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

September 6, 2025

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 two cryptocurrency exchange rates, we illustrate how the simple single-parameter model is competitive in capturing the dynamics of VaR and ES, particularly in the extreme tail.

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

JEL classification: *C22, G11.*

*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. 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.

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 expected shortfall levels 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 99.5% 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 Zakoïan, 2019) and particular location models (Blasques et al., 2024), 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 inferred simply and reliably in- and out-of-sample? 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, 2019) to the EVT setting.

To obtain the time-varying thresholds required for scaling, we use 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 two hourly cryptocurrency exchange rates from May 2018 until August 2025, thus including the so-called "second crypto winter" of 2022. We find that the tail shape parameters vary over time, with the adverse left tail's shape parameter ranging between 0.3 and 0.6, implying the existence of between 1 and 3 integer 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 the crypto intermediary Celsius in November 2022. An out-of-sample evaluation exercise, performed for these data and for different combinations of EVT thresholds' and risk measures' tail probabilities, suggests that our new method performs well in the extreme tails when the thresholds (less far out in the tail) are chosen dynamically by the method of Patton et al. (2019). In addition, despite only having one parameter, our new and more parsimonious EVT model behaves at par and sometimes better in the extreme tail than the two time-varying parameter model of D'Innocenzo et al. (2024) and is also easier to estimate.

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 contains 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 two cryptocurrency 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 asset price losses or negative returns, for $t = 1, \ldots, T$, where T denotes the number of observations of X_t. We are interested in describing the

conditional extreme right 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 values of X_t above some high (possibly time-varying) threshold 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 $\tilde{f}_t > 0$, can be described by

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

for $x_t > 0$, $\tilde{f}_t \in \tilde{\mathcal{F}}_{t-1}$, and $\tilde{\mathcal{F}}_{t-1} = \{X_1, \ldots, X_{t-1}\}$ denoting the conditioning set. We refer to the reciprocal of the tail shape, $1/\tilde{f}_t$, 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.)

Let t_i for $i = 1, \ldots, n_T$, denote the time indices at which we observe a POT, i.e., $X_{t_i} > \tau_{t_i}$, where $n_T < T$ denotes the number of POTs. We now define the scaled POTs Y_i as $Y_i = \left(X_{t_i} - \tau_{t_i}\right)/\tau_{t_i} \Leftrightarrow X_{t_i} = \tau_{t_i} + \tau_{t_i} \, Y_i$. Substituting Y_i into the above limiting result and defining $f_i = \tilde{f}_{t_i}$, and $y_i = \tilde{f}_{t_i} x_{t_i}$, we obtain for $y_i > 0$ that

$$\begin{aligned}
\lim_{\tau_{t_i} \to \infty} \mathbb{P}\left(Y_i > y_i \mid \tilde{\mathcal{F}}_{t_i-1}\right) &= \lim_{\tau_{t_i} \to \infty} \mathbb{P}\left(Y_i > y_i \mid Y_i > 0, \, \tilde{\mathcal{F}}_{t_i-1}\right) \\
&= \lim_{\tau_{t_i} \to \infty} \mathbb{P}\left(\tau_{t_i} + \tau_{t_i} \, Y_i > \tau_{t_i} + \tau_{t_i} \, y_i \mid \tau_{t_i} + \tau_{t_i} \, Y_i > \tau_{t_i}, \, \tilde{\mathcal{F}}_{t_i-1}\right)
\end{aligned}$$

$$= \lim_{\tau_{t_i} \to \infty} \mathbb{P}\left(X_{t_i} > \tau_{t_i} + \tau_{t_i}\, y_i \mid X_{t_i} > \tau_{t_i}, \tilde{\mathcal{F}}_{t_i-1}\right) = (1+y_i)^{-1/f_i}, \quad (1)$$

which yields the generalized Pareto distribution (GPD) as a limiting approximation for the conditional distribution $\mathbb{P}(Y_i > y_i \mid \tilde{\mathcal{F}}_{t_i-1})$ of the extreme tail of the scaled POTs Y_i.

