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

Working Paper

Joint extreme Value-at-Rrisk and Expected Shortfall dynamics with a single integrated tail shape parameter

Sveriges Riksbank Working Paper Series, No. 446

Provided in Cooperation with:
Central Bank of Sweden, Stockholm

Suggested Citation: D'Innocenzo, Enzo; Lucas, André; Schwaab, Bernd; Zhang, Xin (2025) : Joint extreme Value-at-Rrisk and Expected Shortfall dynamics with a single integrated tail shape parameter, Sveriges Riksbank Working Paper Series, No. 446, Sveriges Riksbank, Stockholm

This Version is available at:
https://hdl.handle.net/10419/322319

SVERIGES RIKSBANK
WORKING PAPER SERIES

446

Joint extreme Value-at-Risk and Expected Shortfall dynamics with a single integrated tail shape parameter

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

February 2025

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.

Joint extreme Value-at-Risk and Expected Shortfall dynamics with a single integrated tail shape parameter*

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, Financial Research [(d)] Sveriges Riksbank

February 24, 2025

Sveriges Riksbank Working Paper Series
No. 446

Abstract

We propose a robust semi-parametric framework for persistent time-varying extreme tail behavior, including extreme Value-at-Risk (VaR) and Expected Shortfall (ES). The framework builds on Extreme Value Theory and uses a conditional version of the Generalized Pareto Distribution (GPD) for peaks-over-threshold (POT) dynamics. Unlike earlier approaches, our model (i) has unit root-like, i.e., integrated autoregressive dynamics for the GPD tail shape, and (ii) re-scales POTs by their thresholds to obtain a more parsimonious model with only *one* time-varying parameter to describe the entire tail. We establish parameter regions for stationarity, ergodicity, and invertibility for the integrated time-varying parameter model and its filter, and formulate conditions for consistency and asymptotic normality of the maximum likelihood estimator. Using

*Author information: Enzo D'Innocenzo, University of Bologna, Department of Economics, Piazza Antonio Scaravilli 2, 40122 Bologna, Italy, enzo.dinnocenzo2@unibo.it. André Lucas, Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands, a.lucas@vu.nl. Bernd Schwaab, Financial Research, European Central Bank, Sonnemannstrasse 22, 60314 Frankfurt, Germany, bernd.schwaab@ecb.europa.eu. Xin Zhang, Research Division, Sveriges Riksbank, SE 103 37 Stockholm, Sweden, xin.zhang@riksbank.se. We thank Alessandro Tenderini for excellent research assistance. The views expressed in this paper are those of the authors and do not necessarily reflect the views or policies of the European Central Bank or Sveriges Riksbank.

four exchange rate series, we illustrate how the new model captures the dynamics of extreme VaR and ES.

Keywords: dynamic tail risk, integrated score-driven models, extreme value theory.

JEL classification: *C22, G11.*

1 Introduction

For reliable inference on extreme tail behavior, Extreme Value Theory (EVT) is statistics' favorite approach. It allows the researcher to infer the distribution's extreme tail scale, shape, quantiles, and expectiles by focusing only on the tail area and abstracting from the density's center; see, for example, Balkema and de Haan (1974), Pickands (1975), Hill (1975), and Davidson and Smith (1990) for early key contributions, Embrechts et al. (1997), Coles (2001), de Haan and Ferreira (2006), and McNeil et al. (2010, Ch. 7) for textbook treatments, and Rocco (2014) for a survey. The key insights of EVT have by now been extended from the i.i.d. cross-sectional setting to time series applications; see, for example, Chavez-Demoulin et al. (2005), Chavez-Demoulin and Embrechts (2010), Einmahl et al. (2016), Hoga (2017), Massacci (2017), de Haan and Zhou (2021), and D'Innocenzo et al. (2024). This is particularly useful if EVT is applied for risk and capital buffer determination in finance and economics, for instance, for the calculation of a predictive density's 99.9% Value-at-Risk (VaR) or Expected Shortfall (ES). Such measures may change rapidly under changing market circumstances and distress.

This paper concentrates on modeling the time variation in the extreme tails of conditional loss distributions. Thus far, models for the dynamics of extreme tail behavior have had to deal with at least three major challenges: First, time variation in tail behavior requires dynamic models for the tail's starting point, its shape, and its scale, all of which are important ingredients for the computation of high distribution quantiles. A joint dynamic model for all these three ingredients quickly becomes quite complex, however, and one would benefit from an approach that reduces the number of time-varying parameters to be modeled. Second,

a model for the dynamics of extreme tail behavior should ideally only concentrate on the distribution's tail area and avoid making assumptions about the behavior of the center of the distribution. Finally, empirical estimates of continuously changing tail shapes typically indicate that such time variation is a highly persistent phenomenon (see, e.g., Massacci, 2017; de Haan and Zhou, 2021; D'Innocenzo et al., 2024) with autoregressive dynamics that often have (near) 'unit root' like behavior. This finding seems at odds with the typical stationarity or mean-reverting assumptions made in the same papers. Thus far, a theory for so-called 'integrated' models for tail risk dynamics seems to be lacking. Though results are available for integrated volatility models (such as iGARCH, see e.g. Li et al., 2018 and Francq and Zakoian, 2019) and particular location models (Blasques et al., 2022), no results are available for the behavior of integrated models for shape parameters. In such a challenging setting, it is natural to ask: can time variation in a time series' extreme tail still be estimated consistently? Can VaR and ES at high confidence levels be estimated simply and reliably? And does standard likelihood inference still apply or is it affected if tail shape dynamics are highly persistent, i.e., integrated? Despite its practical relevance for fields such as financial economics and actuarial sciences as well as its theoretical importance, a tractable comprehensive framework to address such first-order questions is currently missing.

It is here that the current paper makes its main contribution. We propose a novel robust, semi-parametric, and dynamic framework with 'integrated' (i.e., persistent) time variation in tail fatness for long univariate time series. The framework builds on results from the EVT literature and uses a conditional Generalized Pareto Distribution (GPD) to approximate the tail beyond a given threshold. The time-varying tail shape in our model is 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 parameter is 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. Score-driven dynamics are known to be optimal in the sense of Blasques

et al. (2015).

Our approach is different from previous EVT studies (including D'Innocenzo et al., 2024) in at least two important ways. First, we do not apply the limiting GPD result from EVT to the peaks-over-thresholds (POTs), but to *scaled* POTs, where the scaling is done by the threshold value. This stands in sharp contrast to – virtually all – earlier applications that typically focus on unscaled POTs (see, for example, McNeil et al., 2010, Ch. 7, Christoffersen, 2012, Ch. 6, Andersen et al., 2013, Hoga (2017), and Massacci, 2017). The use of scaled POTs has been remarkably under-explored in risk management, yet has an important advantage: the limiting GPD approximation is now characterized by a *single* shape parameter and no longer needs both a tail shape and tail scale parameter. The resulting statistical model is parsimonious and much simpler to study theoretically and empirically.

Second, our model deviates from previous approaches in that we consider an *integrated* score-driven filtering equation for the time-varying shape parameter. Empirically, when studying for instance daily or intra-daily financial data, estimates of the autoregressive parameter for the tail shape dynamics are often indistinguishably close to unity, implying highly persistent dynamics (see, e.g., Massacci (2017) and D'Innocenzo et al., 2024). We study the asymptotic properties of such an integrated model in detail, including stationarity and ergodicity properties of the model and the filter, and consistency and asymptotic normality of the model's static parameters. This extends the work on integrated models for higher-order moments from the volatility case (for instance, Jensen and Rahbek, 2004; Francq and Zakoïan, 2012; Francq and Zakoian, 2019) to the EVT setting.

To obtain the time-varying thresholds for scaling, we adapt the recent approach of Patton et al. (2019), which elicits VaR and ES simultaneously in a semi-parametric way, concentrating only on tail observations and not making assumptions about the center of the distribution. The method is therefore extremely useful for estimating threshold values less far out in the tails, such as at 90% or 95% confidence levels. The method faces more challenges for more extreme quantiles. It is here that our dynamic EVT-based GPD approximation

perfectly complements the approach of Patton et al. (2019), as our approach is precisely geared towards modeling the *extreme* tail quantiles.

We obtain two theoretical results. First, we show that under mild regularity conditions the tail shape parameter and the data are asymptotically stationary and ergodic. Moreover, we characterize the invertibility region for the tail shape filter. Interestingly, while the tail shape parameter has integrated dynamics and no finite unconditional mean to revert to, the *ratio* of the true and the estimated tail shape parameter is well-behaved, asymptotically stationary and ergodic, and has a finite unconditional first moment. We also show that the intercept in the DGP for the time-varying tail shape needs to be strictly positive to rule out degenerate limiting behavior of the tail shape. Second, we show that the maximum likelihood estimator for the model's static parameters is strongly consistent and asymptotically normally distributed under mild regularity conditions, despite the integrated dynamics. We confirm the theoretical findings in simulation experiments.

We illustrate the model using several exchange rate examples. First, we consider daily log-returns for the Euro (EUR) and Russian Ruble (RUB) viz-a-viz the U.S. dollar (USD). The adverse right tail is the relevant tail for any firm such as a typical financial intermediary, which obtains revenues or holds assets in local currency and has financed its operations, at least in part, through foreign currency (USD) liabilities. We find that both tail shape parameters vary significantly over time, taking values between approximately 0.2 and 0.5 for EUR/USD and between approximately 0.2 and 1.3 for RUB/USD. These values imply a maximum number of conditional moments in the range 2–5 for EUR/USD, and 0.8–5 for RUB/USD. Market risk estimates for EUR/USD, such as VaR and ES at a 99% confidence level, spike during the global financial crisis between 2008 and 2009 and the euro area sovereign debt crisis between 2010 and 2012. By contrast, market risk estimates for RUB/USD increase markedly following Russia's war of aggression against Ukraine starting in 2014 and 2022. These are higher than the increases after the outbreak of Covid-19 in 2020, but lower than those following the so-called Ruble crisis of 1998.

Second, we consider hourly log-returns for Bitcoin (BTC/USD) and Ether (ETH/USD). Both time series include the so-called "second crypto winter" of 2022. Again, we find that the tail shape parameters vary significantly over time. The adverse left tail's shape parameter varies between 0.3 and 0.6 for BTC and ETH, implying the existence of between 1 up to 3 conditional moments, depending on the period, and thus the presence of extremely fat tails. Tail market risk estimates responded strongly to the collapse of the Terra/Luna cryptocurrency in May 2022 (see, e.g. Uhlig, 2022), the collapse of the cryptocurrency intermediary FTX in June 2022, and the collapse of crypto intermediary Celsius in November 2022.

The four papers closest to ours are Massacci (2017), Patton et al. (2019), de Haan and Zhou (2021), and D'Innocenzo et al. (2024). de Haan and Zhou (2021) propose a fully non-parametric approach to estimating a continuously-changing extreme value index locally from independent but non-identically distributed POTs. Our paper is different in that we adopt a semi-parametric perspective, using a parametrized, integrated filtering recursion to recover persistent time variation in the tail's shape. Massacci (2017) and D'Innocenzo et al. (2024) both study score-driven approaches to filtering the extreme tail's scale *and* shape. Our paper is different in two important ways, in that we propose a particularly parsimonious statistical model (featuring only a single time-varying parameter), and focus on the case of *integrated* time variation in the tail. 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. Our score-driven dynamics contain weights for extreme observations, which 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. Formulated differently, our approach and that of Patton et al. (2019) complement each other. Whereas Patton et al. (2019) provide an appropriate semi-parametric framework to estimate time-varying thresholds less far out in the tails, our approach enables the identification of time variation in risk measures in the extreme tails beyond these thresholds.

Section 2 presents the statistical model. Section 3 discusses the asymptotic properties

of the model and of the maximum likelihood estimator. Section 4 studies the model's performance in simulation experiments. Section 5 applies the model to different exchange rate returns. Section 6 concludes. Proofs and additional results are provided in a web appendix.

2 Statistical model

2.1 A scaled conditional EVT framework

Consider a random variable X_t, such as an asset price or return. We are interested in describing the conditional extreme tail behavior of X_t. We do this by adopting an Extreme Value Theory (EVT) perspective. This allows us to only concentrate on the conditional tail behavior of X_t, leaving any changes in the center of the distribution unmodeled as these are not relevant for our prime objective: estimating the extreme conditional quantiles of X_t. For our main result, we formulate a conditional version of the Pickands-Balkema-de Haan Peaks-Over-Threshold (POT) theorem. This theorem describes the behavior of X_t in the far-out tail area, i.e., for X_t values above some high (possibly time-varying) threshold value τ_t. Following Theorem 1.2.5 of de Haan and Ferreira (2006), the conditional extremal behavior of a random variable X_t that lies in the domain of attraction of a (heavy-tailed) Fréchet law with tail shape $f_t > 0$, can be described by

$$\lim_{\tau_t \to \infty} \mathbb{P}\left(X_t > \tau_t + \tau_t\, f_t\, x_t \mid X_t > \tau_t,\, \mathcal{F}_{t-1}\right) = \left(1 + f_t\, x_t\right)^{-1/f_t},$$

for $x_t > 0$, $f_t \in \mathcal{F}_{t-1}$, and $\mathcal{F}_{t-1} = \{X_1, \ldots, X_{t-1}\}$ denoting the conditioning set. We refer to the reciprocal of the tail shape, $1/f_t$, as the tail index. Distributions that satisfy this condition comprise most fat-tailed distributions used in economics and finance, such as the Student's t distribution, the (generalized) Pareto distribution, the log-gamma distribution, the F distribution, and many more (for further discussion, see e.g. Johnson et al., 1994, Embrechts et al., 1997, and McNeil et al., 2010, Ch. 7.3.)

We define the scaled POTs Y_t as $Y_t = (X_t - \tau_t)/\tau_t \Leftrightarrow X_t = \tau_t + \tau_t Y_t$. Substituting Y_t into the above limiting result and defining $y_t = f_t x_t$, we obtain

$$\begin{aligned}\lim_{\tau_t \to \infty} \mathbb{P}\left(Y_t > y_t \mid Y_t > 0,\, \mathcal{F}_{t-1}\right) &= \lim_{\tau_t \to \infty} \mathbb{P}\left(\tau_t + \tau_t Y_t > \tau_t + \tau_t y_t \mid \tau_t + \tau_t Y_t > \tau_t,\, \mathcal{F}_{t-1}\right) \\ &= \lim_{\tau_t \to \infty} \mathbb{P}\left(X_t > \tau_t + \tau_t y_t \mid X_t > \tau_t,\, \mathcal{F}_{t-1}\right) = (1 + y_t)^{-1/f_t}, \quad (1)\end{aligned}$$

which yields the generalized Pareto distribution (GPD) as a limiting approximation for the conditional distribution $\mathbb{P}(Y_t > y_t \mid Y_t > 0,\, \mathcal{F}_{t-1})$ of the extreme tail of the scaled POTs Y_t.

The advantage of using scaled POTs $Y_t = (X_t - \tau_t)/\tau_t$ rather than their unscaled counterparts $(X_t - \tau_t)$ is that it considerably simplifies the resulting expression for the distribution function in (1) for the extreme tails compared to, for instance, Massacci (2017) or D'Innocenzo et al. (2024). In particular, the expression in (1) is characterized by only *one* time-varying conditional tail shape parameter f_t, rather than by both a tail shape and tail scale parameter as in earlier papers. This simplification proves particularly helpful when deriving the asymptotic properties of the model and the maximum likelihood estimator later.

2.2 Filtering the conditional tail shape parameter f_t

We expect the tail shape to possibly change gradually over time. Therefore, we introduce score-driven dynamics for f_t as in Creal et al. (2013), with $f_{t+1} = \omega + \beta f_t + \alpha s_t$, where s_t is the inverse information scaled derivative of the log predictive GPD tail density. Transforming the cdf expression in (1) into a conditional pdf $p(y_t \mid \mathcal{F}_{t-1}) = f_t^{-1}\,(1 + y_t)^{-f_t^{-1}-1}$ for $y_t > 0$, we obtain the following expression for the scaled score,

$$\nabla_t = \frac{\partial\left(-\ln(f_t) - \left(f_t^{-1} + 1\right)\ln\left(1 + y_t\right)\right)}{\partial f_t} = \frac{1}{f_t^2}\ln\left(1 + y_t\right) - \frac{1}{f_t},$$

$$\mathcal{I}_{t-1} = \mathbb{E}\left[\nabla_t^2 \;\middle|\; \mathcal{F}_{t-1}\right] = f_t^{-2}, \qquad s_t = \mathcal{I}_{t-1}^{-1}\,\nabla_t = \ln(1 + y_t) - f_t.$$

In this paper, we are particularly interested in filtering the tail shape parameter f_t from the data using an *integrated* score-driven filtering equation,

$$f_{t+1} = \omega + f_t + \alpha\, s_t, \tag{2}$$

i.e., a model where we have set β, the coefficient in front of f_t on the right-hand side of (2), equal to unity rather than a value inside the unit interval as is commonly done in previous literature. Note that if $f_1, \omega > 0$ and $0 < \alpha < 1$, then f_t is non-negative for all $t \in \mathbb{N}$ by construction. With a slight abuse of notation, we only change the value of f_t if a POT materializes. It is then, and only then, that we obtain information about the tail shape behavior of Y_t and thus X_t. In all other cases, we only obtain information about the center of the distribution of X_t, which is irrelevant for the time variation in the extreme quantiles.

We use the terminology *integrated* score-driven model similarly as in the integrated GARCH (iGARCH) literature. Whereas iGARCH models have been well-studied (see, e.g., Francq and Zakoian, 2019, and references therein), integrated score-driven models have thus far received hardly any attention. This is remarkable given the fact that empirical estimates of β for score-driven models are often quite close to unity. In a recent paper, Blasques et al. (2022) study an integrated score-driven filter in the particular setting of a time-varying location model for a mixture of two normals. Properties of integrated score-driven models for time-varying parameters beyond the location-scale setting, however, are to the best of our knowledge absent from the current literature. Integrated tail shape dynamics and thus a slowly time-varying tail shape f_t make perfect empirical sense, however, particularly for longer time spans of highly frequent data such as daily or intra-daily data.

The presence of an intercept ω in (2) when f_t has a unit autoregressive coefficient may seem strange at first sight. It is not always standard (see, for example, the ZD-GARCH models studied in Li et al., 2018), but has been shown to be important before in a time-varying volatility setting (see e.g. Francq and Zakoïan, 2012; Francq and Zakoian, 2019).

We show in Section 3 that a non-zero intercept $\omega > 0$ is crucial if one wishes to interpret the score-driven tail shape model as a data generating process: without it, convergence of f_t to its stationary and ergodic limit always results in degenerate, thin-tailed tail behavior with $f_t = 0$ for all t.

