the covariate's impact on the tail parameters is different from zero.

1. estimate the model for the dynamic thresholds τ_t, and save the "hit times" and the values of x_t.
2. estimate model under H_1.
3. compute $\hat{\xi}_t$ and $\hat{\delta}_t$ for all t.
4. compute PITs, $u_t = 1(1 + \xi_t x_t/\delta_t)^{-1/\xi_t}$, using the GPD cdf (2). Use the x_t from step 0 for this.
5. estimate model parameters under H_0 to compute $\hat{\xi}_t^{\,0}$ and $\hat{\delta}_t^{\,0}$.
6. sample with replacement T values from u_t to obtain u_t^* for $t = 1 \cdots T$.
7. compute $x_t^* = invCDF^{GPD}(u_t^*) = \hat{\delta}_t^{\,0}/\hat{\xi}_t^{\,0} \cdot [(1 - u_t^*)^{-\hat{\xi}_t^{\,0}} - 1]$.
8. with x_t^*, $t = 1, \ldots, T$, estimate model under H_1.
9. repeat steps 6–8 many times, storing all estimates of $a^{(.)}$ and $c^{(.)}$. Then compute the standard deviation of those estimates, and/or use them to compute t- and p-values directly.

L VaR impact estimates for changes in bond yields

The deterministic parameter estimates presented in the final two columns of Table 2 are difficult to interpret in economic (or probabilistic) terms when considered in isolation. This section addresses the economic question how market risk measures, such as VaR, responded on average to a €1 bn bond purchase intervention. To this end, we note that the sensitivity of $\text{VaR}^\gamma(y_t)$ to bond purchases z_{t-1} is given by

$$\begin{aligned}\frac{\mathrm{dVaR}^\gamma(y_t)}{\mathrm{d}z_{t-1}} &= \frac{\partial \text{VaR}}{\partial \tau_t}\frac{\mathrm{d}\tau_t}{\mathrm{d}z_{t-1}} + \frac{\partial \text{VaR}}{\partial \delta_t}\frac{\mathrm{d}\delta_t}{\mathrm{d}f_t^\delta}\frac{\mathrm{d}f_t^\delta}{\mathrm{d}z_{t-1}} + \frac{\partial \text{VaR}}{\partial \xi_t}\frac{\mathrm{d}\xi_t}{\mathrm{d}f_t^\xi}\frac{\mathrm{d}f_t^\xi}{\mathrm{d}z_{t-1}} \\ &= c^\tau + \frac{\text{VaR}^\gamma(y_t) - \tau_t}{\delta_t}\ \delta_t\ c^\delta - \frac{\text{VaR}^\gamma(y_t) - \tau_t + \delta_t\left(\frac{1-\gamma}{t^*/t}\right)^{-\xi_t}\ln\left(\frac{1-\gamma}{t^*/t}\right)}{\xi_t}\ \xi_t\ c^\xi \\ &= c^\tau + (\text{VaR}^\gamma(y_t) - \tau_t)\ c^\delta - \left(\text{VaR}^\gamma(y_t) - \tau_t + \delta_t\left(\frac{1-\gamma}{t^*/t}\right)^{-\xi_t}\ln\left(\frac{1-\gamma}{t^*/t}\right)\right)\ c^\xi,\end{aligned} \tag{L.1}$$

where c^τ is defined in (10) and c^δ, c^ξ are given in (8) with $C = (c^\delta, c^\xi)'$. The expression is intuitive: upper tail quantiles can change if bond purchases z_{t-1} affect the conditional quantile τ_t, the conditional tail scale δ_t, or the conditional tail shape ξ_t. The total impact is obtained as the weighted average

$$\text{VaR impact}^\gamma = (1/\sum_t^T z_t)\sum_t^T (\mathrm{dVaR}^\gamma(y_t)/\mathrm{d}z_{t-1})\, z_{t-1}. \tag{L.2}$$

References

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

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

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

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

Brown, B. M. (1971). Martingale central limit theorems. *The Annals of Mathematical Statistics 42*(1), 59–66.

