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

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

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).

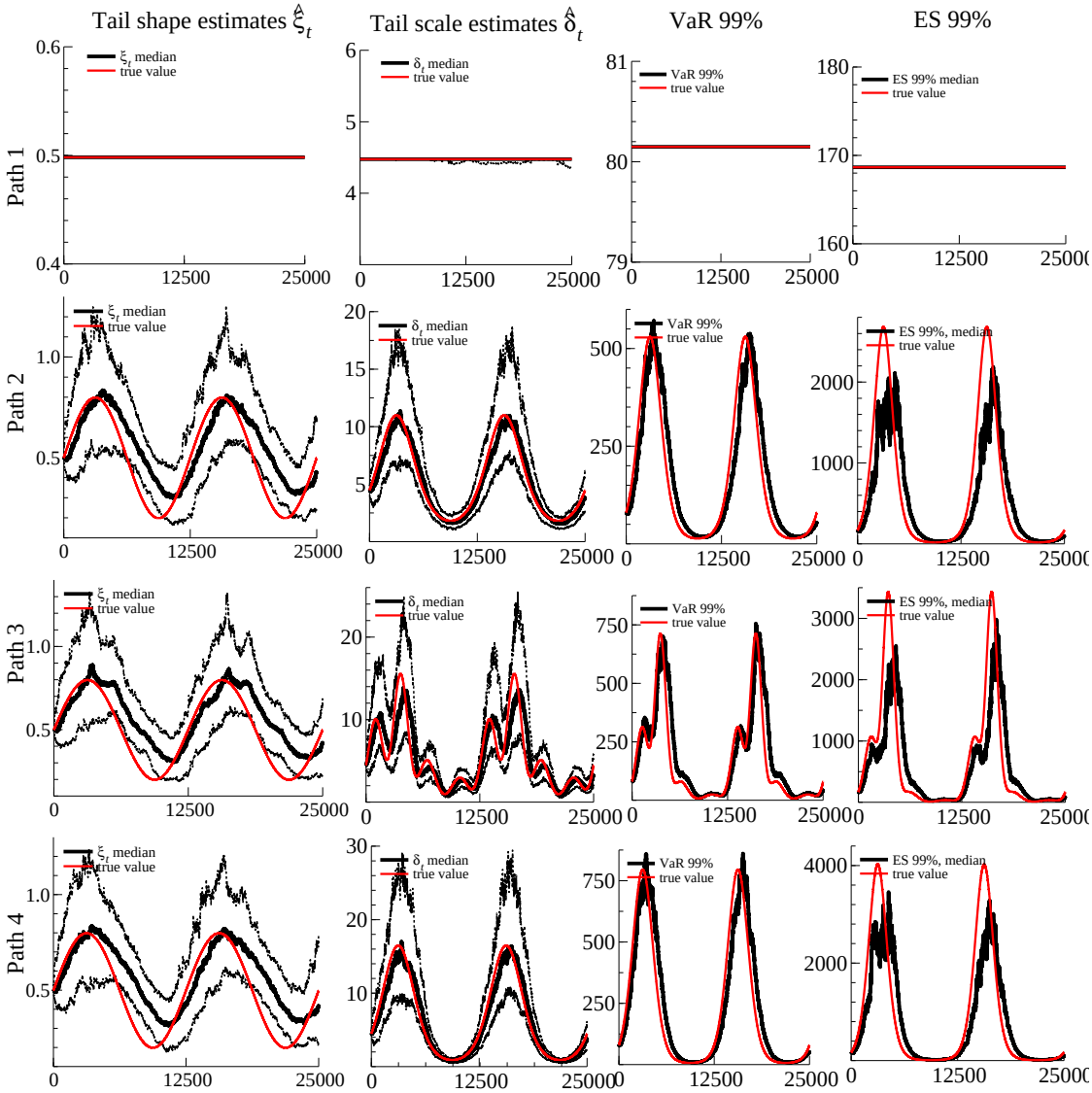
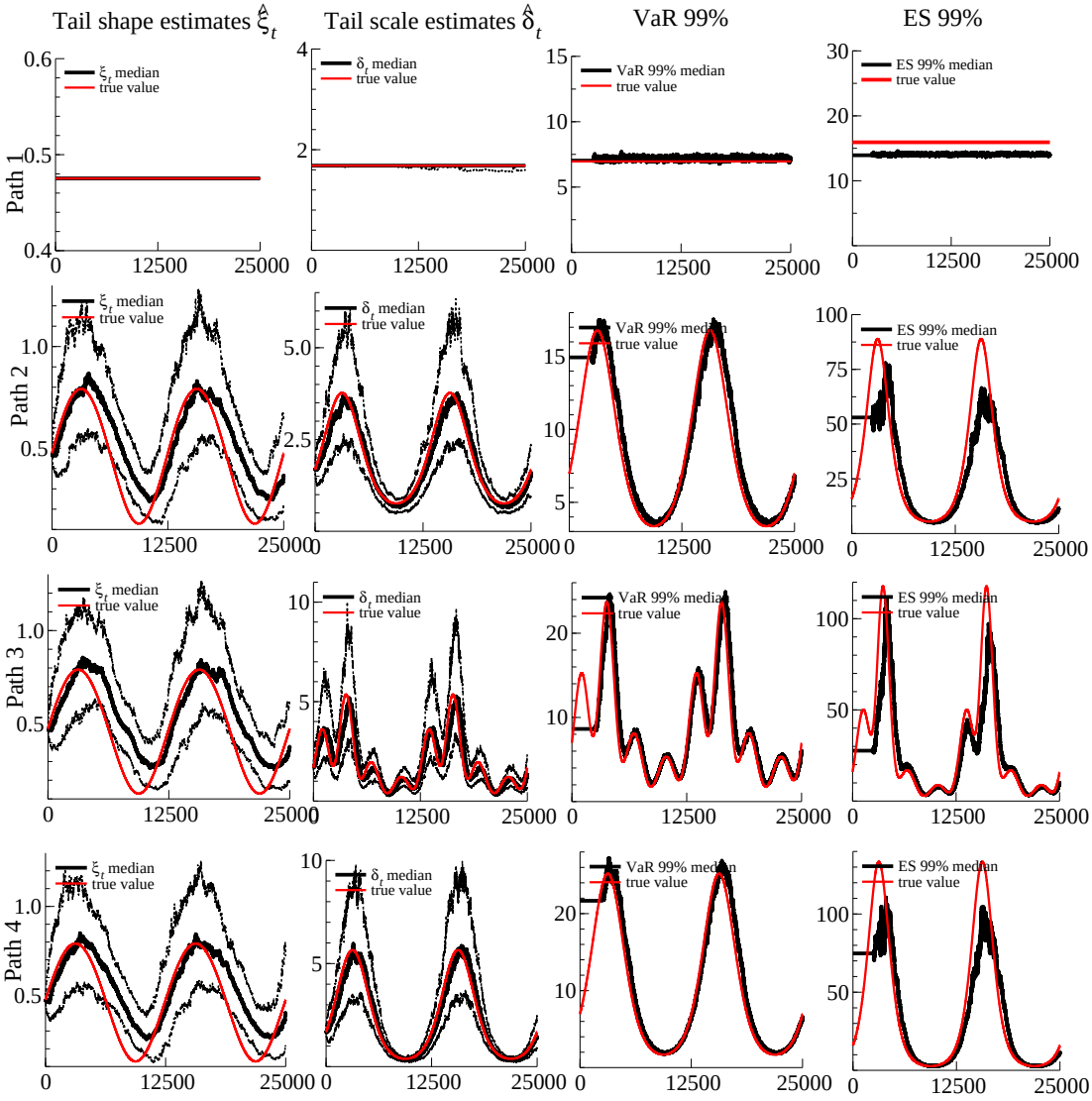