Given the observation-driven nature of the filtering equation (2), an explicit expression is available for the likelihood function. Estimates of the model's static parameters can then be obtained by standard maximum likelihood methods. We gather all the model's static parameters in a parameter vector $\theta = (\omega, \alpha)$. To write down the likelihood function, we only use the POT observations, i.e., the observations with $X_t > \tau_t$ or $Y_t > 0$. Let $T^\star$ denote the number of POTs in the sample. We can then write the maximum likelihood estimator (MLE) $\hat{\theta}_T$ as

$$\hat{\theta}_T = \arg\max_{\theta \in \Theta} \hat{L}_T(\theta), \qquad \hat{L}_T(\theta) = \frac{1}{T^\star} \sum_{y_t > 0} \hat{\ell}_t(\theta), \tag{3}$$
$$\hat{\ell}_t(\theta) = -\ln \hat{f}_t(\theta) - \left(1 + \frac{1}{\hat{f}_t(\theta)}\right) \ln(1 + y_t),$$

where we use the slightly more precise notation $\hat{f}_t(\theta)$ to denote the filtered outcomes as a function of the static parameters,

$$\hat{f}_{t+1}(\theta) = \omega + \hat{f}_t(\theta) + \alpha \left(\ln(1 + y_t) - \hat{f}_t(\theta)\right), \tag{4}$$

evaluated at some θ inside the parameter space Θ and initialized at some $\hat{f}_1 > 0$, and where $\hat{f}_t(\theta)$ only changes upon the realization of a POT $y_t > 0$.

2.3 Filtering extreme Value-at-Risk and Expected Shortfall

Evaluated at their maximum likelihood estimates, the filtered values $\hat{f}_t(\hat{\theta}_T)$ can be used to compute familiar risk quantities like extreme tail VaR or extreme tail ES using the GPD

approximation; see, for instance, McNeil and Frey (2000), McNeil et al. (2010) and Rocco (2014). For an extreme quantile with a tail probability $\gamma < (1+\tau_t)^{-1/\hat{f}_t(\hat{\theta}_T)}$ lying beyond a high threshold $\tau_t > 0$, we can estimate the VaR and ES by

$$\mathrm{VaR}^{\gamma}(x_t \mid \mathcal{F}_{t-1}) = \tau_t \cdot \left(\frac{\gamma}{n_t/t}\right)^{-\hat{f}_t(\hat{\theta}_T)}, \tag{5}$$

$$\mathrm{ES}^{\gamma}(x_t \mid \mathcal{F}_{t-1}) = \frac{\mathrm{VaR}^{\gamma}(x_t \mid \mathcal{F}_{t-1})}{1-\hat{f}_t(\hat{\theta}_T)}, \tag{6}$$

where n_t is the number of POTs up to time t, such that n_t/t estimates the tail probability $\mathbb{P}(X_t > \tau_t)$. The expressions (5) and (6) differ from those in, for instance, McNeil et al. (2010) and D'Innocenzo et al. (2024) owing to the use of scaled POTs $Y_t = (X_t - \tau_t)/\tau_t$ rather than their unscaled counterparts $(X_t - \tau_t)$; see Web Appendix C for derivations. In particular, our expressions only require the estimation of the tail shape parameter $\hat{f}_t(\hat{\theta}_T)$, and not of any auxiliary tail scale.

Given the straightforward formulation of the filter for $f_t(\theta)$ in (4), (5)–(6) directly yield our desired filtered extreme risk measures. These filtered VaR and ES estimates differ in an important way from those in Patton et al. (2019). Whereas the approach of Patton et al. (2019) is very useful for less extreme quantiles, the current EVT-based filtering model with integrated dynamics has important advantages for capturing time variation farther out in the tail. A major advantage of the elicitation function used by Patton et al. (2019) for characterizing the tail quantile and expectile is that it does not hinge on any (possibly incorrect) distributional assumptions. At the same time, the approach comes with a similar limitation for risk assessment as historical simulation has compared to EVT-based methods: historical simulation cannot capture the extreme tail shape beyond the highest realization; see, for instance, McNeil et al. (2010). A similar risk exists for the time variation in VaR and ES using Patton et al. (2019) at extreme quantiles: there, only few POTs are available (if any at all), resulting in only few changes in the dynamic VaR and ES estimates using their approach.

We still consider the methodology of Patton et al. (2019), or one similar to it, as key to our analysis, but only for the determination of the (less extreme) thresholds τ_t; see Sections 2.4 and 5. Its advantages are (i) its reliance on very few distributional assumptions, and (ii) its use of much higher exceedance probabilities (10% or 5%) and thus the occurrence of a sizable number of POTs to capture the time variation in the thresholds τ_t. Beyond these 'less extreme' thresholds τ_t, however, we exploit the shape of the EVT-based GPD of our current paper to go much deeper into the tail. The latter has two advantages. First, we lean on the theoretical insight that the POTs of fat-tailed distributions (i.e., that lie in the domain of attraction of a Fréchet law) are themselves fat-tailed. Using the score-driven dynamics in (4), this information is directly exploited when filtering the tail shape values from the data, resulting in a milder impact of extreme exceedances on the tail shape. Second, we can use all the observed POTs at these less extreme exceedance levels for filtering the time-varying tail shape and thus the extreme ES. As a result, our filtered extreme ES values based on the GPD EVT approximation may vary much more over time compared to when they would have been estimated directly using Patton et al. (2019).

Both features are illustrated in Figure 1 using a small simulation experiment for a high 99.9% confidence level. We simulate a large sample of $T = 100,000$ observations from a standard Student's t distribution with an inverse degrees of freedom that follows a sinusoidal pattern between $\nu = 3$ and $\nu = 15$, and a matching time-varying scale such that the (non-extreme) true 5% VaR has a different pattern over time, non-synchronous with the sinusoidal pattern for ν. Panel (a) provides the true ES, the ES as estimated using Patton et al. (2019) and labeled PZC from now on, and the ES using the EVT-based methodology proposed in this paper. The EVT approach here bases its 5% tail area thresholds τ_t on Patton et al. (2019). Note that we cast our EVT-based VaR and ES to the negative outcome space to make them comparable to those of Patton et al.As expected, the dynamics of the EVT-based approach follow the true ES dynamics much more closely in terms of the up and downward movements. For the extreme 99.9% quantiles, there are simply too few POTs to induce

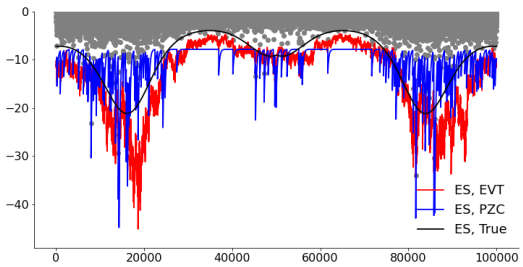
sufficient time variation in the original approach of Patton et al. (2019). This causes that approach to exhibit large jumps followed by quick reversals to an 'equilibrium level' closer to zero. This is even clearer if we focus on a shorter interval (Panel (b)): the PZC curve only jumps occasionally, as expected, and then geometrically converges to its upper bound. By contrast, the EVT-based curve behaves much more smoothly, and follows the true extreme ES more closely by extrapolating the behavior of the 5% tail observations into the extreme 0.1% tail area. The pattern for the VaR is very similar (Panel (d)).

Figure 1's Panel (c) reports the News Impact Curve (NIC) associated with each method. The robustness feature of the new EVT-based approach is readily apparent. The NIC for the VaR (not shown) looks very similar. The parameters used for the NIC are the ones estimated for the simulated data, and the curves are shifted vertically to both start in the origin. The NIC of PZC for extreme quantiles is flat until the (extreme) 99.9% VaR is exceeded. Only upon an exceedance of the extreme VaR, the ES reacts linearly and quite steeply to data. This results in the sharp peaks down and subsequent exponential reversals seen in Figure 1's Panels (b) and (d). The EVT approach is based on the less extreme 5% quantiles of PZC. Therefore, the EVT approach's extreme ES reacts much earlier to data, namely to the PoTs exceeding the less extreme τ_t. It also reacts in a milder, concave way. The concave reaction follows from the core of the EVT's NIC expression, which reduces to $C \cdot |x_t|^c$ for $c = \alpha \log(\gamma_\tau/\gamma)$ and for some C that does not depend on x_t, and where γ_τ is the tail probability of τ; see Web Appendix A for a derivation. As long as $c < 1$, the EVT approach reacts to extreme POTs in a concave, robust way to the data: it acknowledges that outliers may occur deep into the tail area if f_t is high. As a result, f_t reacts less strongly to such outliers, resulting in a more stable pattern for extreme VaR and ES (Panels (a), (b), and (d)). For typical empirical estimates (Section 5), α is estimated at a low value, such that the inequality $c < 1$ is easily satisfied and the robustness of the EVT approach is achieved.

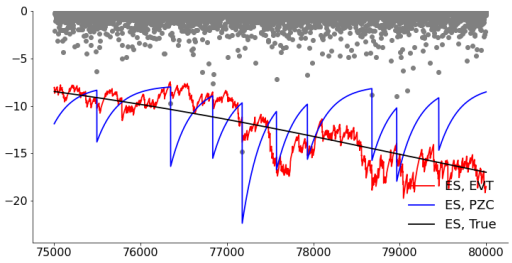
Figure 1: News Impact Curves (NIC) and EVT-based versus PZC filtering results at 99.9%

Panels (a), (b), (d): Time series plots of the 99.9% ES and VaR for the PZC method of Patton et al. (2019) and the EVT-based methodology proposed in this section. The thresholds τ_t for the EVT approach are here based on Patton et al. (2019), but using the 5% tail. The results for the EVT approach (τ, VaR, and ES) are made negative to make them comparable to the PZC results. Data are simulated from a unit scale Student's t distribution with time-varying inverse degrees of freedom ν_t^{-1} that moves sinusoidal between 0.067 ($\nu = 15$) and 0.4 ($\nu = 2.5$) and a 5% VaR that moves in a triangular way from -3 up to -1 and down to -3 again. Panel (a) shows the results for the full sample of $T = 100,000$ observations; panels (b) and (d) zoom in on a data segment to better visualize the patterns. Panel (c) plots the news impact curve associated with each method.

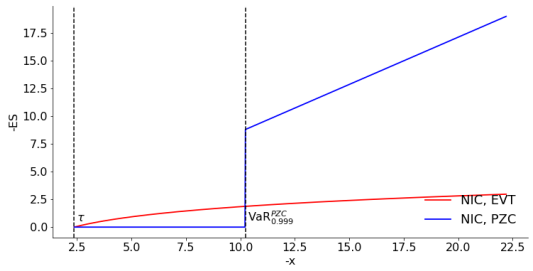
(a) ES dynamics: full sample



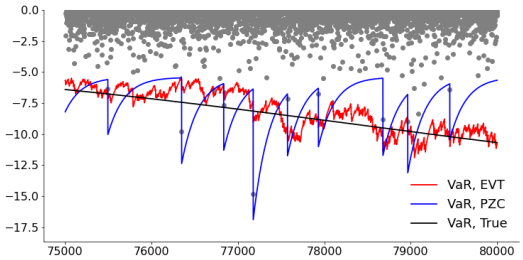
(b) ES dynamics: zoomed in



(c) NIC of ES



(d) VaR dynamics: zoomed in



2.4 Time-varying thresholds τ_t

A substantial number of time series models are available to estimate the thresholds τ_t. Suitable models include, among others, the CaViaR model of Engle and Manganelli (2004), the joint model for VaR and ES of Patton et al. (2019), and the score-driven models of Catania and Luati (2023) and D'Innocenzo et al. (2024).

Unless indicated otherwise, we estimate the time-varying thresholds τ_t using a score-driven model similar to D'Innocenzo et al. (2024) and Patton et al. (2019). The model's

implied VaRs turn out to have good coverage in the illustrations of Section 5, and its parameters are straightforward to estimate. In our specification,

$$\tau_{t+1} = \omega^\tau + a_1^\tau \cdot \big(\mathbb{1}\{x_t > \tau_t\} - (1-\kappa)\big) + a_2^\tau \cdot \big(\mathbb{1}\{x_t > \tau_t\} - (1-\kappa)\big)(x_t - \tau_t) + b^\tau \tau_t, \quad (7)$$

with $\omega^\tau \equiv (1-b^\tau)\cdot \hat{q}^\kappa$, and where $\mathbb{1}\{A\}$ is the indicator function for the event A, $\hat{q}^\kappa$ is the (observed) unconditional κ-quantile of x_t, $a_1^\tau > 0$, $a_2^\tau > 0$, and $0 < b^\tau < 1$ are three 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%. Note that $\mathbb{E}[\mathbb{1}\{x_t > \tau_t\} \mid \mathcal{F}_{t-1}, \theta^\tau] = (1-\kappa)$. 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 $a_1^\tau \kappa$ if $x_t > \tau_t$, i.e., if the previous quantile was exceeded, and a negative shock of $-a_1^\tau(1-\kappa)$ otherwise. Following Patton et al. (2019), the presence of $a_2^\tau > 0$ implies that the adjustment is stronger if the threshold exceedance $(x_t - \tau_t)$ is larger. If $a_2 = 0$, then (7) reduces to the model used in D'Innocenzo et al. (2024). For $0 < b^\tau < 1$, the empirical unconditional quantile $\hat{q}^\kappa$ serves as a long-term attractor for (7).

The parameters a_1^τ, a_2^τ, and b^τ in (7) cannot be estimated using (3). Another objective function is needed for this. Given we only need the threshold value τ_t and not the ES in this step, we use a simpler objective function than that of Patton et al. (2019), which can be used to elicit the threshold together with the ES. In particular, we use the standard tick-loss function that is also used for quantile regression (Koenker, 2005, Ch. 3). The optimization problem can then be formulated as

$$\begin{aligned} \min_{\{a_1^\tau, a_2^\tau, b^\tau\}} \frac{1}{T}\sum_{t=1}^{T} \rho_\kappa\left(x_t - \tau_t\right) \iff \min_{\{a_1^\tau, a_2^\tau, b^\tau\}} \frac{1}{T}\sum_{t=1}^{T} (x_t - \tau_t)\left(\kappa - 1\{x_t < \tau_t\}\right) \\ \iff \max_{\{a_1^\tau, a_2^\tau, b^\tau\}} \frac{1}{T}\sum_{t=1}^{T} (x_t - \tau_t)\left((1-\kappa) - 1\{x_t > \tau_t\}\right), \end{aligned} \quad (8)$$

where $\rho_\kappa(u_t) = u_t\left(\kappa - 1\{u_t < 0\}\right)$, and τ_t evolves as in (7). Engle and Manganelli (2004)

and Catania and Luati (2023) also use this objective function, albeit in different dynamic contexts. In practice, we estimate all thresholds τ_t via (8) using a similar numerical approach as Patton et al. (2019, Appendix C) before maximizing (3). All our results are conditional on the time-varying thresholds τ_t. The effects of this two-step approach are studied through simulations in Section 4.

3 Asymptotic behavior

This section studies the asymptotic properties of the model (1) and (4). We first derive conditions for stationarity and ergodicity of the model and the model-implied filter. These can then be used to establish the consistency and asymptotic normality of the maximum likelihood estimator for the model's static parameters.

We first define two random variables, namely a standard uniform $u_t \sim U(0,1)$, and a standard unit exponential $\epsilon_t = -\ln(1-u_t) \sim \mathcal{E}xp(1)$. Define $G(y_t \mid f_t) = 1-(1+y_t)^{-1/f_t}$ as the expression for the tail GPD approximation from (1). We let $f_t(\theta_0)$ denote the true time-varying tail shape parameter in the DGP as characterized by the true static parameter vector θ_0. We show later that $f_t(\theta_0)$ is the unique stationary and ergodic limit of its initialized counterpart $\hat{f}_t(\theta_0)$ from (4), initialized at $\hat{f}_1$.

Using these definitions, we obtain

$$G(y_t \mid f_t(\theta_0)) = u_t = 1-(1+y_t)^{-1/f_t(\theta_0)} \iff \frac{1}{f_t(\theta_0)}\ln(1+y_t) = -\ln(1-u_t) = \epsilon_t, \quad (9)$$

and thus

$$f_{t+1}(\theta_0) = \omega_0 + f_t(\theta_0) + \alpha_0\left(\ln(1+y_t) - f_t(\theta_0)\right) = \omega_0 + \left(1+\alpha_0\left(\epsilon_t - 1\right)\right) f_t(\theta_0). \quad (10)$$

We make the following assumptions.

Assumption 1. $\{\epsilon_t\}_{t\in\mathbb{Z}}$ *is an independent and identically distributed (IID) noise sequence where each* ϵ_t *has a unit exponential distribution.*

Assumption 2. *The parameter space satisfies* $\Theta = \{\theta \mid 0 < \underline{\omega} \leq \omega \leq \bar{\omega} < \infty,\ 0 < \underline{\alpha} \leq \alpha \leq \bar{\alpha} < 1\}$*, and the true value* $\theta_0 \in$ *int*(Θ)*.*

Both assumptions are mild and quite standard. Assumption 1 postulates that we can generate the non-linear time series dynamics for the tail shape model by feeding an *IID* noise process to the inverse cdf of the GPD to obtain realizations of y_t. These realizations then feed into the next tail shape parameter via the recursion (10). Assumption 2 is standard and establishes that the parameter space is compact and that the true parameter lies in its interior. Again, the restrictions on the parameter space are unsurprising: $\omega > 0$ and $0 < \alpha < 1$ jointly ensure that the tail shape parameter remains non-negative for all values of t. We now obtain the following theorem, which establishes (i) stationarity and ergodicity of the data y_t and of the uninitialized true time-varying parameter $f_t(\theta_0)$; (ii) invertibility of the filter $\hat{f}_t(\theta)$ started at $\hat{f}_1$ and evaluated at a generic value $\theta \in \Theta$; and (iii) the existence of appropriate moments to establish the consistency of the maximum likelihood estimator later on. All proofs are found in Web Appendices A and B.

Theorem 1. *Under Assumptions 1–2:*

(i) the model is stationary and ergodic, i.e., there exists a unique stationary and ergodic solution $f_t(\theta_0)$ *and* y_t *to* (9) *and* (10)*; moreover, there exists some small* $r > 0$ *such that* $\mathbb{E}|f_t(\theta_0)|^r < \infty$ *and* $\mathbb{E}\left|\ln(1+y_t)\right|^r < \infty$*;*

(ii) if $\mathbb{E}\ln^+\ln(1+y_t) < \infty$*, then the model-implied filter is invertible, i.e.,* $\hat{f}_t(\theta)$ *as generated by* (4) *and initialized at* $\hat{f}_1$ *converges to a unique stationary and ergodic solution* $f_t(\theta)$ *uniformly over* Θ*;*

(iii) the ratio process $\hat{z}_t^f(\theta) = f_t(\theta_0)/\hat{f}_t(\theta)$ *converges to a unique stationary and ergodic solution* $z_t^f(\theta) = f_t(\theta_0)/f_t(\theta)$*. Moreover,* $z_t^f(\theta)$ *and* $1/z_t^f(\theta)$ *have finite n-th order*

moment, uniformly over Θ*, for any* $n > 0$*.*

Given the model's structure, the two simple Assumptions 1 and 2 suffice to obtain stationarity and ergodicity of the time-varying tail shape in the DGP. Interestingly, the result also gives rise to the following corollary.