The advantage of using scaled POTs $Y_i = (X_{t_i} - \tau_{t_i})/\tau_{t_i}$ rather than their unscaled counterparts $(X_{t_i} - \tau_{t_i})$ is that it considerably simplifies the resulting expression for the distribution function in (1) for the extreme tails compared to, for example, 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_i, rather than by both a tail shape and tail scale parameter as in earlier papers. This simplification proves particularly helpful when later deriving the asymptotic properties of the model and the maximum likelihood estimator. Empirically, the result requires that we look sufficiently far into the tail, as the result builds on the limiting result of de Haan and Ferreira (2006), which states that the tail scale can be written as $\tilde{f}_t \tau_t$ for sufficiently large τ_t.

2.2 Filtering the conditional tail shape parameter

We allow the tail shape to change gradually over time. In particular, we assume that $\tilde{f}_t$ possibly changes each time a POT materializes: $\tilde{f}_t = f_i$ if $t = t_i$, and $\tilde{f}_t = \tilde{f}_{t-1}$ otherwise. It is then, and only then, that we obtain information about the tail shape behavior of Y_i and thus X_{t_i}. 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 make the simplifying assumption $P(Y_i > y_i \mid \tilde{\mathcal{F}}_{t_i-1}) = P(Y_i > y_i \mid \mathcal{F}_{i-1})$ with $\mathcal{F}_{i-1} = \tilde{\mathcal{F}}_{t_{i-1}}$, i.e., the conditional distribution of the extremes is not affected by intermediate, non-extreme observations X_t for $t = t_{i-1}+1, \ldots, t_i - 1$. We can then introduce score-driven dynamics for f_i as in Creal et al. (2013), with $f_{i+1} = \omega + \beta f_i + \alpha s_i$, for $i \in \mathbb{Z}$, where s_i is the inverse information scaled derivative of the log predictive GPD tail density. Transforming the cdf expression in (1)

into a conditional pdf $p(y_i \mid \mathcal{F}_{i-1}) = f_i^{-1}\,(1+y_i)^{-f_i^{-1}-1}$ for $y_i > 0$, we obtain the following expression for the scaled score,

$$\nabla_i = \frac{\partial\left(-\ln(f_i) - \left(f_i^{-1}+1\right)\ln\left(1+y_i\right)\right)}{\partial f_i} = \frac{1}{f_i^2}\ln\left(1+y_i\right) - \frac{1}{f_i},$$
$$\mathcal{I}_{i-1} = \mathbb{E}\left[\nabla_i^2 \;\middle|\; \mathcal{F}_{i-1}\right] = f_i^{-2}, \qquad s_i = \mathcal{I}_{i-1}^{-1}\,\nabla_i = \ln(1+y_i) - f_i.$$

In this paper, we are particularly interested in filtering the tail shape parameter f_i from the data using an *integrated* ($\beta = 1$) score-driven filtering equation,

$$f_{i+1} = \omega + \beta\, f_i + \alpha\, s_i = \omega + f_i + \alpha\, s_i = \omega + (1-\alpha)\, f_i + \alpha\, \ln(1+y_i), \tag{2}$$

for $i \in \mathbb{Z}$, i.e., a model where we have set β equal to unity rather than to a value inside the unit interval as is commonly done in the literature. The last equation highlights this further by spelling out the scaled score $s_i = \ln(1+y_i) - f_i$ and rewriting the expression such that the coefficients $1-\alpha$ and α in front of f_i and $\ln(1+y_i)$, respectively, add up to 1 as in the iGARCH literature (see, e.g., Francq and Zakoïan, 2019). Note that if $f_{i_0} > 0$ for some $i_0 \in \mathbb{Z}$, and $\omega > 0$ and $0 < \alpha < 1$, then f_i is non-negative for all $i \geq i_0$.