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

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

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

Jensen, S. T. and A. Rahbek (2004). Asymptotic Inference for nonstationary GARCH. *Econometric Theory 20*(6), 1203–1226.

Krengel, U. (2011). *Ergodic theorems*, Volume 6. Walter de Gruyter.

Straumann, D. and T. Mikosch (2006). Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: A stochastic recurrence equations approach. *The Annals of Statistics 34*(5), 2449–2495.

Earlier Working Papers:

For a complete list of Working Papers published by Sveriges Riksbank, see www.riksbank.se

Estimation of an Adaptive Stock Market Model with Heterogeneous Agents *by Henrik Amilon*	2005:177
Some Further Evidence on Interest-Rate Smoothing: The Role of Measurement Errors in the Output Gap *by Mikael Apel and Per Jansson*	2005:178
Bayesian Estimation of an Open Economy DSGE Model with Incomplete Pass-Through by Malin Adolfson, Stefan Laséen, Jesper Lindé and Mattias Villani	2005:179
Are Constant Interest Rate Forecasts Modest Interventions? Evidence from an Estimated Open Economy DSGE Model of the Euro Area *by Malin Adolfson, Stefan Laséen, Jesper Lindé and Mattias Villani*	2005:180
Inference in Vector Autoregressive Models with an Informative Prior on the Steady State *by Mattias Villani*	2005:181
Bank Mergers, Competition and Liquidity *by Elena Carletti, Philipp Hartmann and Giancarlo Spagnolo*	2005:182
Testing Near-Rationality using Detailed Survey Data *by Michael F. Bryan and Stefan Palmqvist*	2005:183
Exploring Interactions between Real Activity and the Financial Stance *by Tor Jacobson, Jesper Lindé and Kasper Roszbach*	2005:184
Two-Sided Network Effects, Bank Interchange Fees, and the Allocation of Fixed Costs *by Mats A. Bergman*	2005:185
Trade Deficits in the Baltic States: How Long Will the Party Last? *by Rudolfs Bems and Kristian Jönsson*	2005:186
Real Exchange Rate and Consumption Fluctuations follwing Trade Liberalization *by Kristian Jönsson*	2005:187
Modern Forecasting Models in Action: Improving Macroeconomic Analyses at Central Banks *by Malin Adolfson, Michael K. Andersson, Jesper Lindé, Mattias Villani and Anders Vredin*	2005:188
Bayesian Inference of General Linear Restrictions on the Cointegration Space *by Mattias Villani*	2005:189
Forecasting Performance of an Open Economy Dynamic Stochastic General Equilibrium Model *by Malin Adolfson, Stefan Laséen, Jesper Lindé and Mattias Villani*	2005:190
Forecast Combination and Model Averaging using Predictive Measures *by Jana Eklund and Sune Karlsson*	2005:191
Swedish Intervention and the Krona Float, 1993-2002 *by Owen F. Humpage and Javiera Ragnartz*	2006:192
A Simultaneous Model of the Swedish Krona, the US Dollar and the Euro *by Hans Lindblad and Peter Sellin*	2006:193
Testing Theories of Job Creation: Does Supply Create Its Own Demand? *by Mikael Carlsson, Stefan Eriksson and Nils Gottfries*	2006:194
Down or Out: Assessing The Welfare Costs of Household Investment Mistakes *by Laurent E. Calvet, John Y. Campbell and Paolo Sodini*	2006:195
Efficient Bayesian Inference for Multiple Change-Point and Mixture Innovation Models *by Paolo Giordani and Robert Kohn*	2006:196
Derivation and Estimation of a New Keynesian Phillips Curve in a Small Open Economy *by Karolina Holmberg*	2006:197
Technology Shocks and the Labour-Input Response: Evidence from Firm-Level Data *by Mikael Carlsson and Jon Smedsaas*	2006:198
Monetary Policy and Staggered Wage Bargaining when Prices are Sticky *by Mikael Carlsson and Andreas Westermark*	2006:199
The Swedish External Position and the Krona *by Philip R. Lane*	2006:200