Corollary 2. *Under Assumptions 1 and 2:*

(i) if $\omega_0 = 0$*, the stationary and ergodic solution for* $f_t(\theta_0)$ *satisfies* $f_t(\theta_0) = 0$ *for all* t*;*

(ii) for $\omega_0 > 0$*, the stationary solution for* $f_t(\theta_0)$ *does not have a finite first moment.*

The non-zero intercept ω_0 in the DGP is thus needed to obtain a non-degenerate limiting behavior of $f_t(\theta_0)$. The intuition for this is immediately clear from the recursion (10), which is a contracting autoregression of order one with a random coefficient. Following Bougerol (1993), Theorem 1 establishes that it has a stationary and ergodic solution. Filling out $f_t(\theta_0) = 0$, we can see that this obviously is a candidate solution in case $\omega_0 = 0$. The corollary then follows immediately from the uniqueness of the stationary and ergodic limit, as shown by Straumann and Mikosch (2006). By taking unconditional expectations of the left and middle part of (10) for $\omega_0 > 0$, and using the fact that the scaled score has conditional expectation zero, we have $\mathbb{E}[f_{t+1}(\theta_0)] = \omega_0 + \mathbb{E}[f_t(\theta_0)]$. It then also follows directly that $f_t(\theta_0)$ cannot have a finite mean if $\omega_0 > 0$.

The second part of Theorem 1 establishes the invertibility of the filter under a $\log^+$-log-moment condition, which is very weak. As the first part of the theorem already established that the data generated by the model is stationary and ergodic and that $\ln(1 + y_t)$ has a small moment r, the filter is invertible at the DGP. Invertibility, however, holds for generic stationary and ergodic y_t with a $\log^+$-log-moment, meaning that it continues to hold if the model is mis-specified. Also note that $\hat{f}_t(\theta)$ does not have a finite first unconditional moment at the DGP, as it is driven by the innovation $\alpha \ln(1 + y_t) = \alpha f_t(\theta_0)\epsilon_t$, where ϵ_t has a unit exponential distribution. The first part of the theorem already implied that $f_t(\theta_0)$ does not have a first moment.

Finally, the third part of Theorem 1 considers the scaled process $\hat{z}_t^f(\theta) = f_t(\theta_0)/\hat{f}_t(\theta)$. This process has finite moments and inverse moments of arbitrary large order, even though $f_t(\theta)$ and $\hat{f}_t(\theta)$ do not, neither at θ_0 nor at $\theta \in \Theta \backslash \theta_0$. The interest in the process $\hat{z}_t^f(\theta)$ stems from considering the centralized log-likelihood function under correct specification,

$$\begin{aligned}\hat{\mathcal{Q}}_T(\theta) &= \hat{L}_T(\theta) - L_T(\theta_0) = \sum_{t=1}^{T} \hat{Q}_t(\theta) \\ &= \sum_{t=1}^{T} -\ln \hat{f}_t(\theta) - (1 + \hat{f}_t(\theta)^{-1}) \ln(1 + y_t) + \ln f_t(\theta_0) + (1 + f_t(\theta_0)^{-1}) \ln(1 + y_t) \\ &= \sum_{t=1}^{T} \ln \frac{f_t(\theta_0)}{\hat{f}_t(\theta)} - \frac{f_t(\theta_0)}{\hat{f}_t(\theta)} \epsilon_t + \epsilon_t = \sum_{t=1}^{T} \ln \hat{z}_t^f(\theta) \; - \epsilon_t \left(\hat{z}_t^f(\theta) - 1 \right). \end{aligned} \tag{11}$$

We also define $\mathcal{Q}_T(\theta) = L_T(\theta) - L_T(\theta_0)$, and define $Q_t(\theta)$ similar to $\hat{Q}_t(\theta)$, but with $\hat{f}_t(\theta)$ replaced by its stationary and ergodic limit $f_t(\theta)$. As the maximizer of $\hat{L}_T(\theta)$ is the same as that of $\hat{\mathcal{Q}}_T(\theta)$, the properties of $\hat{z}_t^f(\theta)$ can be used to derive the properties of the MLE. In particular, (11) clarifies that consistency results can be obtained if a first moment exists for $\hat{z}_t^f(\theta)$. This is precisely what the last part of Theorem 1 establishes. Though $f_t(\theta_0)$ does not have a finite first moment for $\omega_0 > 0$, the *normalized* process $\hat{z}_t^f(\theta) = f_t(\theta_0)/\hat{f}_t(\theta)$ has finite moments up to arbitrary (positive) order.

We can now establish the following result for the maximum likelihood estimator of the model's static parameters using the integrated, score-driven filter.

Theorem 3. *Under Assumptions 1–2, the MLE is strongly consistent,* $\hat{\theta}_T \xrightarrow{a.s.} \theta_0$ *and asymptotically normally distributed with* $\sqrt{T}\left(\hat{\theta}_T - \theta_0\right) \to \mathcal{N}\left(\mathbf{0} \; , \; \boldsymbol{\mathcal{I}}(\theta_0)^{-1}\right)$ *for* $T \to \infty$*, where* $\boldsymbol{\mathcal{I}}(\theta_0) = -\mathbb{E}\left[\partial^2 Q_t(\theta_0)/\partial\theta\partial\theta^\top\right]$ *denotes the non-singular Fisher information matrix.*

Theorem 3 allows for an inferential framework for the key filtering parameter α if the model is correctly specified. Note that the correct specification immediately follows from the EVT perspective and the limiting result in (1) as long as the original data lie in the domain of attraction of a Fréchet law. This is much weaker than in usual settings, where the

assumption of correct specification might be deemed overly restrictive. Still, we can allow for some form of mis-specification due to for instance the use of finite thresholds $\hat{\tau}_t$ in the following way. The main arguments in the proof of Theorem 3 continue to hold as long as $F(y_t) = 1 - (1 + y_t)^{-1/f_t(\theta_0)} = 1 - \exp(-\epsilon_t)$ for some *IID* $\{\epsilon_t\}_{t\in\mathbb{Z}}$ that is not necessarily unit exponentially distributed as in the correctly specified case, but still has $\mathbb{E}[\epsilon_t] = 1$ and $\mathbb{E}[\epsilon_t^4] < \infty$. The result of Theorem 3 then needs to be slightly adapted by replacing the asymptotic covariance matrix $\boldsymbol{\mathcal{I}}(\theta_0)^{-1}$ by its usual sandwich form $\boldsymbol{\mathcal{I}}(\theta_0)^{-1}\boldsymbol{\mathcal{J}}(\theta_0)\boldsymbol{\mathcal{I}}(\theta_0)^{-1}$, where $\boldsymbol{\mathcal{J}}(\theta_0) = \mathbb{E}\left[(\partial Q_t(\theta_0)/\partial\theta)\ (\partial Q_t(\theta_0)/\partial\theta)^\top\right]$ denotes the expected outer product of gradients. The simulation Section 4 investigates even more severe forms of mis-specification and shows that the asymptotic normality approximation with the sandwich covariance matrix continues to give adequate results for inference in such settings.

4 Simulation study

4.1 Simulation design

This section investigates the performance of our dynamic EVT model in a controlled setting. We focus on the quality of the estimates of α and ω and the adequacy of the asymptotic normal approximation. The simulation study involves three settings (three sets of experiments). In all three settings, we generate draws x_t from a mixture distribution. With probability κ, we draw from a standard Gaussian GARCH(1,1) with parameters $(0.01, 0.07, 0.92)'$. The parameters are chosen close to the values found by estimating a Gaussian GARCH(1,1) model for the EUR/USD exchange rates used in the next section. With probability $1-\kappa$, we draw a tail observation as $x_t = \tau_t \exp(f_t\epsilon_t)$, where τ_t is the correct quantile from the underlying standard Gaussian GARCH(1,1) that is used for generating the observations from the center of the distribution, f_t is the tail shape, and ϵ_t is a standard unit exponential random variable.

In settings 1 and 2, we let the time-varying tail shape be generated by the model from

Section 2 using $\omega = 1.5 \cdot 10^{-5}$, and $\alpha = 0.01$ in line with parameters found for the BTC/USD and ETH/USD exchange rates in the empirical illustrations of Section 5. The difference between the first two settings is that we use the true thresholds τ_t in the first setting and the estimated thresholds $\hat{\tau}_t$ based on Section 2.4 in the second experiment. This allows us to study the effect of the estimation of the thresholds on the results.

In setting 3, we let the logarithm of the time-varying tail shape parameter be generated by a autoregressive model of order 1, with the following state equation:

$$\log(f_t) = -0.01 + 0.99 \cdot \log(f_{t-1}) + 0.06 \cdot \eta_{t-1},$$

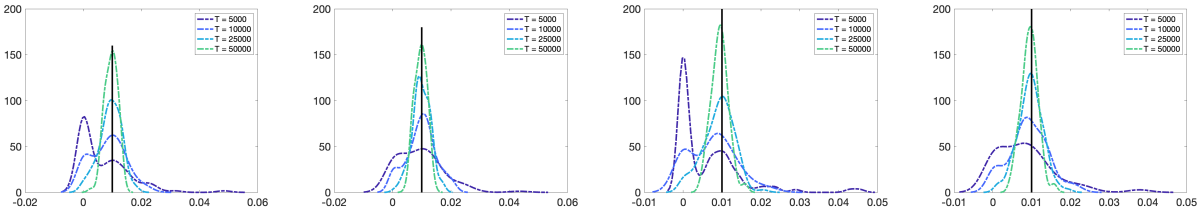
where $\{\eta_t\}_{t\in\mathbb{Z}}$ is an *IID* noise sequence where each η_t has a standard Gaussian distribution. The model is thus mis-specified, and we investigate whether the model can still reliably track the true time-varying parameter f_t and whether the estimated $\hat{\alpha}$ still behaves well compared to the correctly specified setting; see the discussion after Theorem 3. In this setting our filter is still invertible, but the estimator of the static parameters only converges to a pseudo-true value. The latter is chosen such that the mis-specified filtered model matches the unknown DGP as closely as possible (compare Blasques et al., 2015).

In all three settings, we consider the performance of the models estimated with and without the intercept parameter ω. This allows us to investigate the effect of including or excluding this parameter, which is typically estimated (very) close to zero. We consider four different sample sizes: $T \in \{5000, 10000, 25000, 50000\}$ and $1 - \kappa = 10\%$ observations coming from the tail. Sample sizes in our empirical illustrations are between about 6000 and 40000, matching the simulation setup. Together with the two different specifications (with and without ω), this yields a total of $3 \times 4 \times 2 = 24$ simulation experiments. Each experiment is repeated $S = 1,000$ times.

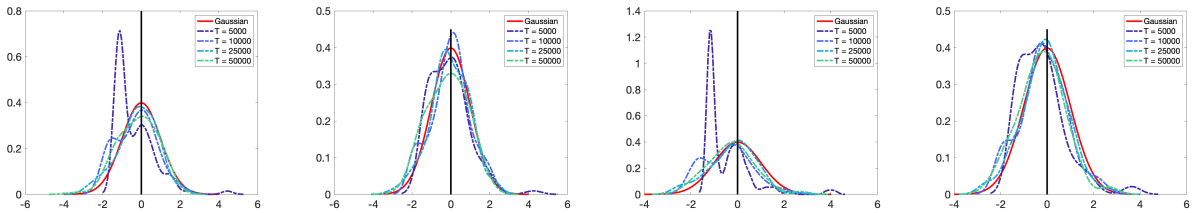
Figure 2: Simulation results for scenarios 1 and 2

Kernel density estimates of the distribution of the MLE for scenarios 1 and 2. In scenario 1, the true thresholds τ_t are used, denoted by τ_t in the subfigure heading. In scenario 2 the estimated thresholds are used, denoted as $\hat{\tau}_t$. We present results for estimated $\hat{\omega}$ as well as for $\hat{\omega}$ fixed at zero (denoted as $\hat{\omega} = 0$). The POTs have a GPD distribution with the correctly specified tail shape dynamics using the model from Section 2. Kernel density estimates are provided for $\hat{\alpha}$ and for its t-statistic using $S = 1,000$ simulations.

(a) $\hat{\alpha}$ ($\hat{\omega}$ free, τ_t) (b) $\hat{\alpha}$ ($\hat{\omega} = 0$, τ_t) (c) $\hat{\alpha}$ ($\hat{\omega}$ free, $\hat{\tau}_t$) (d) $\hat{\alpha}$ ($\hat{\omega} = 0$, $\hat{\tau}_t$)



(e) $t_{\hat{\alpha}}$ ($\hat{\omega}$ free, τ_t) (f) $t_{\hat{\alpha}}$ ($\hat{\omega} = 0$, τ_t) (g) $t_{\hat{\alpha}}$ ($\hat{\omega}$ free, $\hat{\tau}_t$) (h) $t_{\hat{\alpha}}$ ($\hat{\omega} = 0$, $\hat{\tau}_t$)



4.2 Simulation results

Figure 2 presents the results for the first two sets of experiments. We clearly see that the sample size matters for the results. If the sample size is too small ($1-\kappa = 10\%$ of $T = 5000$), the parameter $\hat{\alpha}$ is regularly estimated on the edge of the parameter space, i.e., at zero, if α and ω are estimated jointly. As the model has integrated dynamics, a non-zero ω combined with an α of zero results in a trending pattern for the tail shape parameter f_t. If the number of observations is too small, it is apparently difficult for the model to distinguish between a trending f_t and an integrated random f_t. The effect is obviously inherited by the t-statistic of $\hat{\alpha}$, as seen in the lower panel. As the sample size grows, the additional peak at 0 shrinks and the distribution of $\hat{\alpha}$ and of its t-statistic becomes more and more normal.

If ω is fixed at zero in the model (second column of figures compared to the first column),

the results appear very similar to the setting with estimated ω (first column of figures). For smaller samples, the estimator of α even appears to behave in a more stable way, collapsing to the edge of the parameter space less often compared to the setting with estimated ω. This suggests that, for practical purposes, ω may be set to zero (or an arbitrarily small positive number) during estimation without much of an effect on the estimated α. This is convenient, as it further simplifies the estimation problem to estimating a single parameter α, akin to the estimation of the single smoothing parameter in a RiskMetrics model for time-varying volatility.

Comparing the first versus the third column of graphs, or the second versus the fourth, we see that the effect of using estimated ($\hat{\tau}_t$) rather than true (τ_t) thresholds only has a mild effect on the distribution of $\hat{\alpha}$. For small sample sizes, the effect of $\hat{\alpha}$ collapsing to the edge of the parameter space are more severe when the thresholds are estimated. As in the case of true thresholds τ_t, however, these degenerate cases disappear quickly as the sample size increases. If $\hat{\omega}$ is fixed at zero rather than estimated, we again see that the behavior of $\hat{\alpha}$ is more stable in small samples, without seriously affecting its behavior in large samples. Again, fixing $\hat{\omega}$ to zero (or some small number) may be preferable from a stability point of view, particularly if the sample is not overly large.

Figure E.1 in the Appendix presents the results for scenario 3 where the dynamics for the tail shape are fully mis-specified in the model compared to the DGP. The results are consistent with the previous findings. For small sample sizes, the model has sometimes difficulty in finding a non-zero $\hat{\alpha}$ if $\hat{\alpha}$ is estimated jointly with $\hat{\omega}$, though less drastic than in Figure 2. For larger sample sizes, the problem disappears. The problem is less pronounced if $\hat{\omega}$ is fixed at zero rather than estimated, similar as in scenarios 1 and 2. The bottom panels in Figure E.1 in the appendix give the fit of the filter to the true time-varying tail shape parameter $f_t(\theta_0)$ and show that the model fits the true, unobserved process $f_t(\theta_0)$ quite well, despite mis-specification. Note that the filtered parameter $\hat{f}_t(\hat{\theta})$ differs in at least three ways from the true parameter dynamics: it uses the estimated $\hat{\theta}$, it is initialized, and

most importantly, it uses the incorrect score-driven dynamics for a DGP that is actually a pure autoregression for $\log(f_t(\theta_0))$. Despite this severe mis-specification, the filter still tracks the salient dynamics of the true $f_t(\theta_0)$. Results are similar for the other sample sizes.

The simulation experiments thus lead to two main suggestions. First, the asymptotic distributional results seem to hold up in finite samples and the filtered tail shape parameter tracks the true tail shape dynamics quite well, whether the model is correctly specified or not. Second, fixing ω to zero or some other small number during estimation may simplify the model and estimation problem even further without visibly affecting the distributional results for $\hat{\alpha}$ in large samples, and stabilizing them in small samples. We exploit these findings in the empirical analysis in the next section.

5 Empirical illustrations

5.1 EUR/USD and RUB/USD exchange rates

To illustrate our approach, we obtain end-of-day prices for the Euro (EUR) and the Russian Ruble (RUB) viz-a-viz the U.S. dollar (USD) as two easily available time series from Eikon (formerly Thomson/Reuters). The EUR/USD currency pair is extremely liquid, as both the euro area and the United States are comprised of large, developed, and market-oriented economies. By contrast, the RUB/USD currency pair is much less liquid, pointing to a relatively less-developed and less market-oriented economy. The RUB/USD currency pair became particularly illiquid following wars of aggression by Russia against Ukraine in 2014 and early 2022, and the subsequent economic sanctions. The EUR/USD time series spans the period January, 4 1999 to December, 1 2023, yielding 6,500 daily observations. The RUB/USD data range from January 2, 1996 to September 6, 2023, yielding 7,222 daily observations. We consider log-returns $x_t = 100 \times (\ln s_t - \ln s_{t-1})$, where s_t is the spot exchange rate in units of local currency, before applying our methodology. We focus on

Table 1: Parameter estimates

Parameter estimates for the dynamic EVT model. The first illustration considers EUR/USD and RUB/USD daily exchange rate log-returns. The second illustration considers BTC/USD and ETH/USD hourly cryptocurrency log-returns. Standard error estimates are in round brackets, and p-values are in square brackets.

	First illustration		Second illustration	
	EUR/USD	RUB/USD	BTC/USD	ETH/USD
α	0.034	0.039	0.008	0.011
	(0.016)	(0.020)	(0.003)	(0.004)
	[0.03]	[0.05]	[0.03]	[0.02]
T	6501	7223	41193	41258
T^*	638	720	2057	2058
loglik	-158.764	-657.267	-1377.416	-1190.192
AIC	319.527	1316.535	2758.831	2384.384

the right tail of the exchange rate returns. The right tail is the relevant tail for any firm that holds assets, or obtains revenues, in local currency and has financed these assets, at least in part, through USD liabilities. Large non-US financial institutions typically fit this description.

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 threshold process τ_t evolves according to (7) with $(1-\kappa) = 10\%$. It is initialized using the 90% empirical quantile of x_t over the first two years of data, and then evolves dynamically using the approach presented in Section 2.4, where we fix b^τ close to one to impose further smoothness. The process for the dynamic tail shape parameter f_t is initialized at $\hat{f}_1$, for which we use the maximum likelihood estimate of a static GPD tail shape model from an early subsample (1996–1997 for RUB/USD, January to March 1999 for EUR/USD). In line with the simulation outcomes from Section 4, we fix $\omega = 10^{-7}$ to a small positive number to stabilize the estimates of α that govern the magnitude of the tail shape dynamics. We estimate α by maximum likelihood. Table 1's first two columns present the results. Both for the EUR/USD and RUB/USD, the estimates of α are small and positive and comparable in size.