Figure H.2: Simulation results for Student's t data

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.

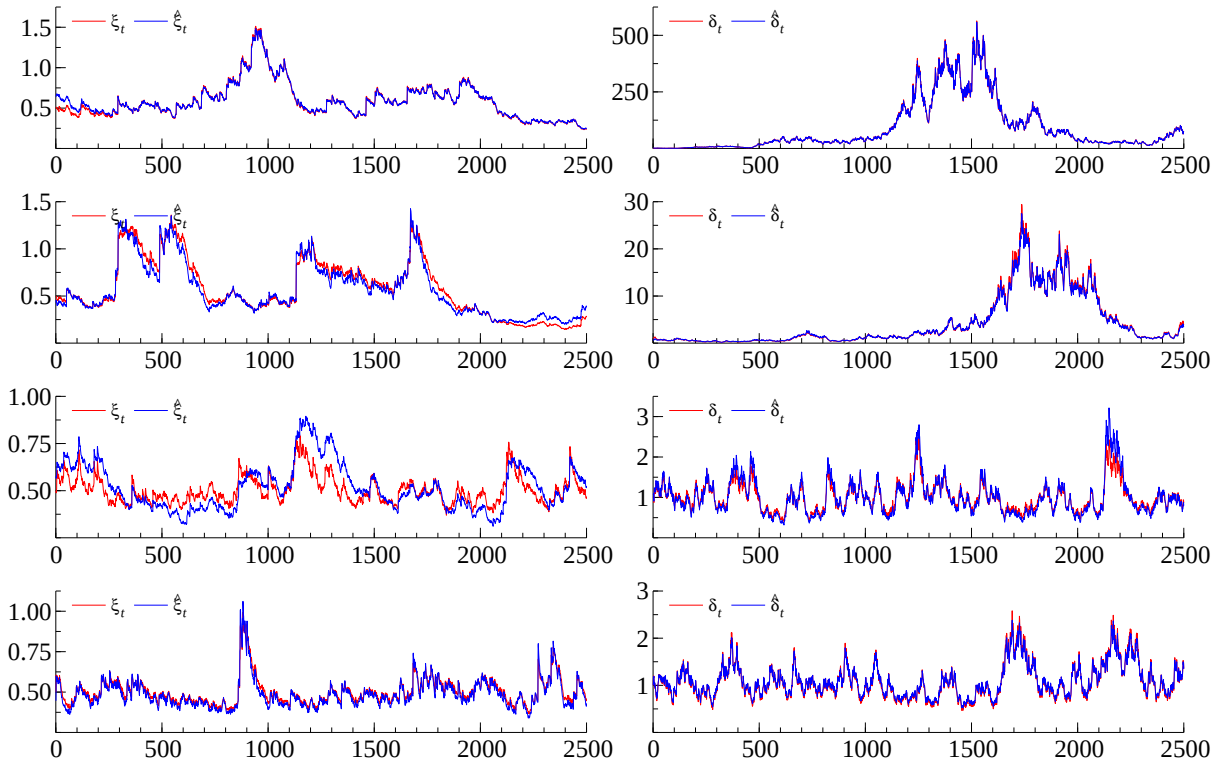


Table H.2: Simulation results: standard errors using different methods

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.

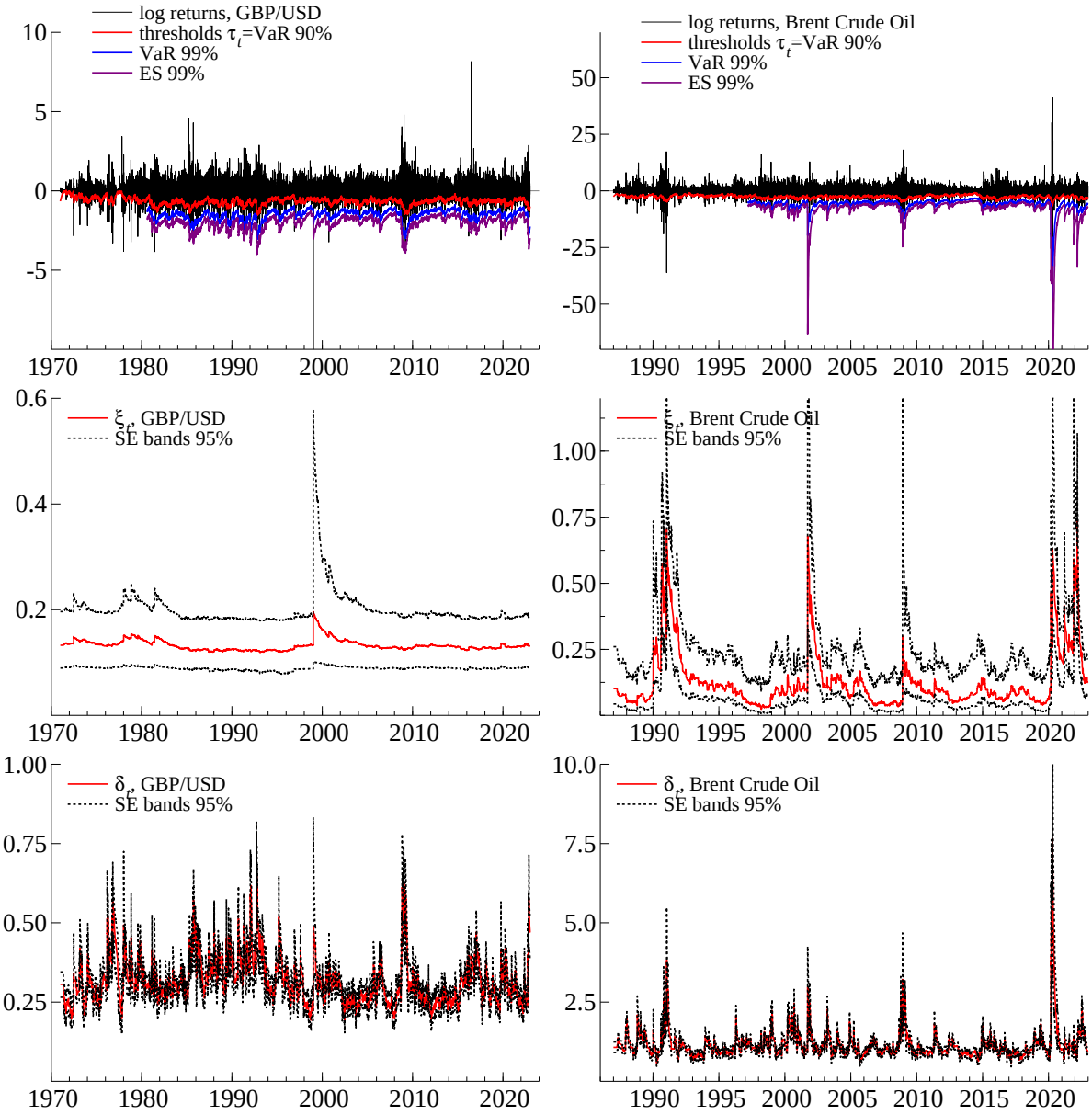
Table I.1: Parameter estimates

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.