Price Setting Transactions and the Role of Denominating Currency in FX Markets *by Richard Friberg and Fredrik Wilander*	2007:201
The geography of asset holdings: Evidence from Sweden *by Nicolas Coeurdacier and Philippe Martin*	2007:202
Evaluating An Estimated New Keynesian Small Open Economy Model *by Malin Adolfson, Stefan Laséen, Jesper Lindé and Mattias Villani*	2007:203
The Use of Cash and the Size of the Shadow Economy in Sweden *by Gabriela Guibourg and Björn Segendorf*	2007:204
Bank supervision Russian style: Evidence of conflicts between micro- and macro-prudential concerns *by Sophie Claeys and Koen Schoors*	2007:205
Optimal Monetary Policy under Downward Nominal Wage Rigidity *by Mikael Carlsson and Andreas Westermark*	2007:206
Financial Structure, Managerial Compensation and Monitoring *by Vittoria Cerasi and Sonja Daltung*	2007:207
Financial Frictions, Investment and Tobin's q *by Guido Lorenzoni and Karl Walentin*	2007:208
Sticky Information vs Sticky Prices: A Horse Race in a DSGE Framework *by Mathias Trabandt*	2007:209
Acquisition versus greenfield: The impact of the mode of foreign bank entry on information and bank lending rates *by Sophie Claeys and Christa Hainz*	2007:210
Nonparametric Regression Density Estimation Using Smoothly Varying Normal Mixtures *by Mattias Villani, Robert Kohn and Paolo Giordani*	2007:211
The Costs of Paying – Private and Social Costs of Cash and Card *by Mats Bergman, Gabriella Guibourg and Björn Segendorf*	2007:212
Using a New Open Economy Macroeconomics model to make real nominal exchange rate forecasts *by Peter Sellin*	2007:213
Introducing Financial Frictions and Unemployment into a Small Open Economy Model *by Lawrence J. Christiano, Mathias Trabandt and Karl Walentin*	2007:214
Earnings Inequality and the Equity Premium *by Karl Walentin*	2007:215
Bayesian forecast combination for VAR models *by Michael K. Andersson and Sune Karlsson*	2007:216
Do Central Banks React to House Prices? *by Daria Finocchiaro and Virginia Queijo von Heideken*	2007:217
The Riksbank's Forecasting Performance *by Michael K. Andersson, Gustav Karlsson and Josef Svensson*	2007:218
Macroeconomic Impact on Expected Default Freqency *by Per Åsberg and Hovick Shahnazarian*	2008:219
Monetary Policy Regimes and the Volatility of Long-Term Interest Rates *by Virginia Queijo von Heideken*	2008:220
Governing the Governors: A Clinical Study of Central Banks *by Lars Frisell, Kasper Roszbach and Giancarlo Spagnolo*	2008:221
The Monetary Policy Decision-Making Process and the Term Structure of Interest Rates *by Hans Dillén*	2008:222
How Important are Financial Frictions in the U S and the Euro Area *by Virginia Queijo von Heideken*	2008:223
Block Kalman filtering for large-scale DSGE models *by Ingvar Strid and Karl Walentin*	2008:224
Optimal Monetary Policy in an Operational Medium-Sized DSGE Model *by Malin Adolfson, Stefan Laséen, Jesper Lindé and Lars E. O. Svensson*	2008:225
Firm Default and Aggregate Fluctuations *by Tor Jacobson, Rikard Kindell, Jesper Lindé and Kasper Roszbach*	2008:226
Re-Evaluating Swedish Membership in EMU: Evidence from an Estimated Model *by Ulf Söderström*	2008:227