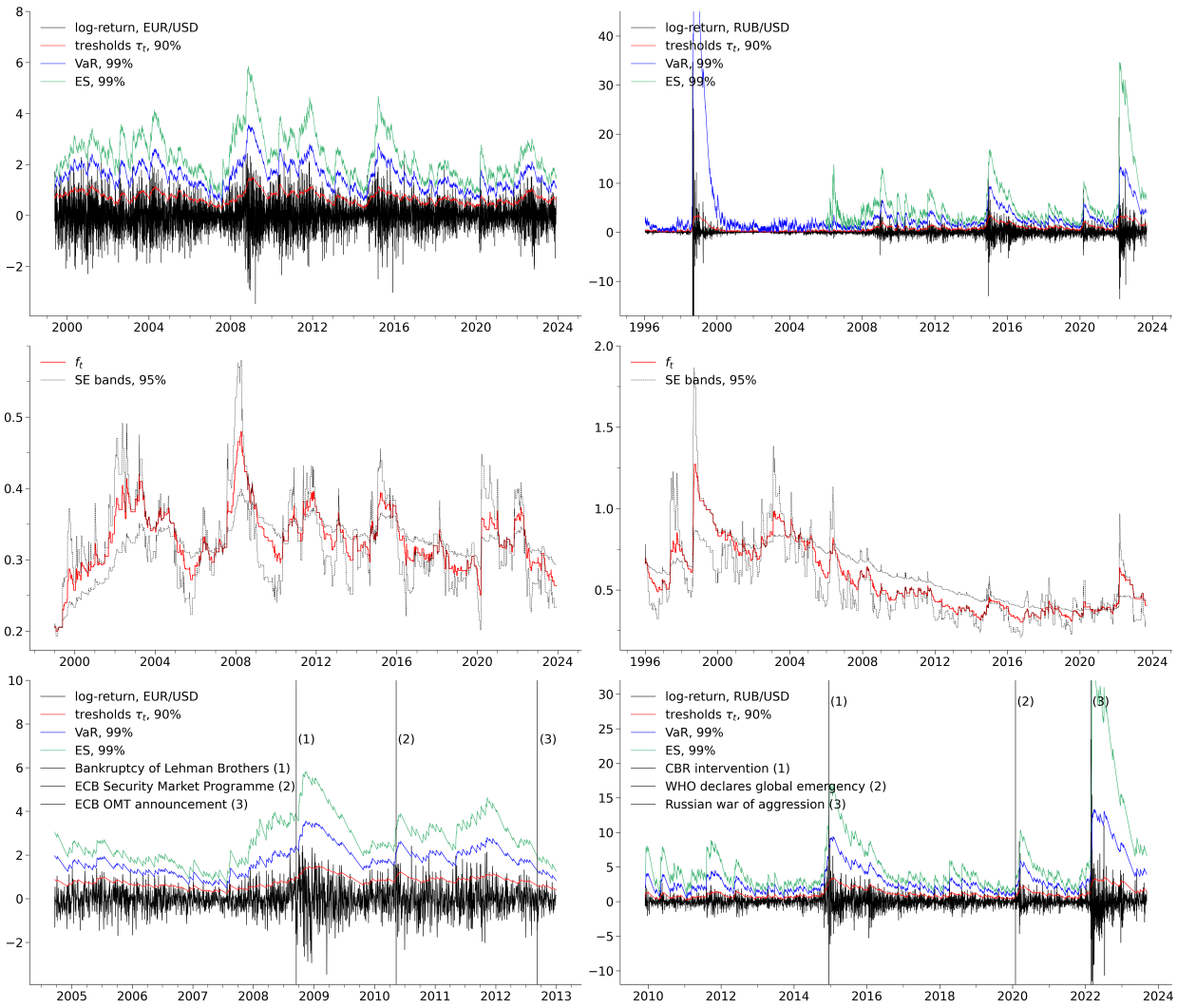
The top panels in Figure 3 present EUR/USD and RUB/USD log-returns. Thresholds τ_t are reported at the 90% confidence level, while VaR and ES are plotted at the more extreme 99% confidence level. If one wants to obtain even more extreme risk quantiles, this is easily done using (5)–(6) at the expense of obtaining a less well-scaled plot. The time variation in extreme market risks is visible from both top panels. The variation is particularly large for the RUB/USD rate, and especially around the Russian financial crisis and the collapse of LTCM in Fall 1998; see also the discussion further below. Also, note that we cannot plot the ES for the RUB/USD rate over the entire sample period given that f_t is estimated to be above one repeatedly in the first half of the sample. If the tail index $1/f_t$ falls below one, the first conditional moment of x_t (and therefore the ES) no longer exists.

The middle panels in Figure 3 present filtered estimates of f_t along with 95% confidence bands based on the methodology of Blasques et al. (2016) and are conditional on the estimated thresholds $\hat{\tau}_t$. The tail shape parameter varies between approximately 0.2 and 0.5 for EUR/USD and 0.2 and 1.3 for RUB/USD. Values of $f_t > 0.5$ imply that the second moment of the one-day-ahead predictive density does not exist. While second moments have always existed for the EUR/USD rate, they have rarely, if ever, existed for the RUB/EUR rate between 1996 and 2008. Maximal tail fatness for the Ruble is observed during the 1998 Russian financial crisis, sometimes also referred to as the Ruble crisis. The confidence bands around f_t suggest that it is reasonably precisely estimated and that it is far from zero (the thin-tailed Gumbel case). The confidence bands are asymmetric, owing to the parameter restriction $0 < \alpha < 1$, which we have imposed during estimation via re-parameterization.

The bottom left panel in Figure 3 provides zoomed-in estimates of extreme EUR/USD market risks during the global financial crisis between 2008 and 2009 and the euro area sovereign debt crisis between 2010 and 2012. We observe rapidly rising extreme foreign exchange rate risks following the bankruptcy of Lehman Brothers on September 15, 2008 (first vertical line); the euro area sovereign debt crisis, to which the European Central Bank (ECB) responded with sovereign bond purchases within its Securities Markets Programme

Figure 3: Exchange rate log-returns, tail shape, and extreme risks

Top panels: EUR/USD (left) and RUB/USD (right) daily log-returns (× 100). We focus on the adverse right tail, corresponding to a *depreciation* of the local currency viz-a-viz the U.S. dollar. Thresholds τ_t are reported at a 90% confidence level. VaR and ES are plotted at an extreme 99% confidence level. Middle panels: filtered tail shape parameter f_t with asymmetric 95% confidence band. Bottom panels: Zoomed-in extreme risks with key events.



(announced on May 10th, 2010; second vertical line); and preceding the ECB's announcement on September 6th, 2012 of its Outright Monetary Transactions program (third vertical line), following then-ECB president Mario Draghi's speech vowing to do "whatever it takes" to preserve the integrity of the euro area. 99% VaR and ES increased markedly following the first event, remained elevated around the second event, increased towards the third event,

Table 2: VaR violation rates

VaR coverage rates and p-value for the nominal coverage test of Kupiec (2000). Top and bottom panels consider VaR at non-extreme (90% and 95%) and more extreme (99%) confidence levels.

	First illustration		Second illustration	
	EUR/USD	RUB/USD	BTC/USD	ETH/USD
	$\hat{\tau}_t = \text{VaR}_t^{0.9}$		$\hat{\tau}_t = \text{VaR}_t^{0.95}$	
Violation rate	9.81%	9.96%	4.99%	4.99%
p-value	0.62	0.93	0.91	0.95
	$\text{VaR}_t^{0.99}$		$\text{VaR}_t^{0.99}$	
Violation rate	0.72%	0.97%	1.04%	1.02%
p-value	0.02	0.79	0.43	0.71

and gradually declined afterwards.

The bottom right panel in Figure 3 provides zoomed-in estimates of extreme market risks following Russia's invasions of Ukraine in 2014 and 2022 (the first and third vertical line, respectively). We observe heightened exchange rate volatility and unprecedented extreme tail risks following both events, both taking about two years to return down to their long-term levels. Extreme tail risks exert pressure on any firm with domestic assets and USD liabilities, and particularly so if debt ratios are high. The 99% VaR and ES in 2014 and 2022 are higher than in 2020 (following the outbreak of Covid-19, second vertical line), but lower than in 1998 (the Ruble crisis), though the latter took a shorter time to return to is long-term lower levels.

Columns 2 and 3 in Table 2 report EUR/USD and RUB/USD VaR violation percentages and p-values to test whether the violation rates coincide with the nominal rates. The VaR estimates are considered at a conventional (top panel, 90%) and a more extreme (bottom panel, 99%) confidence level. For 90% VaR, the VaR violation rates are close to 10%, and both tests are passed safely at conventional significance levels. For 99% VaR, the VaR violation rate is close to 1%, and both tests are passed at the 0.01 significance level. The statistical model adequately captures the relevant market risk dynamics in the extreme tail.

Web Appendix D presents our empirical results when Patton et al. (2019)'s joint model of ES and VaR is used to obtain the thresholds $\hat{\tau}_t$. The extreme market risk estimates are approximately similar. The thresholds inferred using (7) have slightly better VaR violation rates, possibly owing to its single-focused (VaR) objective function (8).

5.2 Cryptocurrency exchange rates

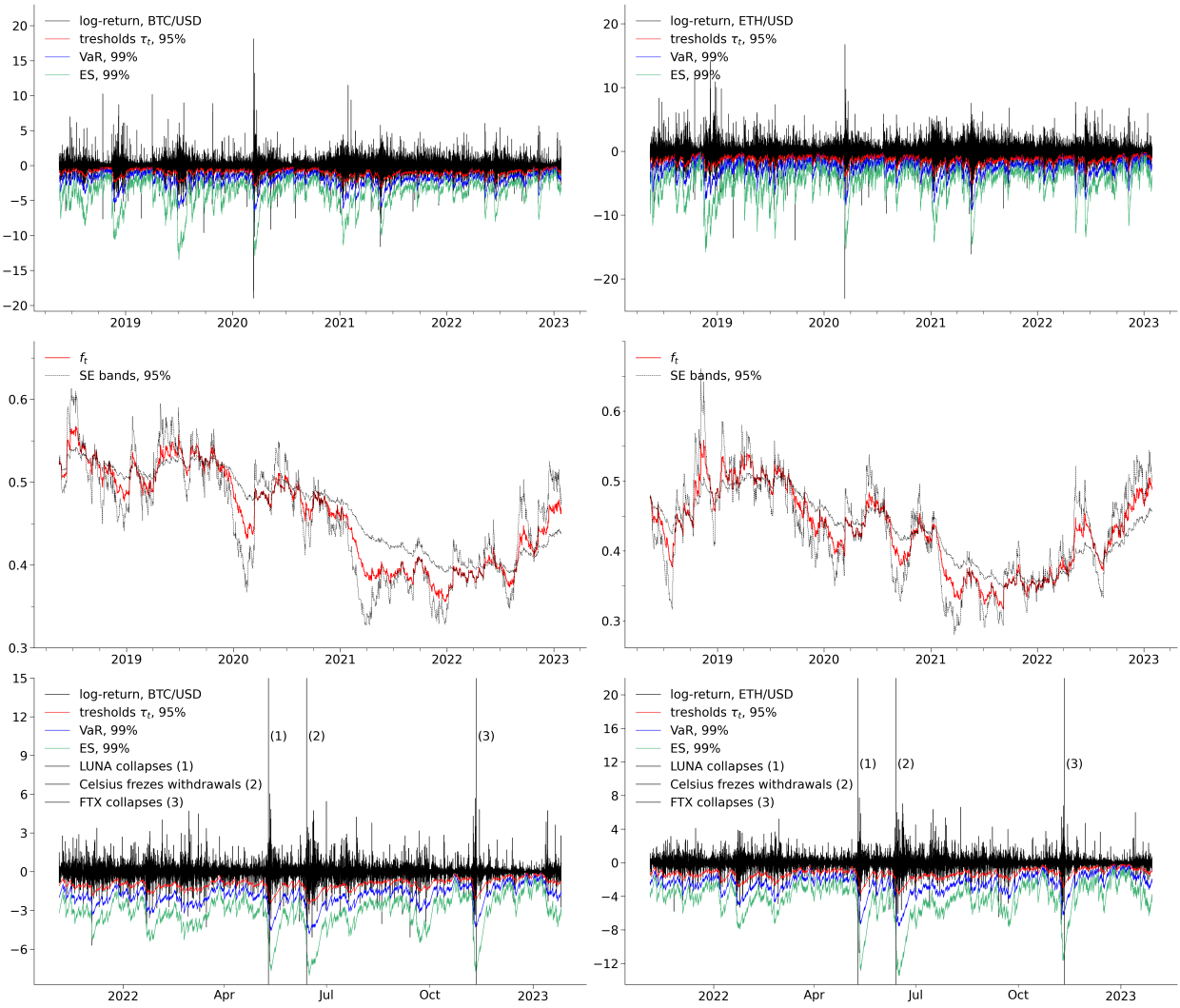
We obtain hourly prices for Bitcoin (BTC) and Ether (ETH) in USD from Coinbase. The BTC time series ranges from May, 15 2018 to January, 25 2023, yielding 41,192 observations. The ETH data range from May 15, 2018 to January 28, 2023, yielding 41,257 observations. To model the left tail of cryptocurrency returns, we consider *negative* log-returns $x_t = -100 \times (\ln s_t - \ln s_{t-1})$, where s_t is the price of either BTC or ETH in USD. The left tail is the economically relevant tail for an investor that holds cryptocurrency.

As for the EUR/USD and RUB/USD rates, we again rely on the time variation in the thresholds τ_t to accommodate time variation in any of the parameters describing the center of the distribution. Given the higher number of observations, we use $(1-\kappa) = 5\%$ to estimate the thresholds τ_t and initialize $\hat{\tau}_1$ and $\hat{f}_1$ using the first half year of data. Table 1 last two columns present our estimates of α. The estimate $\hat{\alpha}$ is positive, and statistically significantly different from zero at the 5% confidence level, for both BTC and ETH.

Figure 4 visualizes the tail risk outcomes. The top panels present BTC's and ETH's log-returns and the 95% thresholds τ_t. We also plot the 99% VaR and ES, but more extreme confidence percentages are easily computed via (5)–(6). There is clear time variation in extreme market risks. This is also underlined by the plots of f_t in the middle panels of Figure 4. The tail shape parameter varies between approximately 0.3 and 0.6 for both BTC and ETH, though their dynamics and the time points of their peaks and troughs are substantially different. This implies that the second moment of the one-hour-ahead predictive density exists most, but not all of the time. In particular, in the first part of the

Figure 4: Cryptocurrency log-returns, tail shape, and extreme risks

Top panels: Bitcoin/USD (left) and Ether/USD (right) hourly log-returns (× 100). Thresholds τ_t are reported at a 95% confidence level. VaR and ES are plotted at an extreme 99% confidence level. The thresholds τ_t, VaR, and ES are mirrored at the horizontal axis to correspond to log-returns (instead of log-losses). Middle panels: filtered tail shape parameter f_t with asymmetric 95% confidence band. Bottom panels: Zoomed-in extreme risks with key events.



sample, the tail risk of both cryptocurrencies seems to be very large. Also towards the end of the sample, f_t moves towards 0.5 where the second moment fails to exist. The confidence bands around f_t suggest that the tail shape parameter is precisely estimated. In particular, the tail parameter seems to be far from zero, the thin-tailed Gumbel case.

The bottom panels in Figure 4 provide zoomed-in estimates of extreme market risks

during the so-called "second crypto winter" of 2022, with vertical lines indicating three key events. VaR and ES at the 99% confidence level are particularly volatile during 2022. Both market risk measures respond strongly to the collapse of the Terra/Luna cryptocurrency on May 10th, 2022 (first vertical line), see e.g. Uhlig (2022); the collapse of FTX, a major cryptocurrency intermediary and shadow bank on June 13th, 2022 (second vertical line); and the collapse of Celsius, another cryptocurrency intermediary and shadow bank on Nov 11th, 2022 (third vertical line). ES 99% approximately tripled around each of these events, before reverting to more "normal" levels later on.

Columns 4 and 5 of Table 2 report BTC and ETH VaR violation rates and tests. The VaR estimates are considered at a non-extreme (top panel, 95%) and a more extreme (bottom panel, 99%) confidence level. For both 95% and 99% VaR, the violation rates are very close to 5% and 1% nominal levels, respectively, and the tests for nominal coverage are passed safely at conventional significance levels.

6 Conclusion

We introduced a robust semi-parametric modeling framework for studying persistent time variation in tail parameters for long univariate time series. To this end, we modeled the time variation in the shape parameter of the Generalized Pareto Distribution, which approximates the tail of most heavy-tailed densities found in economics and finance. By re-scaling the peaks-over-thresholds by their respective thresholds, we obtained a new single factor model to capture the time variation in extreme tails.

By endowing the time variation in the tail parameter with integrated score-driven dynamics, we obtained a simple filter for extreme tail risk that required only one or two static parameters to be estimated from the data. In this way, the paper complements standard well-known and widely used integrated filters for price volatility with a similar filter for the extreme tail risk. Given its reliance on Extreme Value Theory (EVT), the filter is less prone

to over-react to incidental large asset returns, thus augmenting less robust semi-parametric filters for Expected Shortfall such as the one of Patton et al. (2019), and insensitive to changes in the *center* as opposed to the *tails* of the distribution by only reacting to the Peaks-over-Threshold (PoT) observations; compare Massacci (2017).

As a theoretical contribution, we established parameter regions for stationarity, ergodicity, and invertibility of the filter process, and considered conditions for consistency and asymptotic normality of the maximum likelihood estimator of the model's static parameters. This complements the emerging literature on score-driven volatility filters with unit coefficient (integrated) dynamics and the existing literature on iGARCH filters with results for integrated dynamics for time-varying parameters describing higher-order properties of the distribution. The ease of the method's applicability and its implications were illustrated by studying the time variation in the tails of regular and cryptocurrency exchange rate returns over both quiet and turbulent times.

References

Andersen, T., T. Bollerslev, P. Christoffersen, and F. Diebold (2013). *Financial risk measurement for financial risk management*. Handbook of Economics of Finance. Elsevier.

Balkema, A. A. and L. de Haan (1974). Residual life time at great age. *The Annals of Probability 2*(5), 792–804.

Berkes, I., L. Horváth, and P. Kokoszka (2003). GARCH processes: structure and estimation. *Bernoulli 9*(2), 201 – 227.

Billingsley, P. (1961). The lindeberg-levy theorem for martingales. *Proceedings of the American Mathematical Society 12*(5), 788–792.

Blasques, F., S. J. Koopman, K. Łasak, 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*(3), 875–887.

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

Blasques, F., J. van Brummelen, P. Gorgi, and S. J. Koopman (2022). Maximum likelihood estimation for non-stationary location models with mixture of normal distributions. Technical report, Tinbergen Institute Discussion Paper.

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

Catania, L. and A. Luati (2023). Semiparametric modeling of multiple quantiles. *Journal of Econometrics 237*, 105365.

Chavez-Demoulin, V., A. C. Davison, and A. J. McNeil (2005). A point process approach to value-at-risk estimation. *Quantitative Finance 5*(2), 227–234.

Chavez-Demoulin, V. and P. Embrechts (2010). Revisiting the edge, ten years on. *Communications in StatisticsTheory and Methods 39*(8-9), 1674–1688.

Christoffersen, P. (2012). *Elements of Financial Risk Management, 2nd edition*. Academic Press.

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.

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 A. Ferreira (2006). *Extreme Value Theory: An Introduction*. Springer.

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

D'Innocenzo, E., A. Lucas, B. Schwaab, and X. Zhang (2024). Modeling extreme events: time-varying extreme tail shape. *Journal of Business and Economic Statistics 42*, 903–917.