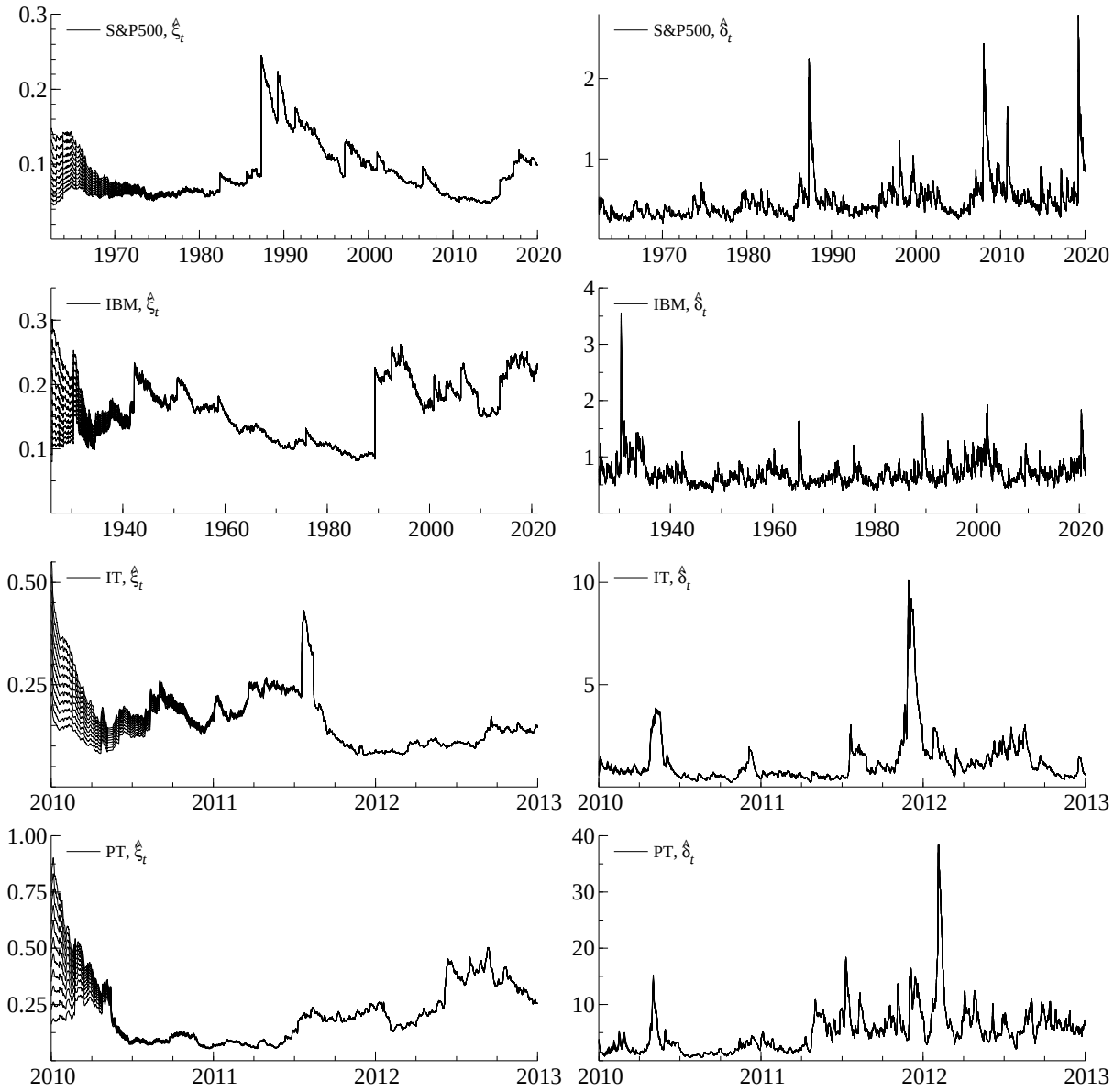


J Diagnostic checks for filter invertibility

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, $u_t = 1\text{–}(1 + \xi_t x_t/\delta_t)^{-1/\xi_t}$, 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

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

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.