The Effect of Cash Flow on Investment: An Empirical Test of the Balance Sheet Channel *by Ola Melander*	2009:228
Expectation Driven Business Cycles with Limited Enforcement *by Karl Walentin*	2009:229
Effects of Organizational Change on Firm Productivity *by Christina Håkanson*	2009:230
Evaluating Microfoundations for Aggregate Price Rigidities: Evidence from Matched Firm-Level Data on Product Prices and Unit Labor Cost *by Mikael Carlsson and Oskar Nordström Skans*	2009:231
Monetary Policy Trade-Offs in an Estimated Open-Economy DSGE Model *by Malin Adolfson, Stefan Laséen, Jesper Lindé and Lars E. O. Svensson*	2009:232
Flexible Modeling of Conditional Distributions Using Smooth Mixtures of Asymmetric Student T Densities *by Feng Li, Mattias Villani and Robert Kohn*	2009:233
Forecasting Macroeconomic Time Series with Locally Adaptive Signal Extraction *by Paolo Giordani and Mattias Villani*	2009:234
Evaluating Monetary Policy *by Lars E. O. Svensson*	2009:235
Risk Premiums and Macroeconomic Dynamics in a Heterogeneous Agent Model *by Ferre De Graeve, Maarten Dossche, Marina Emiris, Henri Sneessens and Raf Wouters*	2010:236
Picking the Brains of MPC Members *by Mikael Apel, Carl Andreas Claussen and Petra Lennartsdotter*	2010:237
Involuntary Unemployment and the Business Cycle *by Lawrence J. Christiano, Mathias Trabandt and Karl Walentin*	2010:238
Housing collateral and the monetary transmission mechanism *by Karl Walentin and Peter Sellin*	2010:239
The Discursive Dilemma in Monetary Policy *by Carl Andreas Claussen and Øistein Røisland*	2010:240
Monetary Regime Change and Business Cycles *by Vasco Cúrdia and Daria Finocchiaro*	2010:241
Bayesian Inference in Structural Second-Price common Value Auctions *by Bertil Wegmann and Mattias Villani*	2010:242
Equilibrium asset prices and the wealth distribution with inattentive consumers *by Daria Finocchiaro*	2010:243
Identifying VARs through Heterogeneity: An Application to Bank Runs *by Ferre De Graeve and Alexei Karas*	2010:244
Modeling Conditional Densities Using Finite Smooth Mixtures *by Feng Li, Mattias Villani and Robert Kohn*	2010:245
The Output Gap, the Labor Wedge, and the Dynamic Behavior of Hours *by Luca Sala, Ulf Söderström and Antonella Trigari*	2010:246
Density-Conditional Forecasts in Dynamic Multivariate Models *by Michael K. Andersson, Stefan Palmqvist and Daniel F. Waggoner*	2010:247
Anticipated Alternative Policy-Rate Paths in Policy Simulations *by Stefan Laséen and Lars E. O. Svensson*	2010:248
MOSES: Model of Swedish Economic Studies *by Gunnar Bårdsen, Ard den Reijer, Patrik Jonasson and Ragnar Nymoen*	2011:249
The Effects of Endogenuos Firm Exit on Business Cycle Dynamics and Optimal Fiscal Policy *by Lauri Vilmi*	2011:250
Parameter Identification in a Estimated New Keynesian Open Economy Model *by Malin Adolfson and Jesper Lindé*	2011:251
Up for count? Central bank words and financial stress *by Marianna Blix Grimaldi*	2011:252
Wage Adjustment and Productivity Shocks *by Mikael Carlsson, Julián Messina and Oskar Nordström Skans*	2011:253