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.

Francq, C. and Zakoian (2019). *GARCH Models, 2nd Edition*. John Wiley & Sons.

Francq, C. and J.-M. Zakoïan (2012). Strict stationarity testing and estimation of explosive and stationary generalized autoregressive conditional heteroscedasticity models. *Econometrica 80*(2), 821–861.

Harvey, A. C. (2013). *Dynamic models for Volatility and Heavy Tails*. Cambridge University Press.

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

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

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.

Krengel, U. (1985). *Ergodic theorems*. De Gruyter studies in Mathematics.

Kupiec, P. H. (2000). Stress tests and risk capital. *Journal of Risk 2*, 27–40.

Lee, S.-W. and B. E. Hansen (1994). Asymptotic theory for the garch (1, 1) quasi-maximum likelihood estimator. *Econometric theory 10*(1), 29–52.

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

Massacci, D. (2017). Tail risk dynamics in stock returns: Links to the macroeconomy and global markets connectedness. *Management Science 63*(9), 3072–3089.

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

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

Patton, A. 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.

Pickands, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics 3*(1), 119–131.

Rao, R. R. (1962). Relations between weak and uniform convergence of measures with applications. *The Annals of Mathematical Statistics 33*(2), 659–680.

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.

Uhlig, H. (2022). A Luna-tic stablecoin crash. Working Paper No. 2022-95, University of Chicago, Becker Friedman Institute for Economics.

White, H. (1994). *Estimation, Inference and Specification Analysis*. Cambridge University Press.

Web Appendix to

"Joint extreme Value-at-Risk and Expected Shortfall dynamics with a single integrated tail shape parameter"

Web Appendix A: Proofs

NIC for the EVT-based VaR

Let the conditional exceedance probability of τ_t be equal to γ_τ, and let $0 < \gamma < \gamma_\tau$ denote our extreme tail probability. Consider modeling the right-hand extreme tail. For a given values of $f_t = f$ and $\tau_t = \tau$, we have

$$f_{t+1} = \omega + f + \alpha \left(\ln(1 + y_t) - f \right),$$

$$\begin{aligned}
VaR_{t+1} &= \tau \cdot \left(\frac{\gamma}{\gamma_\tau} \right)^{-f_{t+1}} = \tau \cdot \left(\frac{\gamma}{\gamma_\tau} \right)^{-\omega - (1-\alpha) f - \alpha \ln(1+y_t)} \\
&= \tau \left(\frac{\gamma}{\gamma_\tau} \right)^{-\omega - (1-\alpha) f} \cdot \left(\frac{\gamma}{\gamma_\tau} \right)^{-\alpha \ln(1+y_t)} = C \cdot \left(\frac{\gamma}{\gamma_\tau} \right)^{-\alpha \ln(1+y_t)} \\
&= C \cdot \exp\left(\ln\left(\frac{\gamma}{\gamma_\tau} \right) \right)^{-\alpha \ln(1+y_t)} = C \cdot \exp\left(-\alpha \ln(1 + y_t) \ln\left(\frac{\gamma}{\gamma_\tau} \right) \right) \\
&= C \cdot \exp\left(\ln(1 + y_t) \right)^{-\alpha \ln\left(\gamma_\tau / \gamma \right)} = C \cdot (1 + y_t)^{-\alpha \ln\left(\gamma_\tau / \gamma \right)} \\
&= C \cdot \left(x_t / \tau \right)^{\alpha \ln\left(\gamma_\tau / \gamma \right)} = \tilde{C} \cdot x_t^{-\alpha \ln\left(\gamma_\tau / \gamma \right)},
\end{aligned}$$

for x_t exceeding the threshold τ, i.e., $x_t > \tau > 0$. The shape of the news impact curve for the VaR based on the EVT approach is thus concave as long as $\alpha \ln(\gamma_\tau / \gamma) < 1$. Note that for the plots in Section 2.3 we have re-cast our EVT approach to the extreme left-hand tail to make it directly comparable to the approach of Patton et al. (2019).

Preliminary results

Lemma A.1. Under Assumptions 1 and 2, the inequality

$$\mathbb{E}\left[\ln \left| 1 + \alpha \left(\epsilon - 1 \right) \right| \right] < 0,$$

is always satisfied.

Proof. For $b = \alpha/(1-\alpha) > 0$, we have

$$\begin{aligned}
\mathbb{E}\left[\ln|1+\alpha(\epsilon-1)|\right] &= \ln(1-\alpha) + \int_0^\infty \ln(1+bx)e^{-x}\mathrm{d}x \\
&= \ln(1-\alpha) - \left[\ln(1+bx)e^{-x}\right]_0^\infty + \int_0^\infty \frac{b}{1+bx}e^{-x}\mathrm{d}x \\
&= \ln(1-\alpha) + \int_0^\infty \frac{b}{1+bx}e^{-x}\mathrm{d}x \\
&= \ln(1-\alpha) + e^{1/b}\int_{1/b}^\infty \frac{e^{-x}}{x}\mathrm{d}x \\
&= \ln(1-\alpha) - e^{\alpha^{-1}-1}\ \mathrm{Ei}\left(1-\alpha^{-1}\right) \\
&< 0,
\end{aligned}$$

for $0 < \alpha < 1$, where $\mathrm{Ei}(z) = -\int_{-z}^\infty t^{-1}\,e^{-t}\,\mathrm{d}t$ denotes the exponential integral. ∎

Proof of Theorem 1

Since $\ln(1+y_t) = f_t(\theta_0)\epsilon_t$, we can write the score-driven filter as

$$\begin{aligned}
\hat{f}_{t+1}(\theta) =& \omega + \hat{f}_t(\theta) + \alpha\left(\ln(1+y_t) - \hat{f}_t(\theta)\right) \\
=& \omega + \hat{f}_t(\theta) + \alpha\left(f_t(\theta_0)\epsilon_t - \hat{f}_t(\theta)\right) \\
=& \omega + (1-\alpha)\,\hat{f}_t(\theta) + \alpha\epsilon_t f_t(\theta_0),
\end{aligned}$$

When evaluating the process above at the true parameter vector θ_0, we note that the unobserved process $\{f_{t+1}(\theta_0)\}_{t\in}$ satisfies

$$f_{t+1}(\theta_0) = \omega_0 + (1-\alpha_0+\alpha_0\epsilon_t)\,f_t(\theta_0).$$

Note that both $\hat{f}_t(\theta)$ and $f_t(\theta_0)$ are embedded in the stochastic recurrence equations (SREs) of the form $\hat{f}_{t+1}(\theta) = \hat{\phi}(\hat{f}_t(\theta), y_t, \theta)$ and $f_{t+1}(\theta_0) = \phi(f_t(\theta_0), \epsilon_t, \theta_0)$, respectively.

Part (i): To prove stationarity and ergodicty (SE) of $f_t(\theta_0)$, we apply Theorem 3.1 of Bougerol (1993). We first check the log-moment condition, which is easily satisfied since

$$\begin{aligned}\mathbb{E}\left[\ln^+\left|\phi(\bar{f}_1,\epsilon_t,\theta_0)\right|\right] &\leq \mathbb{E}\left[\ln\omega_0+\ln\left(1+\frac{(1-\alpha_0+\alpha_0\epsilon_t)}{\omega_0}\bar{f}_1\right)\right]\\ &= \ln\omega_0+\mathbb{E}\left[\frac{(1-\alpha_0+\alpha_0\epsilon_t)}{\omega_0}\bar{f}_1\right]=\ln\omega_0+\frac{1}{\omega_0}\bar{f}_1<\infty,\end{aligned}$$

for all $\bar{f}_1\in(0,\infty)$, where we have used the fact that ϵ_t is *IID* exponentially distributed with unit mean following Assumption 1. The contraction condition of Bougerol (1993) follows directly, as

$$\mathbb{E}\left[\sup_{\bar{f}}\ln\left|\frac{\partial\phi(\bar{f},\epsilon_t,\theta_0)}{\partial\bar{f}}\right|\right]=\mathbb{E}\left[\ln\left|1+\alpha_0(\epsilon_t-1)\right|\right]<0,$$

for $\alpha_0\in(0,1)$ using Lemma A.1 above. Hence, all the conditions of Theorem 3.1 of Bougerol (1993) are satisfied and we conclude that an SE solution $f_t(\theta_0)$ exists and that any initialized sequence converges exponentially fast almost surely (*e.a.s.*) to this unique SE limit. Given $y_t=\exp(f_t(\theta_0)\epsilon_t)-1$, it follows immediately that y_t is SE by Proposition 4.3 of Krengel (1985).

The existence of moments follows from Lemma 2.4 of Straumann and Mikosch (2006). The almost sure SE representation of $f_t(\theta_0)$ equals

$$f_t(\theta_0)=\omega_0\sum_{i=0}^{\infty}\prod_{j=0}^{i-1}\left(1+\alpha_0(\epsilon_{t-j}-1)\right)>0. \tag{A.1}$$

Note that $\mathbb{E}\left[(1+\alpha_0(\epsilon_t-1))^q\right]<\infty$ for any finite $q>0$ given that ϵ_t has a unit exponential distribution. Following to Lemma 2.4 of Straumann and Mikosch (2006), there exists an $0<\eta<1$ and a sufficiently small $0<r\leq q$ such that $\mathbb{E}[(1+\alpha_0(\epsilon_t-1))^r]=\eta$ and thus

$\mathbb{E}[\prod_{j=0}^{i-1}(1+\alpha_0(\epsilon_{t-j}-1))^r] = \eta^i$. Using this, we obtain

$$\mathbb{E}[f_t(\theta_0)^r] = \omega_0^r \sum_{i=0}^{\infty} \mathbb{E}\left[\prod_{j=0}^{i-1}\left(1+\alpha_0(\epsilon_{t-j}-1)\right)^r\right] = \omega_0^r \sum_{i=0}^{\infty} \eta^i < \infty.$$

As $\ln(1+y_t) = f_t(\theta_0)\epsilon_t$, this also directly establishes the existence of a log-moment for $\ln(1+y_t)$ and thus proves the first part of the theorem.

Part (ii): To prove that the filter $\hat{f}_t(\theta)$ is SE, we again apply Theorem 3.1 of Bougerol (1993). The existence of a log-moment is ensured because

$$\begin{aligned}\mathbb{E}\left[\ln^+ \sup_{\theta\in\Theta}\left|\hat{\phi}(\bar{f}_1, y_t, \theta)\right|\right] &\leq C + \ln^+ \sup_{\theta\in\Theta}\omega + \ln^+ \sup_{\theta\in\Theta}(1-\alpha) + \ln^+ \bar{f}_1 + \sup_{\theta\in\Theta}\alpha\mathbb{E}\left[\ln^+ \ln(1+y_t)\right] \\ &< \infty,\end{aligned}$$

for any $\bar{f}_1 \in (0,\infty)$, and where C is a finite constant. The last inequality follows from the assumed $\log^+$ moment for $\ln(1+y_t)$ and is automatically satisfied via part (i) of the theorem if the model is correctly specified.

To establish the contraction property, note that

$$\mathbb{E}\left[\sup_{\theta\in\Theta}\sup_{\bar{f}} \ln\left|\frac{\partial\hat{\phi}(\bar{f}, y_t, \theta)}{\partial\bar{f}}\right|\right] = \mathbb{E}\left[\sup_{\theta\in\Theta}\ln(1-\alpha)\right] = \sup_{\theta\in\Theta}\ln(1-\alpha) < 0,$$

as $0 < \underline{\alpha} \leq \alpha \leq \bar{\alpha} < 1$. We can now use Theorem 3.1 of Bougerol (1993) and conclude that $\hat{f}_t(\theta)$ is asymptotically SE, and converges *e.a.s.* to a unique SE limit $f_t(\theta)$, i.e., $\sup_{\theta\in\Theta}|\hat{f}_t(\theta) - f_t(\theta)| \overset{e.a.s.}{\to} 0$.

This establishes the second part of the theorem.

Part (iii): First note that $\hat{f}_t(\theta) \geq \underline{\omega}_f$, and thus

$$\sup_{\theta\in\Theta}\left|\frac{1}{\hat{f}_t(\theta)}-\frac{1}{f_t(\theta)}\right| \quad \leq \quad \underline{\omega}_f^2 \cdot \sup_{\theta\in\Theta}\left|\hat{f}_t(\theta)-f_t(\theta)\right| \xrightarrow{e.a.s.} 0.$$

It then follows directly from Lemma 2.1 of Straumann and Mikosch (2006) that $\hat{z}_t^f(\theta) \xrightarrow{e.a.s.} z_t^f(\theta)$ uniformly over $\theta \in \Theta$ if $\mathbb{E}[\ln^+ f_t(\theta_0)] < \infty$. The latter follows immediately from Part (i) above.

The boundedness of the moments follows along the same lines as Lemma A3 of Francq and Zakoïan (2012) by replacing their $a(\eta_t) = \beta_{FZ} + \alpha_{FZ}\eta_t^2$ for *IID* η_t with zero mean, unit variance, and $P(\eta_t^2 = 1) < 1$, by our $1 - \alpha + \alpha\epsilon_t$ for *IID* unit exponential ϵ_t, such that $0 < \beta_{FZ} = 1 - \alpha < 0$ and $\alpha_{FZ} = \alpha$, where β_{FZ} and α_{FZ} denote the parameters in the parameterization of Francq and Zakoïan (2012). Similarly, the boundedness of the inverse moment follows directly along the lines of Lemma 6 of Lee and Hansen (1994).

Proof of Theorem 3

Consistency: We show consistency by verifying the conditions in Theorem 3.4 of White (1994) with respect to the sequence $\{\hat{\mathcal{Q}}_T(\theta)\}_{T\in\mathbb{N}}$ as defined in equation (11). Specifically: (i) The parameter space Θ is compact; (ii) $\{\hat{\mathcal{Q}}_T(\theta)\}_{T\in\mathbb{N}}$ is a sequence of random functions continuous on Θ almost surely; (iii) $\hat{\mathcal{Q}}_T(\theta) = T^{-1}\sum_{t=1}^T \hat{Q}_t(\theta) \to \bar{\mathcal{Q}}(\theta) := \mathbb{E}\left[Q_t(\theta)\right]$ as $T \to \infty$ almost surely; and (iv) $\{\bar{\mathcal{Q}}(\theta) : \Theta \mapsto \mathbb{R}\}$ has an identifiably unique maximizer $\theta_0 \in \Theta$, that is, $\bar{\mathcal{Q}}(\theta_0) > \bar{\mathcal{Q}}(\theta)\ \forall\theta \neq \theta_0$.

Condition (i) holds by assumption, whereas (ii) trivially follows by continuity of $\{\hat{z}_t^f(\theta)\}_{t\in\mathbb{N}}$ and $\{z_t^f(\theta)\}_{t\in\mathbb{Z}}$. Furthermore, from Theorem 1 we obtain that $\mathbb{E}\left[\sup_{\theta\in\Theta}\left|\hat{Q}_t(\theta)\right|\right] < \infty$ and $\mathbb{E}\left[\sup_{\theta\in\Theta}\left|Q_t(\theta)\right|\right] < \infty$. Theorem 1 also ensures that the process $\{\hat{z}_t^f(\theta)\}_{t\in\mathbb{N}}$ converges e.a.s. to its stationary and ergodic limit $\{z_t^f(\theta)\}_{t\in\mathbb{Z}}$. We thus have

$$\sup_{\theta\in\Theta}\left|\hat{Q}_t(\theta)-Q_t(\theta)\right| \leq \sup_{\theta\in\Theta}\left|\ln\hat{z}_t^f(\theta)-\ln z_t^f(\theta)\right| - \epsilon_t \cdot \sup_{\theta\in\Theta}\left|\hat{z}_t^f(\theta)-z_t^f(\theta)\right|.$$

By the mean value theorem, there exist an intermediate point $\hat{f}_t^\star(\theta)$ between $\hat{f}_t(\theta)$ and $f_t(\theta)$ such that, using Lemma 2.1 of Straumann and Mikosch (2006), we obtain that

$$\begin{aligned}\sup_{\theta\in\Theta}\left|\ln\hat{z}_t^f(\theta)-\ln z_t^f(\theta)\right| &= \sup_{\theta\in\Theta}\left|\ln\hat{f}_t\theta)-\ln f_t(\theta)\right| = \sup_{\theta\in\Theta}\left|\frac{1}{f_t^\star(\theta)}\right|\sup_{\theta\in\Theta}\left|\hat{f}_t(\theta)-f_t(\theta)\right| \\ &\leq \frac{1}{\underline{\omega}_f}\sup_{\theta\in\Theta}\left|\hat{f}_t(\theta)-f_t(\theta)\right| \xrightarrow{e.a.s.} 0,,\end{aligned}$$

Since $\mathbb{E}[\epsilon_t]=1$ by Assumption 1, we can again apply Lemma 2.1 of Straumann and Mikosch (2006) to get

$$\epsilon_t\cdot\sup_{\theta\in\Theta}\left|\hat{z}_t^f(\theta)-z_t^f(\theta)\right| \xrightarrow{e.a.s.} 0.$$

It thus follows that

$$\sup_{\theta\in\Theta}\left|\hat{Q}_t(\theta)-Q_t(\theta)\right| \xrightarrow{e.a.s.} 0, \tag{A.2}$$

with $\mathbb{E}\left[\sup_{\theta\in\Theta}\left|Q_t(\theta)\right|\right]<\infty$. Let $\mathcal{Q}_T(\theta)=\sum_{t=1}^T \ln z_t^f(\theta)-\epsilon_t\left(z_t^f(\theta)-1\right)$ be the SE limit of $\hat{\mathcal{Q}}_T(\theta)$. Now, from the triangle inequality

$$\sup_{\theta\in\Theta}\left|\frac{1}{T}\hat{\mathcal{Q}}_T(\theta)-\bar{\mathcal{Q}}(\theta)\right| \leq \sup_{\theta\in\Theta}\left|\frac{1}{T}\hat{\mathcal{Q}}_T(\theta)-\mathcal{Q}_T(\theta)\right| + \sup_{\theta\in\Theta}\left|\frac{1}{T}\mathcal{Q}_T(\theta)-\bar{\mathcal{Q}}(\theta)\right|.$$

The first term on the RHS vanishes almost surely using Lemma 2.1 of Straumann and Mikosch (2006) and (A.2). For the second term, we can apply the ULLN for stationary and ergodic sequences of Rao (1962). As a result, we have

$$\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^T\hat{Q}_t(\theta)=\bar{\mathcal{Q}}(\theta)=1+\mathbb{E}\left[\ln z_t^f(\theta)-z_t^f(\theta)\right], \tag{A.3}$$

almost surely. For the last equality we have used the fact that ϵ_t is independent of $z_t^f(\theta)$ and $\mathbb{E}[\epsilon_t]=1$, as implied by Assumption 1.