The timing of uncertainty shocks in a small open economy *by Hanna Armelius, Isaiah Hull and Hanna Stenbacka Köhler*	2016:334
Quantitative easing and the price-liquidity trade-off *by Marien Ferdinandusse, Maximilian Freier and Annukka Ristiniemi*	2017:335
What Broker Charges Reveal about Mortgage Credit Risk *by Antje Berndt, Burton Hollifield and Patrik Sandåsi*	2017:336
Asymmetric Macro-Financial Spillovers *by Kristina Bluwstein*	2017:337
Latency Arbitrage When Markets Become Faster *by Burton Hollifield, Patrik Sandås and Andrew Todd*	2017:338
How big is the toolbox of a central banker? Managing expectations with policy-rate forecasts: Evidence from Sweden *by Magnus Åhl*	2017:339
International business cycles: quantifying the effects of a world market for oil *by Johan Gars and Conny Olovsson l*	2017:340
Systemic Risk: A New Trade-Off for Monetary Policy? *by Stefan Laséen, Andrea Pescatori and Jarkko Turunen*	2017:341
Household Debt and Monetary Policy: Revealing the Cash-Flow Channel *by Martin Flodén, Matilda Kilström, Jósef Sigurdsson and Roine Vestman*	2017:342
House Prices, Home Equity, and Personal Debt Composition *by Jieying Li and Xin Zhang*	2017:343
Identification and Estimation issues in Exponential Smooth Transition Autoregressive Models *by Daniel Buncic*	2017:344
Domestic and External Sovereign Debt *by Paola Di Casola and Spyridon Sichlimiris*	2017:345
The Role of Trust in Online Lending *by Christoph Bertsch, Isaiah Hull, Yingjie Qi and Xin Zhang*	2017:346
On the effectiveness of loan-to-value regulation in a multiconstraint framework *by Anna Grodecka*	2017:347
Shock Propagation and Banking Structure *by Mariassunta Giannetti and Farzad Saidi*	2017:348
The Granular Origins of House Price Volatility *by Isaiah Hull, Conny Olovsson, Karl Walentin and Andreas Westermark*	2017:349
Should We Use Linearized Models To Calculate Fiscal Multipliers? *by Jesper Lindé and Mathias Trabandt*	2017:350
The impact of monetary policy on household borrowing – a high-frequency IV identification *by Maria Sandström*	2018:351
Conditional exchange rate pass-through: evidence from Sweden *by Vesna Corbo and Paola Di Casola*	2018:352
Learning on the Job and the Cost of Business Cycles *by Karl Walentin and Andreas Westermark*	2018:353
Trade Credit and Pricing: An Empirical Evaluation *by Niklas Amberg, Tor Jacobson and Erik von Schedvin*	2018:354
A shadow rate without a lower bound constraint *by Rafael B. De Rezende and Annukka Ristiniemi*	2018:355
Reduced "Border Effects", FTAs and International Trade *by Sebastian Franco and Erik Frohm*	2018:356
Spread the Word: International Spillovers from Central Bank Communication *by Hanna Armelius, Christoph Bertsch, Isaiah Hull and Xin Zhang*	2018:357
Predictors of Bank Distress: The 1907 Crisis in Sweden *by Anna Grodecka, Seán Kenny and Anders Ögren*	2018:358