Furthermore, $\bar{\mathcal{Q}}(\theta)\leq 0$ with equality if and only if $z_t^f(\theta)=1$ almost surely, because

$\log(z) - z + 1 \leq 0$ for any $z \in \mathbb{R}^+$, with equality only for $z = 1$. Note that $z_t^f(\theta_0) = 1$. This in turn implies that $\bar{\mathcal{Q}}(\theta_0) = 0$. We conclude the consistency proof by showing that if $z_t^f(\theta) = z_t^f(\theta_0) = 1$ almost surely for every t, then it must be that $\omega = \omega_0$ and $\alpha_0 = \alpha$. To show this, note that it suffices to show that the implication holds for $f_t(\theta) = f_t(\theta_0)$. Therefore, let $f_t(\theta) = f_t(\theta_0)$ almost surely for every t. We then have that

$$\begin{aligned} 0 &= f_{t+1}(\theta) - f_{t+1}(\theta_0) \\ &= (\omega - \omega_0) + \big(f_t(\theta) - f_t(\theta_0)\big) - \alpha f_t(\theta) + \alpha_0\, f_t(\theta_0) + (\alpha - \alpha_0)\, \epsilon_t f_t(\theta_0) \\ &= (\omega - \omega_0) + (\alpha - \alpha_0)\ (\epsilon_t - 1)\ f_t(\theta_0), \end{aligned}$$

almost surely. Obviously, from Assumption 1, ϵ_t is an $\mathcal{F}_t$-measurable random variable with a non-degenerate distribution, and from Theorem 1 $f_t(\theta_0)$ also has a non-degenerate distribution. As a result, the equality only holds almost surely if both $\alpha = \alpha_0$ and $\omega = \omega_0$.

The strong consistency of the MLE $\hat{\theta}_T$ in (3) is then guaranteed by noting that all the conditions of Theorem 3.4 in White (1994) are satisfied.

Asymptotic normality: Next, by strong consistency of the MLE $\hat{\theta}_T$, we obtain that, for large enough T the following Taylor expansion is allowed:

$$\nabla^\theta \hat{\mathcal{Q}}_T(\hat{\theta}_T) = \nabla^\theta \hat{\mathcal{Q}}_T(\theta_0) + \nabla^{\theta\theta} \hat{\mathcal{Q}}_T(\theta^\star) \left(\hat{\theta}_T - \theta_0\right), \tag{A.4}$$

where $\hat{\mathcal{Q}}_T(\theta) = \sum_{t=1}^T \hat{Q}_t(\theta)$ and $|\theta^\star - \theta_0| < \left|\hat{\theta}_T - \theta_0\right|$. It is easy to see that since the MLE $\hat{\theta}_T$ is the maximizer of $\hat{\mathcal{Q}}_T(\theta)$ and $\theta_0 \in int(\Theta)$ by Assumption 2, we have $\nabla^\theta \hat{\mathcal{Q}}_T(\hat{\theta}_T) = \mathbf{0}_2$, and hence we can rewrite (A.4) as

$$\frac{1}{T} \nabla^{\theta\theta} \hat{\mathcal{Q}}_T(\theta^\star) \left(\hat{\theta}_T - \theta_0\right) = -\frac{1}{T} \nabla^\theta \hat{\mathcal{Q}}_T(\theta_0). \tag{A.5}$$

To prove the asymptotic normality of the MLE $\hat{\theta}_T$ we verify the conditions given in Theorem 6.2 of White (1994). In particular, we let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space, and verify that: (i) The parameter space Θ is a compact subset of $\mathbb{R}^2$ with non-empty interior, (ii) the random function $\hat{\mathcal{Q}}_T(\theta) : \Omega \times \Theta \mapsto \mathbb{R}$ is continuously differentiable of order 2 on Θ almost surely, (iii) The MLE $\hat{\theta}_T : \Omega \mapsto \Theta$ is $\mathcal{F}$-measurable and strongly consistent, i.e. $\hat{\theta}_T \overset{a.s.}{\to} \theta_0$ where $\theta_0 \in int(\Theta)$; (iv) the score vector satisfies $T^{-1/2}\nabla^\theta \mathcal{Q}_T(\theta_0) \Rightarrow \mathcal{N}\left(\mathbf{0}_2, \mathbb{E}\left[\nabla^\theta Q_t(\theta_0)\nabla^\theta Q_t(\theta_0)^\top\right]\right)$; (v) the uniform stochastic convergence of the Hessian matrix, that is, $\sup_{\theta\in\Theta}\left\|\frac{1}{T}\nabla^{\theta\theta}\hat{\mathcal{Q}}_T(\theta) - \nabla^{\theta\theta}\bar{\mathcal{Q}}(\theta)\right\| \overset{a.s.}{\to} 0$, where $\nabla^{\theta\theta}\bar{\mathcal{Q}}(\theta) = \mathbb{E}\left[\nabla^{\theta\theta}Q_t(\theta)\right]$ is finite; (vi) the limit $\nabla^{\theta\theta}\bar{\mathcal{Q}}(\theta)$ evaluated at the true parameter vector θ_0 satisfies $-\bar{\mathcal{Q}}(\theta_0) = -\mathbb{E}\left[\nabla^{\theta\theta}Q_t(\theta_0)\right] = \boldsymbol{\mathcal{I}}(\theta_0)$, where $\boldsymbol{\mathcal{I}}(\theta_0)$ is the Fisher's information matrix.

Obviously, (i)–(iii) are directly implied by Assumptions 1 and 2.

For (iv) it suffices to prove that $\{\nabla^\theta Q_t(\theta_0)\}_{t\in\mathbb{Z}}$ is a stationary and ergodic zero-mean martingale difference process with respect to the filtration $\{\mathcal{F}_t\}_{t\in\mathbb{N}}$ with $\mathcal{F}_t = \sigma\{\epsilon_t, \epsilon_{t-1}, \epsilon_{t-2}, \dots\}$. In fact, note that

$$\mathbb{E}\left[\nabla^\theta Q_t(\theta_0) \;\middle|\; \mathcal{F}_{t-1}\right] = \nabla^\theta z_t^f(\theta_0)\left(\frac{1}{z_t^f(\theta_0)} - \mathbb{E}\left[\epsilon_t|\mathcal{F}_{t-1}\right]\right) = \mathbf{0}_2, \tag{A.6}$$

which clearly follows from Assumption 1, the fact that $\nabla^\theta z_t^f(\theta)$ are $\mathcal{F}_{t-1}$-measurable, and that $z_t^f(\theta_0) = 1$ for all t.

Moreover, we can also prove that $\nabla^\theta Q_t(\theta_0)$ is square-integrable since we clearly have $z_t^f(\theta_0) = 1$, and therefore $\nabla^\theta z_t^f(\theta_0) = -\nabla^\theta(1/z_t^f(\theta_0))$ and

$$\begin{aligned}
\mathbb{E}\left[\nabla^\theta Q_t(\theta_0)\nabla^\theta Q_t(\theta_0)^\top\right] &= \mathbb{E}\left[\nabla^\theta z_t^f(\theta_0)\nabla^\theta z_t^f(\theta_0)^\top\left(\frac{1}{z_t^f(\theta_0)} - \epsilon_t\right)^2\right] \\
&= \mathbb{E}\left[\nabla^\theta z_t^f(\theta_0)\nabla^\theta z_t^f(\theta_0)^\top\right] \\
&= \mathbb{E}\left[\nabla^\theta\left(\frac{1}{z_t^f(\theta_0)}\right)\nabla^\theta\left(\frac{1}{z_t^f(\theta_0)}\right)^\top\right] < \infty,
\end{aligned}$$

as implied by Assumption 1 together with Lemma B.2. Therefore, we are allowed to apply the CLT for square-integrable martingales of Billingsley (1961) in order to obtain

$$T^{-1/2}\ \nabla^{\theta}\mathcal{Q}_T(\theta_0) \Rightarrow \mathcal{N}\left(\mathbf{0}_2, \mathbb{E}\left[\nabla^{\theta}Q_t(\theta_0)\nabla^{\theta}Q_t(\theta_0)^{\top}\right]\right).$$

Next, we focus on (v) and prove the uniform stochastic convergence of the Hessian matrix. From the triangle inequality

$$\begin{aligned}\sup_{\theta\in\Theta}\left\|\frac{1}{T}\nabla^{\theta\theta}\hat{\mathcal{Q}}_T(\theta) - \nabla^{\theta\theta}\bar{\mathcal{Q}}(\theta)\right\| \leq & \sup_{\theta\in\Theta}\left\|\frac{1}{T}\nabla^{\theta\theta}\hat{\mathcal{Q}}_T(\theta) - \frac{1}{T}\nabla^{\theta\theta}\mathcal{Q}_t(\theta)\right\| \\ & + \sup_{\theta\in\Theta}\left\|\frac{1}{T}\nabla^{\theta\theta}\mathcal{Q}_T(\theta) - \nabla^{\theta\theta}\bar{\mathcal{Q}}(\theta)\right\|, \end{aligned} \tag{A.7}$$

where $\{\nabla^{\theta\theta}Q_t(\theta)\}_{t\in\mathbb{Z}}$ is stationary and ergodic and $\nabla^{\theta\theta}\bar{\mathcal{Q}}(\theta) = \mathbb{E}\left[\nabla^{\theta\theta}Q_t(\theta)\right]$ where, by Lemma B.3, $\mathbb{E}\left[\sup_{\theta\in\Theta}\left\|\nabla^{\theta\theta}Q_t(\theta)\right\|\right]$ exists.

Hence, from the ULLN of Rao (1962) for stationary and ergodic sequences,

$$\sup_{\theta\in\Theta}\left\|\frac{1}{T}\nabla^{\theta\theta}\mathcal{Q}_T(\theta) - \nabla^{\theta\theta}\bar{\mathcal{Q}}(\theta)\right\| \xrightarrow{a.s.} 0.$$

Now, from Theorem 1 and Lemma B.1 together with continuity arguments, we obtain

$$\sup_{\theta\in\Theta}\left\|\nabla^{\theta\theta}\hat{Q}_t(\theta) - \nabla^{\theta\theta}Q_t(\theta)\right\| \xrightarrow{e.a.s.} 0.$$

Combining these results, we conclude that (A.7) vanishes almost surely, that is

$$\sup_{\theta\in\Theta}\left\|\frac{1}{T}\nabla^{\theta\theta}\hat{\mathcal{Q}}_T(\theta) - \nabla^{\theta\theta}\bar{\mathcal{Q}}(\theta)\right\| \xrightarrow{a.s.} 0.$$

Moreover, by the strong consistency of the MLE, and the fact that $\theta \mapsto \nabla^{\theta\theta}\bar{\mathcal{Q}}(\theta)$ is continuous, to complete the proof, we only need to verify (vi) and show that $\nabla^{\theta\theta}\bar{\mathcal{Q}}(\theta_0)$ is

non-singular. By applying the law of iterated expectations and Assumption 1, we get that

$$\begin{aligned}\mathbb{E}\left[\nabla^{\theta\theta}Q_t(\theta_0)\right] &= \mathbb{E}\left[\nabla^{\theta\theta}z_t^f(\theta_0)\left(\frac{1}{z_t^f(\theta_0)}-\epsilon_t\right)-\nabla^{\theta}z_t^f(\theta_0)\nabla^{\theta}z_t^f(\theta_0)^\top\frac{1}{z_t^f(\theta_0)^2}\right]\\ &= \mathbb{E}\left[\mathbb{E}\left[\nabla^{\theta\theta}z_t^f(\theta_0)\left(1-\epsilon_t\right)-\nabla^{\theta}z_t^f(\theta_0)\nabla^{\theta}z_t^f(\theta_0)^\top\;\middle|\;\mathcal{F}_{t-1}\right]\right]\\ &= -\mathbb{E}\left[\nabla^{\theta}z_t^f(\theta_0)\nabla^{\theta}z_t^f(\theta_0)^\top\right],\end{aligned}$$

since $z_t^f(\theta_0)$, $\nabla^{\theta}z_t^f(\theta_0)$ and $\nabla^{\theta\theta}z_t^f(\theta_0)$ are $\mathcal{F}_{t-1}$-measurable, $z_t^f(\theta_0)=1$ for all t, and $\mathbb{E}\left[\epsilon_t\right]=1$.

Note that the process $\{\nabla^{\theta}z_t^f(\theta_0)\}_{t\in\mathbb{Z}}$ can be written as

$$\begin{aligned}\nabla^{\theta}z_{t+1}^f(\theta) &= \nabla^{\theta}\left(\frac{f_{t+1}(\theta_0)}{f_{t+1}(\theta)}\right)=-\frac{f_{t+1}(\theta_0)}{f_{t+1}(\theta)^2}\nabla^{\theta}f_{t+1}(\theta)\\ &= -\frac{f_{t+1}(\theta_0)}{f_{t+1}(\theta)^2}\left(\nabla^{\theta}\omega+\left(\nabla^{\theta}\alpha\right)\left(f_t(\theta_0)\epsilon_t-f_t(\theta)\right)+(1-\alpha)\nabla^{\theta}f_t(\theta)\right)\\ &= \left(\frac{f_{t+1}(\theta_0)}{f_{t+1}(\theta)}\right)^2\left(\frac{f_t(\theta_0)}{f_{t+1}(\theta_0)}\right)\left(w_t(\theta)+(1-\alpha)\nabla^{\theta}z_t^f(\theta)\right),\end{aligned}\tag{A.8}$$

$$w_t(\theta)=\left(-1/f_t(\theta_0)\quad,\quad(1/z_t^f(\theta))-\epsilon_t\right)^\top,$$

where $f_t(\theta_0)/f_{t+1}(\theta_0)=1/(1+\alpha_0(\epsilon_t-1)+\omega_0/f_t(\theta_0))$. Since $\{\nabla^{\theta}z_t(\theta_0)\}_{t\in\mathbb{Z}}$ are stationary and ergodic, if $\nabla^{\theta\theta}\bar{\mathcal{Q}}(\theta)$ were singular, then $\exists\boldsymbol{\lambda}\in\mathbb{R}^2\backslash\{\mathbf{0}_2\}$ such that $\boldsymbol{\lambda}^\top\nabla^{\theta}z_t^f(\theta_0)=\mathbf{0}_2$ almost surely $\forall t\in\mathbb{Z}$. This is obviously ruled out by the functional form of (A.8) and the unit exponential distributional form of ϵ_t and, therefore, it must be that $\boldsymbol{\lambda}^\top\nabla^{\theta}z_t^f(\theta_0)=\mathbf{0}_2\iff\boldsymbol{\lambda}=\mathbf{0}_2$ and thus, $\nabla^{\theta\theta}\bar{\mathcal{Q}}(\theta_0)$ is non-singular. In conclusion, we note that the Fisher's information equality $\mathbb{E}\left[\nabla^{\theta}Q_t(\theta_0)\nabla^{\theta}Q_t(\theta_0)^\top\right]=-\mathbb{E}\left[\nabla^{\theta\theta}Q_t(\theta_0)\right]=\boldsymbol{\mathcal{I}}(\theta_0)$ follows by standard arguments.

Web Appendix B: Technical lemmas

We define the operators $\nabla^{\theta} = \frac{\partial}{\partial \theta}$ and $\nabla^{\theta\theta} = \frac{\partial^2}{\partial \theta \partial \theta^\top}$. In addition, we denote the score vector by $\nabla^{\theta} Q_t(\theta) = \left(\nabla^{\omega} Q_t(\theta), \nabla^{\alpha} Q_t(\theta)\right)^\top \in \mathbb{R}^2$, and the Hessian matrix

$$\nabla^{\theta\theta} Q_t(\theta) = \begin{pmatrix} \nabla^{\omega\omega} Q_t(\theta) & \nabla^{\omega\alpha} Q_t(\theta) \\ \nabla^{\omega\alpha} Q_t(\theta) & \nabla^{\alpha\alpha} Q_t(\theta) \end{pmatrix} \in \mathbb{R}^{2\times 2}.$$

It is important to note that differentiating the log-likelihood difference $Q_t(\theta)$ is equivalent to differentiating $\hat{\ell}_t(\theta)$ as defined in (3) since $\ell_t(\theta_0)$ does not depend on θ. Define $z_t^{f,\text{inv}}(\theta) = 1/z_t^f(\theta)$. The elements of the score vector are given by

$$\nabla^{\theta} Q_t(\theta) = \left(z_t^f(\theta)^{-1} - \epsilon_t\right) \nabla^{\theta} z_t^f(\theta) = \left(\epsilon_t - z_t^{f,\text{inv}}(\theta)\right) \frac{\nabla^{\theta} z_t^{f,\text{inv}}(\theta)}{z_t^{f,\text{inv}}(\theta)^2}, \tag{B.1}$$

and

$$\begin{aligned} \nabla^{\theta\theta} Q_t(\theta) &= \left(z_t^f(\theta)^{-1} - \epsilon_t\right) \nabla^{\theta\theta} z_t^f(\theta) - \frac{\nabla^{\theta} z_t^f(\theta) \nabla^{\theta} z_t^f(\theta)^\top}{z_t^f(\theta)^2} \\ &= \left(\epsilon_t - z_t^{f,\text{inv}}(\theta)\right) \left(\frac{\nabla^{\theta\theta} z_t^{f,\text{inv}}(\theta)}{z_t^{f,\text{inv}}(\theta)^2} - \frac{\nabla^{\theta} z_t^{f,\text{inv}}(\theta) \nabla^{\theta} z_t^{f,\text{inv}}(\theta)^\top}{z_t^{f,\text{inv}}(\theta)^3} \right) \\ &\quad - \frac{\nabla^{\theta} z_t^{f,\text{inv}}(\theta) \nabla^{\theta} z_t^{f,\text{inv}}(\theta)^\top}{z_t^{f,\text{inv}}(\theta)^2}, \end{aligned} \tag{B.2}$$

where the first derivative processes $\nabla^{\theta} z_t^{f,\text{inv}}(\theta) = -z_t^f(\theta)^2 \nabla^{\theta} z_t^f(\theta)$ are defined as $\nabla^{\theta} z_t^{f,\text{inv}}(\theta) = \left(\nabla^{\omega} z_t^{f,\text{inv}}(\theta), \nabla^{\alpha} z_t^{f,\text{inv}}(\theta)\right)^\top \in \mathbb{R}^2$, where

$$\nabla^{\theta} z_{t+1}^{f,\text{inv}}(\theta) = \begin{pmatrix} \nabla^{\omega} z_{t+1}^{f,\text{inv}}(\theta) \\ \nabla^{\alpha} z_{t+1}^{f,\text{inv}}(\theta) \end{pmatrix} = \begin{pmatrix} \dfrac{\nabla^{\omega} f_{t+1}(\theta)}{f_{t+1}(\theta_0)} \\ \dfrac{\nabla^{\alpha} f_{t+1}(\theta)}{f_{t+1}(\theta_0)} \end{pmatrix} = \frac{\nabla^{\theta} f_{t+1}(\theta)}{f_{t+1}(\theta_0)}, \tag{B.3}$$

and

$$\nabla^{\theta} f_{t+1}(\theta) = \phi_t^{\theta}\left(f_t(\theta), \nabla^{\theta} f_t(\theta), \epsilon_t, \theta\right) = \begin{pmatrix} 1 \\ f_t(\theta_0)\epsilon_t - f_t(\theta) \end{pmatrix} + (1-\alpha)\,\nabla^{\theta} f_t(\theta). \tag{B.4}$$

For the second derivative processes $\nabla^{\theta\theta} z_t^{f,\text{inv}}(\theta)$, we have

$$\nabla^{\theta\theta} z_{t+1}^{f,\text{inv}}(\theta) = \frac{\nabla^{\theta\theta} f_{t+1}(\theta)}{f_{t+1}(\theta_0)} = f_{t+1}(\theta_0)^{-1} \begin{pmatrix} \nabla^{\omega\omega} f_{t+1}(\theta) & \nabla^{\omega\alpha} f_{t+1}(\theta) \\ \nabla^{\omega\alpha} f_{t+1}(\theta) & \nabla^{\alpha\alpha} f_{t+1}(\theta) \end{pmatrix}, \tag{B.5}$$

$$\begin{aligned} \nabla^{\theta\theta} f_{t+1}(\theta) &= \phi_t^{\theta\theta}\left(f_t(\theta), \nabla^{\theta} f_t(\theta), \nabla^{\theta\theta} f_t(\theta), \epsilon_t, \theta\right) \\ &= \nabla^{\theta\top}\left(\begin{pmatrix} 1 \\ f_t(\theta_0)\epsilon_t - f_t(\theta) \end{pmatrix} + (1-\alpha)\,\nabla^{\theta} f_t(\theta)\right) \\ &= -\begin{pmatrix} 0 & \nabla^{\omega} f_t(\theta) \\ \nabla^{\omega} f_t(\theta) & 2\nabla^{\alpha} f_t(\theta) \end{pmatrix} + (1-\alpha)\,\nabla^{\theta\theta} f_t(\theta). \end{aligned} \tag{B.6}$$

Similar derivations hold for the initialized counterparts $\hat{z}_{t+1}^{f,\text{inv}}(\theta) = \hat{f}_{t+1}/f_{t+1}(\theta_0)$.