Diversication Advantages During the Global Financial Crisis *by Mats Levander*	2018:359
Towards Technology-News-Driven Business Cycles *by Paola Di Casola and Spyridon Sichlimiris*	2018:360
The Housing Wealth Effect: Quasi-Experimental Evidence *by Dany Kessel, Björn Tyrefors and Roine*	2018:361
Identification Versus Misspecification in New Keynesian Monetary Policy Models *by Malin Adolfson, Stefan Laseén, Jesper Lindé and Marco Ratto*	2018:362
The Macroeconomic Effects of Trade Tariffs: Revisiting the Lerner Symmetry Result *by Jesper Lindé and Andrea Pescatori*	2019:363
Biased Forecasts to Affect Voting Decisions? The Brexit Case *by Davide Cipullo and André Reslow*	2019:364
The Interaction Between Fiscal and Monetary Policies: Evidence from Sweden *by Sebastian Ankargren and Hovick Shahnazarian*	2019:365
Designing a Simple Loss Function for Central Banks: Does a Dual Mandate Make Sense? *by Davide Debortoli, Jinill Kim and Jesper Lindé*	2019:366
Gains from Wage Flexibility and the Zero Lower Bound *by Roberto M. Billi and Jordi Galí*	2019:367
Fixed Wage Contracts and Monetary Non-Neutrality *by Maria Björklund, Mikael Carlsson and Oskar Nordström Skans*	2019:368
The Consequences of Uncertainty: Climate Sensitivity and Economic Sensitivity to the Climate *by John Hassler, Per Krusell and Conny Olovsson*	2019:369
Does Inflation Targeting Reduce the Dispersion of Price Setters' Inflation Expectations? *by Charlotte Paulie*	2019:370
Subsampling Sequential Monte Carlo for Static Bayesian Models *by David Gunawan, Khue-Dung Dang, Matias Quiroz, Robert Kohn and Minh-Ngoc Tran*	2019:371
Hamiltonian Monte Carlo with Energy Conserving Subsampling *by Khue-Dung Dang, Matias Quiroz, Robert Kohn, Minh-Ngoc Tran and Mattias Villani*	2019:372
Institutional Investors and Corporate Investment *by Cristina Cella*	2019:373
The Impact of Local Taxes and Public Services on Property Values *by Anna Grodecka and Isaiah Hull*	2019:374
Directed technical change as a response to natural-resource scarcity *by John Hassler, Per Krusell and Conny Olovsson*	2019:375
A Tale of Two Countries: Cash Demand in Canada and Sweden *by Walter Engert, Ben Fung and Björn Segendorf*	2019:376
Tax and spending shocks in the open economy: are the deficits twins? *by Mathias Klein and Ludger Linnemann*	2019:377
Mind the gap! Stylized dynamic facts and structural models *by Fabio Canova and Filippo Ferroni*	2019:378
Financial Buffers, Unemployment Duration and Replacement Labor Income *by Mats Levander*	2019:379
Inefficient Use of Competitors' Forecasts? *by André Reslow*	2019:380
How Much Information Do Monetary Policy Committees Disclose? Evidence from the FOMC's Minutes and Transcripts *by Mikael Apel, Marianna Blix Grimaldi and Isaiah Hull*	2019:381
Risk endogeneity at the lender/investor-of-last-resort *by Diego Caballero, André Lucas, Bernd Schwaab and Xin Zhang*	2019:382
Heterogeneity in Households' Expectations of Housing Prices – Evidence from Micro Data *by Erik Hjalmarsson and Pär Österholm*	2019:383
Big Broad Banks: How Does Cross-Selling A Affect Lending? *by Yingjie Qi*	2020:384
Unemployment Fluctuations and Nominal GDP Targeting *by Roberto Billi*	2020:385

FAQ: How do I extract the output gap? *by Fabio Canova*	2020:386
Drivers of consumer prices and exchange rates in small open economies *by Vesna Corbo and Paola Di Casola*	2020:387
TFP news, stock market booms and the business cycle: Revisiting the evidence with VEC models *by Paola Di Casola and Spyridon Sichlimiris*	2020:388
The costs of macroprudential deleveraging in a liquidity trap *by Jiaqian Chen, Daria Finocchiaro, Jesper Lindé and Karl Walentin*	2020:389
The Role of Money in Monetary Policy at the Lower Bound *by Roberto M. Billi, Ulf Söderström and Carl E. Walsh*	2020:390
MAJA: A two-region DSGE model for Sweden and its main trading partners *by Vesna Corbo and Ingvar Strid*	2020:391
The interaction between macroprudential and monetary policies: The cases of Norway and Sweden *by Jin Cao, Valeriya Dinger, Anna Grodecka-Messi, Ragnar Juelsrud and Xin Zhang*	2020:392
Withering Cash: Is Sweden ahead of the curve or just special? *by Hanna Armelius, Carl Andreas Claussen and André Reslow*	2020:393
Labor shortages and wage growth *by Erik Frohm*	2020:394
Macro Uncertainty and Unemployment Risk *by Joonseok Oh and Anna Rogantini Picco*	2020:395
Monetary Policy Surprises, Central Bank Information Shocks, and Economic Activity in a Small Open Economy *by Stefan Laséen*	2020:396
Econometric issues with Laubach and Williams' estimates of the natural rate of interest *by Daniel Buncic*	2020:397
Quantum Technology for Economists *by Isaiah Hull, Or Sattath, Eleni Diamanti and Göran Wendin*	2020:398

Sveriges Riksbank
Visiting address: Brunkebergs torg 11
Mail address: se-103 37 Stockholm

Website: www.riksbank.se
Telephone: +46 8 787 00 00, Fax: +46 8 21 05 31
E-mail: registratorn@riksbank.se