The following Lemma shows that the derivative processes $\nabla^{\theta} z_t^{f,\text{inv}}(\theta)$ and $\nabla^{\theta\theta} z_t^{f,\text{inv}}(\theta)$ of the ratio process $\{z_t^{f,\text{inv}}(\theta)\}_{t\in\mathbb{Z}}$ are also asymptotically stationary and ergodic with bounded log-moments.

Lemma B.1. Under the conditions of Theorem 1,

$$\begin{aligned} \sup_{\theta\in\Theta} \left\| \nabla^{\theta} \hat{z}_t^{f,\text{inv}}(\theta) - \nabla^{\theta} z_t^{f,\text{inv}}(\theta) \right\| &\xrightarrow{e.a.s.} 0, \\ \sup_{\theta\in\Theta} \left\| \nabla^{\theta\theta} \hat{z}_t^{f,\text{inv}}(\theta) - \nabla^{\theta\theta} z_t^{f,\text{inv}}(\theta) \right\| &\xrightarrow{e.a.s.} 0, \end{aligned}$$

for stationary and ergodic derivative processes $\nabla^{\theta} z_t^{f,\text{inv}}(\theta)$ and $\nabla^{\theta\theta} z_t^{f,\text{inv}}(\theta)$.

Proof of Lemma B.1

We note that $\nabla^\theta \hat{f}_{t+1}(\theta)$ is a function of both the filter $\hat{f}_t(\theta)$ and its derivative $\nabla^\theta \hat{f}_t(\theta)$. To establish the stationarity and ergodicity, we verify the conditions given in Theorem 2.10 of Straumann and Mikosch (2006) for perturbed stochastic recurrence equations (SREs).

It is immediate to see that the conditions S.1 and S.2 stated in Theorem 2.10 of Straumann and Mikosch (2006) are the same as the log-moment and the contraction condition in Theorem 3.1 of Bougerol (1993), and these are clearly implied by Theorem 1, since the mapping function $\phi_t^\theta\left(\hat{f}_t(\theta), \nabla^\theta \hat{f}_t(\theta), \epsilon_t, \theta\right)$ has finite log-moment and the contraction condition is satisfied because $0 < \underline{\alpha} < \alpha < \bar{\alpha} < 1$. We then only have to check condition S.3 of Straumann and Mikosch (2006), that ensures that the perturbed and unperturbed SRE converge sufficiently fast for the difference between their asymptotic solutions to vanish exponentially fast.

The condition follows by showing that

$$\sup_{\theta\in\Theta}\left\|\phi_t^\theta\left(\hat{f}_t(\theta), \nabla^\theta \bar{f}_1(\theta), \epsilon_t, \theta\right) - \phi_t^\theta\left(f_t(\theta), \nabla^\theta \bar{f}_1(\theta), \epsilon_t, \theta\right)\right\| \xrightarrow{e.a.s.} 0,$$

where $\nabla^\theta \bar{f}_1(\theta)$ is some fixed starting point for the derivative recursion. It is straightforward to see that the norm is given by

$$\sup_{\theta\in\Theta}\left\|\begin{pmatrix} 0 \\ \hat{f}_t(\theta) - f_t(\theta) \end{pmatrix}\right\| \xrightarrow{e.a.s.} 0.$$

As $\nabla^\theta \hat{z}_t^{f,\text{inv}}(\theta) = \nabla^\theta \hat{f}_t(\theta)/f_t(\theta_0)$, the first result now follows immediately.

The second result follows along the same lines, but now using the SRE defined by $\phi_t^{\theta\theta}(f_t(\theta), \nabla^\theta f_t(\theta), \nabla^{\theta\theta} f_t(\theta))$ in (B.6) and using the e.a.s. convergence of both $\hat{f}_t(\theta)$ and $\nabla^\theta \hat{f}_t(\theta)$ to their SE limits. ∎

Next we introduce another lemma that provides a suitable number of bounded moments

for the derivatives of the ratio process $\{z_t^{f,\text{inv}}(\theta)\}_{t\in\mathbb{Z}}$, i.e., $\{\nabla^\theta z_t^{f,\text{inv}}(\theta)\}_{t\in\mathbb{Z}}$ and $\{\nabla^{\theta\theta} z_t^{f,\text{inv}}(\theta)\}_{t\in\mathbb{Z}}$. As it is clear from equations (B.1) and (B.2), this is a necessary step to ensure that the score vector of the log-likelihood is a martingale difference sequence with bounded and constant variance-covariance matrix and, further, that the empirical mean of the negative Hessian matrix converges almost surely to a positive-definite constant matrix.

Lemma B.2. Under the conditions of Theorem 1, the derivatives processes $\{\nabla^\theta z_t^{f,\text{inv}}(\theta)\}_{t\in\mathbb{Z}}$ and $\{\nabla^{\theta\theta} z_t^{f,\text{inv}}(\theta)\}_{t\in\mathbb{Z}}$ have n uniformly bounded moments $\forall n > 0$, that is

$$\mathbb{E}\left[\sup_{\theta\in\Theta}\left\|\nabla^\theta z_t^{f,\text{inv}}(\theta)\right\|^n\right] < \infty, \qquad \mathbb{E}\left[\sup_{\theta\in\Theta}\left\|\nabla^{\theta\theta} z_t^{f,\text{inv}}(\theta)\right\|^n\right] < \infty.$$

Proof of Lemma B.2

Consider the SRE (B.3), then we have $\left\|\nabla^\theta z_t^{f,\text{inv}}(\theta)\right\| = \left\|\nabla^\theta f_t(\theta)/f_t(\theta_0)\right\|$, and

$$\left\|\frac{\nabla^\theta f_{t+1}(\theta)}{f_t(\theta_0)}\right\| \leq (1-\alpha)^t \left\|\frac{\nabla^\theta \bar{f}_1(\theta)}{\omega_0}\right\| + \sum_{i=0}^{t}(1-\alpha)^i \left\|\begin{pmatrix} \frac{1}{\omega_0} \\ \frac{\epsilon_{t-i} f_{t-i}(\theta_0)}{f_t(\theta_0)} - \frac{f_{t-i}(\theta)}{f_t(\theta_0)} \end{pmatrix}\right\|,$$

so that, for t sufficiently large, we get

$$\left\|\frac{\nabla^\theta f_{t+1}(\theta)}{f_t(\theta_0)}\right\| \leq C + \sum_{i=0}^{\infty}(1-\alpha)^i \left\|\frac{\epsilon_{t-i} f_{t-i}(\theta_0)}{f_t(\theta_0)}\right\| + \sum_{i=0}^{\infty}(1-\alpha)^i \left\|\frac{f_{t-i}(\theta)}{f_t(\theta_0)}\right\|.$$

Since Theorem 1 implies that $\mathbb{E}\left[\log^+\left|\epsilon_t f_t(\theta_0)\right|\right] \leq \log 2 + \mathbb{E}\left[\log^+|\epsilon_t|\right] + \mathbb{E}\left[\log^+\left|f_t(\theta_0)\right|\right] < \infty$ and $\mathbb{E}\left[\log^+\left|f_t(\theta)\right|\right] < \infty$, then by Lemma 2.2 of Berkes et al. (2003) and using the exponential decay of the weights $(1-\alpha)$, it holds that $\sum_{i=0}^{t}(1-\alpha)^i\left\|\epsilon_{t-i} f_{t-i}(\theta_0)\right\| < \infty$ and $\sum_{i=0}^{t}(1-\alpha)^i\left\|f_{t-i}(\theta)\right\| < \infty$ with probability one.

Next, we also note that by Assumption 1 it clearly holds that $\mathbb{E}\left[|\epsilon_t|^r\right] < \infty$ for some sufficiently small $r > 0$, whereas in Theorem 1 we already proved that $\mathbb{E}\left[\left|f_t(\theta_0)\right|^r\right] < \infty$. From this, it follows that $\mathbb{E}\left[\sup_{\theta\in\Theta}\left|f_t(\theta)\right|^r\right] < \infty$, because, for t sufficiently large and the

strict stationarity of $\left\{f_t(\theta)\right\}_{t\in\mathbb{Z}}$, we have

$$\begin{aligned} f_{t+1}(\theta) =& \omega + (1-\alpha)\, f_t(\theta) + \alpha \ln(1+y_t) = \omega + (1-\alpha)\, f_t(\theta) + \alpha\epsilon_t f_t(\theta_0) \\ =& \frac{\omega}{\alpha} + \sum_{i=0}^{\infty} (1-\alpha)^i\, \epsilon_{t-i} f_{t-i}(\theta_0), \end{aligned} \tag{B.7}$$

so that, for all $\delta > 0$, an application of the Markov's and Chauchy-Schwartz inqualities yields

$$\mathbb{P}\left(\sup_{\theta\in\Theta} \sum_{i=0}^{\infty} (1-\alpha)^i\, \epsilon_{t-i} f_{t-i}(\theta_0) > \delta\right) \leq \delta^{-r/2} \mathbb{E}\left[|\epsilon_0|^r\right] \mathbb{E}\left[\left|f_0(\theta_0)\right|^r\right] \sup_{\theta\in\Theta} \sum_{i=0}^{\infty} (1-\alpha)^i < \infty.$$

Moreover, using the almost sure representation in (B.7), we have

$$\left\| \nabla^{\omega} \left(\frac{f_{t+1}(\theta)}{f_{t+1}(\theta_0)} \right) \right\| = \left\| \frac{\nabla^{\omega} f_{t+1}(\theta)}{f_{t+1}(\theta_0)} \right\| = \left\| \left(\alpha\, f_{t+1}(\theta_0)\right)^{-1} \right\| \leq \left\| \underline{\alpha}\, \omega_0 \right\|^{-1}, \tag{B.8}$$

and, using $\epsilon_t f_t(\theta_0) \leq \alpha_0^{-1} f_{t+1}(\theta_0)$,

$$\begin{aligned} \left\| \frac{\nabla^{\alpha} f_{t+1}(\theta)}{f_{t+1}(\theta_0)} \right\| &= \left\| \frac{-\frac{\omega}{\alpha^2} + \sum_{i=0}^{\infty} (1-\alpha+i\,\alpha)\,(1-\alpha)^{i-1}\, \epsilon_{t-i} f_{t-i}(\theta_0)}{f_{t+1}(\theta_0)} \right\| \\ &\leq \frac{\overline{\omega}}{\underline{\alpha}^2\, \omega_0} + \left\| \frac{\sum_{i=0}^{\infty} (1-\alpha+i\,\alpha)\,(1-\alpha)^{i-1}\, \epsilon_{t-i} f_{t-i}(\theta_0)}{\omega_0 + \sum_{i=0}^{\infty} (1-\alpha)^i \epsilon_{t-i} f_{t-i}(\theta_0)} \right\| \\ &= \frac{\overline{\omega}}{\underline{\alpha}^2\, \omega_0} + \left\| \frac{\sum_{i=0}^{\infty} (1-\alpha+i\,\alpha)\,(1-\alpha)^{i-1} \ln(1+y_{t-i})}{\omega_0 + \sum_{i=0}^{\infty} (1-\alpha)^i \ln(1+y_{t-i})} \right\|. \end{aligned} \tag{B.9}$$

The rest of the proof now follows along the same lines as Lemma 5.2 in Berkes et al. (2003). A similar argument proves the result for the second derivative process $\{\nabla^{\theta\theta} z_t^{f,\text{inv}}(\theta)\}_{t\in\mathbb{N}}$. ■

Lemma B.3. Under the conditions of Theorem 1, the Hessian processes $\{\nabla^{\theta\theta} z_t^{f,\text{inv}}(\theta)\}_{t\in\mathbb{Z}}$ has a uniformly bounded moment, that is

$$\mathbb{E}\left[\sup_{\theta\in\Theta} \left\| \nabla^{\theta\theta} Q_t(\theta) \right\|\right] < \infty.$$

Proof of Lemma B.3

Using equation (B.2), together with a combination of Hölder and Minkowsky inequalities, we obtain

$$
\begin{aligned}
&\mathbb{E}\left[\sup_{\theta\in\Theta}\left\|\nabla^{\theta\theta}Q_t(\theta)\right\|\right] \\
\leq &\left(\mathbb{E}\left[\sup_{\theta\in\Theta}\left\|\epsilon_t - z_t^{f,\text{inv}}(\theta)\right\|^2\right]\right)^{1/2}\left(\mathbb{E}\left[\sup_{\theta\in\Theta}\left\|\frac{\nabla^{\theta\theta}z_t^{f,\text{inv}}(\theta)}{z_t^{f,\text{inv}}(\theta)^2} - \frac{\nabla^{\theta}z_t^{f,\text{inv}}(\theta)\nabla^{\theta}z_t^{f,\text{inv}}(\theta)^\top}{z_t^{f,\text{inv}}(\theta)^3}\right\|^2\right]\right)^{1/2} \\
&+\mathbb{E}\left[\sup_{\theta\in\Theta}\left\|\frac{\nabla^{\theta}z_t^{f,\text{inv}}(\theta)\nabla^{\theta}z_t^{f,\text{inv}}(\theta)^\top}{z_t^{f,\text{inv}}(\theta)^2}\right\|\right] \\
\leq &C\times\left(\left(\mathbb{E}\left[\epsilon_t^2\right]\right)^{1/2} + \left(\mathbb{E}\left[\sup_{\theta\in\Theta}\left|z_t^{f,\text{inv}}(\theta)\right|^2\right]\right)^{1/2}\right) \\
&\times\left(\left(\mathbb{E}\left[\sup_{\theta\in\Theta}\left\|\frac{\nabla^{\theta\theta}z_t^{f,\text{inv}}(\theta)}{z_t^{f,\text{inv}}(\theta)^2}\right\|^2\right]\right)^{1/2} + \left(\mathbb{E}\left[\sup_{\theta\in\Theta}\left\|\frac{\nabla^{\theta}z_t^{f,\text{inv}}(\theta)\nabla^{\theta}z_t^{f,\text{inv}}(\theta)^\top}{z_t^{f,\text{inv}}(\theta)^3}\right\|^2\right]\right)^{1/2}\right) \\
&+C\times\mathbb{E}\left[\sup_{\theta\in\Theta}\left\|\frac{\nabla^{\theta}z_t^{f,\text{inv}}(\theta)\nabla^{\theta}z_t^{f,\text{inv}}(\theta)^\top}{z_t^{f,\text{inv}}(\theta)^2}\right\|\right].
\end{aligned}
$$

By Assumption 1 we clearly have that $\mathbb{E}\left[\epsilon_t^2\right] = 1$ whereas by Theorem 1(iii) it holds that $\mathbb{E}\left[\sup_{\theta\in\Theta}\left|z_t^{f,\text{inv}}(\theta)\right|^n\right] = \mathbb{E}\left[\sup_{\theta\in\Theta}\left|1/z_t^{f}(\theta)\right|^n\right] < \infty$ for any $n > 0$. Furthermore, in Lemma B.2 we proved that the derivative processes satisfy $\mathbb{E}\left[\sup_{\theta\in\Theta}\left\|\nabla^{\theta}z_t^{f,\text{inv}}(\theta)\right\|^n\right] < \infty$ and $\mathbb{E}\left[\sup_{\theta\in\Theta}\left\|\nabla^{\theta\theta}z_t^{f,\text{inv}}(\theta)\right\|^n\right] < \infty$ for any $n > 0$, and therefore, by combining all these results, we infer that $\mathbb{E}\left[\sup_{\theta\in\Theta}\left\|\nabla^{\theta\theta}Q_t(\theta)\right\|\right] < \infty$, thus concluding the proof of the Lemma. ■

Web Appendix C: Derivation of market risk measures

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

$$\begin{aligned}\overline{G}(x_t) &= 1 - G(x_t) = \mathbb{P}(X_t > x_t) = \mathbb{P}(X_t > \tau_t)\mathbb{P}(X_t > x_t | X_t > \tau_t) \\ &= \mathbb{P}(X_t > \tau_t)\mathbb{P}(X_t > x_t | (X_t - \tau_t)/\tau_t > 0) = \overline{G}(\tau_t)\overline{F}(y_t),\end{aligned}$$

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

$$\begin{aligned}&\overline{G}(x_t) = \overline{G}(\tau_t)\overline{F}(y_t) = \gamma \\ \iff &\frac{n_t}{t}(1 + y_t)^{-1/f_t} = \gamma \\ \iff &1 + \tau_t^{-1}(q_t^\gamma(X_t) - \tau_t) = \left(\frac{\gamma}{n_t/t}\right)^{-f_t} \\ \iff &q_t^\gamma(X_t) = \tau_t \left(\frac{\gamma}{n_t/t}\right)^{-f_t},\end{aligned}$$

where n_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 $\text{ES}^\gamma(X_t)$ is given by

$$\begin{aligned}\text{ES}^\gamma(X_t) &= \frac{1}{\gamma}\int_{1-\gamma}^{1} q_t^s(X_t)\mathrm{d}s \\ &= \frac{\text{VaR}^\gamma(x_t \mid \mathcal{F}_{t-1}, \theta)}{1 - f_t},\end{aligned}$$

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

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

for $f_t < 1$.

Web Appendix D: PZC (2019) thresholds

This section presents our empirical results when Patton et al. (2019) (PZC)'s joint model of ES and VaR is used to obtain the thresholds τ_t. Table D.1 reports the model's static parameter estimates. Table D.2 reports VaR violation rates. Figures D.1 and D.2 plot the four exchange rate log-returns, tail shape parameters, and the corresponding extreme risk measures.

Table D.1: PZC (2019) parameter estimates

Parameter estimates for the dynamic EVT model. The first illustration considers EUR/USD and RUB/USD daily exchange rate log-returns. The second illustration considers BTC/USD and ETH/USD hourly cryptocurrency log-returns. Standard error estimates are in round brackets, p-values are in square brackets.

	First illustration		Second illustration	
	EUR/USD	RUB/USD	BTC/USD	ETH/USD
α	0.007	0.027	0.005	0.004
	(0.005)	(0.014)	(0.002)	(0.002)
	[0.12]	[0.05]	[0.02]	[0.07]
β^τ	0.996	0.997	0.995	0.989
γ^τ	0.009	0.023	0.007	0.010
b^τ	-0.979	-0.562	-1.479	-1.958
a^τ	-0.691	-0.329	-0.815	-1.170
T	6501	7223	41193	41258
T^*	640	786	2217	2211
loglik	-124.503	-496.864	-1493.521	-1161.735
AIC	251.006	995.728	2989.043	2325.469

Table D.2: PZC (2019) VaR violation rates

VaR violation rates and p-values for the Kupiec (2000) test that this rate coincides with the nominal rate. Top and bottom panels consider VaR estimates and their violation rates at conventional non-extreme (90% and 95%) and more extreme (99%) confidence levels.

	First illustration		Second illustration	
	EUR/USD	RUB/USD	BTC/USD	ETH/USD
	$\hat{\tau}_t = \text{VaR}_t^{0.9}$		$\hat{\tau}_t = \text{VaR}_t^{0.95}$	
VaR violations	9.84%	10.88%	5.38%	5.36%
p-value	0.68	0.01	0.00	0.00
	$\text{VaR}_t^{0.99}$		$\text{VaR}_t^{0.99}$	
VaR violations	0.49%	0.84%	1.06%	1.00%
p-value	0.00	0.17	0.26	0.98

Figure D.1: Exchange rate log-returns, tail shape, and extreme risks

Top panels: EUR/USD (left) and RUB/USD (right) daily log-returns (× 100). We focus on the adverse right tail, corresponding to a *depreciation* of the local currency viz-a-viz the U.S. dollar. Thresholds τ_t are reported at a 90% confidence level. VaR and ES are plotted at an extreme 99% confidence level. Middle panels: filtered tail shape parameter f_t with asymmetric 95% confidence band. Bottom panels: Zoomed-in extreme risks with key events.

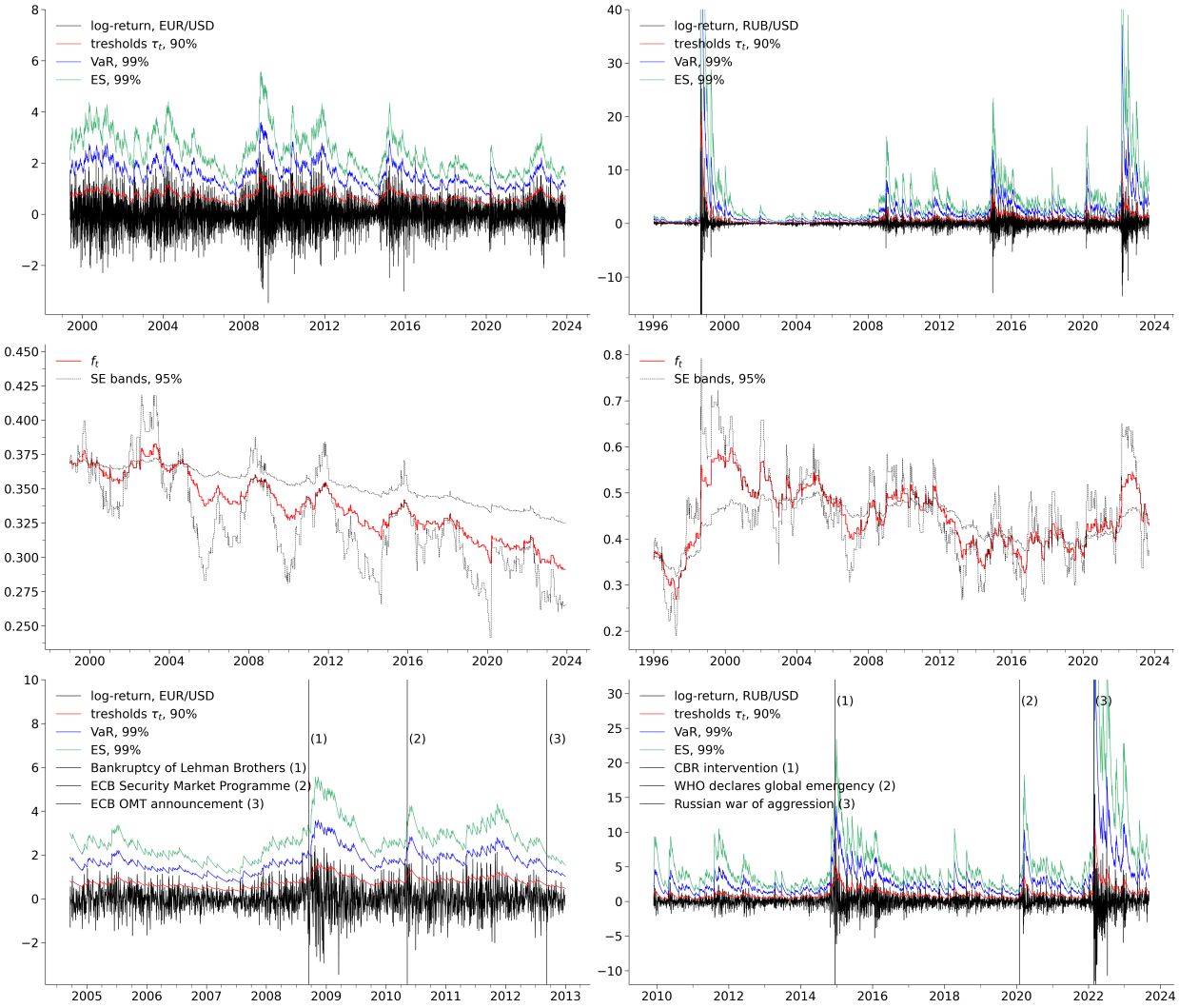
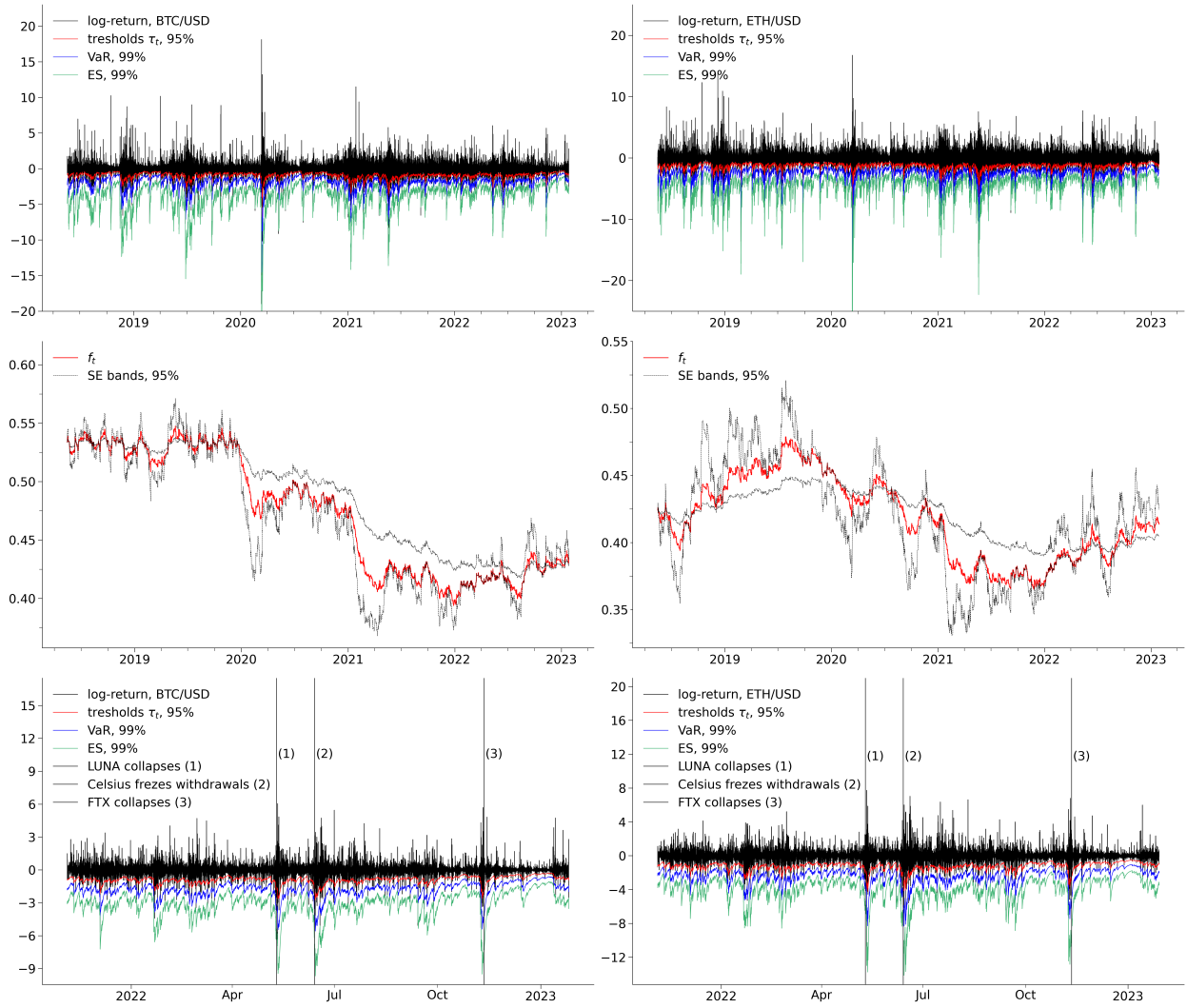


Figure D.2: Cryptocurrency log-returns, tail shape, and extreme risks

Top panels: Bitcoin/USD (left) and Ether/USD (right) hourly log-returns (× 100). Thresholds τ_t are reported at a 95% confidence level. VaR and ES are plotted at an extreme 99% confidence level. The thresholds τ_t, VaR, and ES are mirrored at the horizontal axis to correspond to log-returns (instead of log-losses). Middle panels: filtered tail shape parameter f_t with asymmetric 95% confidence band. Bottom panels: Zoomed-in extreme risks with key events.

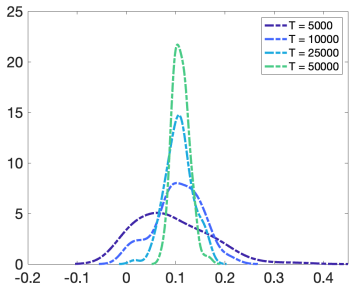


Web Appendix E: Extra simulation results

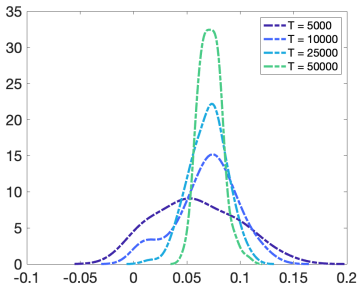
Figure E.1: Simulation results scenario 3

Top panels show kernel density estimates of the distribution of the MLE for $\hat{\alpha}$ for scenario 3 (i.e., a Gaussian AR(1) for the true $\log(f_t)$, such that the model's tail shape dynamics are mis-specified), using estimated thresholds $\hat{\tau}_t$. The kernel density estimates are based on $S = 1,000$ simulations. Lower panels show the fit of the filtered $\hat{f}_t(\hat{\theta})$ to the true $f_t(\theta_0)$ in a typical simulation run for two of the sample sizes. Note that the number of POTs is about $1 - \kappa = 10\%$ of the sample size given the mixture setup of the DGP.

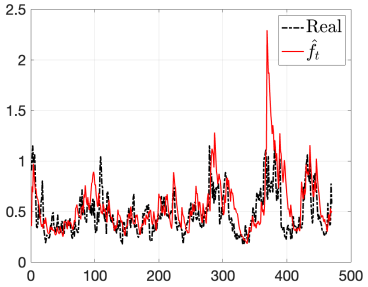
(a) $\hat{\alpha}$ ($\hat{\omega}$ free, $\hat{\tau}_t$)



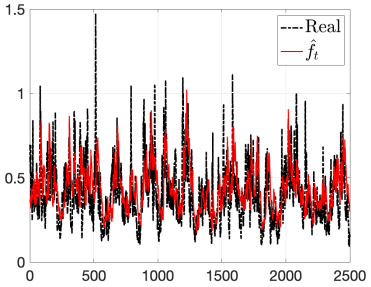
(b) $\hat{\alpha}$ ($\hat{\omega} = 0$, $\hat{\tau}_t$)



(c) fit of filter ($T = 5000$)



(d) fit of filter ($T = 25000$)



Recent Working Papers:

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

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
Modeling extreme events: time-varying extreme tail shape *by Bernd Schwaab, Xin Zhang and André Lucas*	2020:399
The Effects of Government Spending in the Eurozone *by Ricardo Duque Gabriel, Mathias Klein and Ana Sofia Pessoa*	2020:400
Narrative Fragmentation and the Business Cycle *by Christoph Bertsch, Isaiah Hull and Xin Zhang*	2021:401
The Liquidity of the Government Bond Market – What Impact Does Quantitative Easing Have? Evidence from Sweden *by Marianna Blix Grimaldi, Alberto Crosta and Dong Zhang*	2021:402
Five Facts about the Distributional Income Effects of Monetary Policy *by Niklas Amberg, Thomas Jansson, Mathias Klein and Anna Rogantini Picco*	2021:403
When domestic and foreign QE overlap: evidence from Sweden *by Paola Di Casola and Pär Stockhammar*	2021:404
Dynamic Macroeconomic Implications of Immigration *by Conny Olovsson, Karl Walentin, and Andreas Westermark*	2021:405
Revisiting the Properties of Money *by Isaiah Hull and Or Sattath*	2021:406
The cost of disinflation in a small open economy vis-à-vis a closed economy *by Oleksandr Faryna, Magnus Jonsson and Nadiia Shapovalenko*	2021:407
On the Performance of Cryptocurrency Funds *by Daniele Bianchi and Mykola Babiak*	2021:408
The low-carbon transition, climate commitments and firm credit risk *by Sante Carbone, Margherita Giuzio, Sujit Kapadia, Johannes Sebastian Krämer, Ken Nyholm and Katia Vozian*	2022:409
Seemingly Irresponsible but Welfare Improving Fiscal Policy at the Lower Bound *by Roberto M. Billi and Carl E. Walsh*	2022:410
Pension Reform and Wealth Inequality: Evidence from Denmark *by Torben M. Andersen, Joydeep Bhattacharya, Anna Grodecka-Messi and Katja Mann*	2022:411

Inflation Targeting or Fiscal Activism? *by Roberto M. Billi*	2022:412
Trading volume and liquidity provision in cryptocurrency markets *by Daniele Bianchi, Mykola Babiak and Alexander Dickerson*	2022:413
DISPERSION OVER THE BUSINESS CYCLE: PASSTHROUGH, PRODUCTIVITY, AND DEMAND *by Mikael Carlsson, Alex Clymo and Knut-Eric Joslin*	2022:414
Electoral Cycles in Macroeconomic Forecasts *by Davide Cipullo and André Reslow*	2022:415
The Curious Incidence of Monetary Policy Across the Income Distribution *by Tobias Broer, John Kramer and Kurt Mitman*	2022:416
Central Bank Mandates and Monetary Policy Stances: through the Lens of Federal Reserve Speeches *by Christoph Bertsch, Isaiah Hull, Robin L. Lumsdaine, and Xin Zhang*	2022:417
The Political Costs of Austerity *by Ricardo Duque Gabriel, Mathias Klein and Ana Sofia Pessoa*	2022:418
Central bank asset purchases: Insights from quantitative easing auctions of government bonds *by Stefan Laséen*	2023:419
Greenflation? *by Conny Olovsson and David Vestin*	2023:420
Effects of foreign and domestic central bank government bond purchases in a small open economy DSGE model: Evidence from Sweden before and during the coronavirus pandemic *by Yildiz Akkaya, Carl-Johan Belfrage, Paola Di Casola and Ingvar Strid*	2023:421
Dynamic Credit Constraints: Theory and Evidence from Credit Lines* *by Niklas Amberg, Tor Jacobson, Vincenzo Quadrini and Anna Rogantini Picco*	2023:422
Stablecoins: Adoption and Fragility *by Christoph Bertsch*	2023:423
CBDC: Lesson from a Historical Experience *by Anna Grodecka-Messi and Xin Zhang*	2023:424
Do Credit Lines Provide Reliable Liquidity Insurance? Evidence from Commercial-Paper Backup Lines *by Niklas Amberg*	2023:425
Price Pass-Through Along the Supply Chain: Evidence from PPI and CPI Microdata *by Edvin Ahlander, Mikael Carlsson and Mathias Klein*	2023:426
Cash for Transactions or Store-of-Value? A comparative study on Sweden and peer countries *by Carl Andreas Claussen, Björn Segendorf and Franz Seitz*	2023:427
Fed QE and bank lending behaviour: a heterogeneity analysis of asset purchases *by Marianna Blix Grimaldi and Supriya Kapoor*	2023:428
Monetary policy in Sweden after the end of Bretton Woods *by Emma Bylund, Jens Iversen and Anders Vredin*	2023:429
Banking Without Branches *by Niklas Amberg and Bo Becker*	2024:430
Climate impact assessment of retail payment services *by Niklas Arvidsson, Fumi Harahap, Frauke Urban and Anissa Nurdiawati*	2024:431
Four Facts about International Central Bank Communication by *Christoph Bertsch, Isaiah Hull, Robin L. Lumsdaine, and Xin Zhang*	2024:432
Optimal Monetary Policy with r* < 0 *by Roberto Billi, Jordi Galí, and Anton Nakov*	2024:433
Quantitative Easing, Bond Risk Premia and the Exchange Rate in a Small Open Economy *by Jens H. E. Christensen and Xin Zhang*	2024:434
Supply-Chain Finance: An Empirical Evaluation of Supplier Outcomes *by Niklas Amberg, Tor Jacobson and Yingjie Qi*	2024:435
Optimal Contracts and Inflation Targets Revisited *by Torsten Persson and Guido Tabellini*	2024:436
Potential Climate Impact of Retail CBDC Models *by Niklas Arvidsson, Fumi Harahap, Frauke Urban and Anissa Nurdiawati*	2024:437
Do we need firm data to understand macroeconomic dynamics? *by Michele Lenza and Ettore Savoia*	2024:438

Inflation-Dependent Exchange Rate Pass-Through in Sweden: Insights from a Logistic Smooth Transition VAR Model *by Gabriella Linderoth and Malte Meuller*	2024:439
Quantitative Easing and the Supply of Safe Assets: Evidence from International Bond Safety Premia *by Jens H. E. Christensen, Nikola N. Mirkov and Xin Zhang*	2024:440
Bank fragility and the incentives to manage risk *by Toni Ahnert, Christoph Bertsch, Agnese Leonello and Robert Marquez*	2024:441
A Traffic-Jam Theory of Growth *by Daria Finocchiaro and Philippe Weil*	2024:442
Intertemporal MPC and Shock Size *by Tullio Jappelli, Ettore Savoia and Alessandro Sciacchetano*	2024:443
Plundered or profitably pumped-up? The effects of private equity takeover *by Anders Kärnä and Samantha Myers*	2024:444
Measuring Riksbank Monetary Policy: Shocks and Macroeconomic Transmission *by Jakob Almerud, Dominika Krygier, Henrik Lundvall and Mambuna Njie*	2024:445

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