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 Zakoïan, 2019, and references therein), integrated score-driven models have thus far hardly received 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. (2024) 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_i 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_i 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, 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_i to its stationary and ergodic limit always results in degenerate, thin-tailed tail behavior with $f_i = 0$ for all $i \in \mathbb{Z}$.

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 $\theta = (\omega, \alpha)$ and write down the likelihood function and the maximum likelihood estimator (MLE) $\hat{\theta}_{n_T}$ for the POTs only,

$$\hat{\theta}_{n_T} = \arg\max_{\theta \in \Theta} \hat{L}_{n_T}(\theta), \qquad \hat{L}_{n_T}(\theta) = \frac{1}{n_T} \sum_{i=1}^{n_T} \hat{\ell}_i(\theta), \tag{3}$$
$$\hat{\ell}_i(\theta) = -\ln \hat{f}_i(\theta) - \left(1 + \frac{1}{\hat{f}_i(\theta)}\right) \ln(1 + y_i),$$

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

$$\hat{f}_{i+1}(\theta) = \omega + \hat{f}_i(\theta) + \alpha \left(\ln(1 + y_i) - \hat{f}_i(\theta)\right), \tag{4}$$

evaluated at some θ inside the parameter space Θ and initialized at some $\hat{f}_1 > 0$.

2.3 Filtering extreme Value-at-Risk and Expected Shortfall

Evaluated at their maximum likelihood estimates, the filtered values $\hat{\tilde{f}}_t(\hat{\theta}_{n_T})$ 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). Let κ denote the right-tail probability used to define the right-tail threshold τ_t, i.e., $\mathbb{P}(X_t > \tau_t \mid \tilde{\mathcal{F}}_{t-1}) = \kappa$. Then we can estimate the conditional VaR and ES at a more extreme right-tail probability level $\gamma < \kappa$ by

$$\mathrm{VaR}^{1-\gamma}(X_t \mid \tilde{\mathcal{F}}_{t-1}) = \tau_t \cdot \left(\frac{\gamma}{\zeta_t}\right)^{-\hat{\tilde{f}}_t(\hat{\theta}_{n_T})}, \tag{5}$$

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

where $t = 1, \ldots, T$ and where ζ_t is an estimator of the tail probability $\mathbb{P}(X_t > \tau_t \mid \tilde{\mathcal{F}}_{t-1})$, e.g., the percentage of POT observations up to time t. Alternatively, one can replace ζ_t directly by the postulated nominal POT probability κ. 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 $Y_i = (X_{t_i} - \tau_{t_i})/\tau_{t_i}$ rather than their unscaled counterparts $(X_{t_i} - \tau_{t_i})$; see Web Appendix C for derivations. In particular, our expressions only require the estimation of the tail shape parameter $\hat{\tilde{f}}_t(\hat{\theta}_{n_T})$, and not of any auxiliary tail scale. Due to the result of de Haan and Ferreira (2006), the expressions coincide again if one uses $\hat{\tilde{f}}_t(\hat{\theta}_{n_T})\tau_t$ as a tail scale parameter in the former papers.

Given the straightforward formulation of the filter for $f_i(\theta)$ 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 expected shortfall 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.

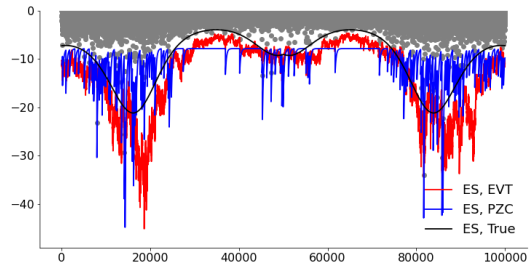
We still consider the methodology of Patton et al. (2019), or one similar to it, as key to our analysis, but mainly for the determination of the (less extreme) thresholds τ_t; see Sections 4 and 5. Its advantages are (i) its reliance on very few distributional assumptions, and (ii) its use of much higher 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 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 smoothly 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-

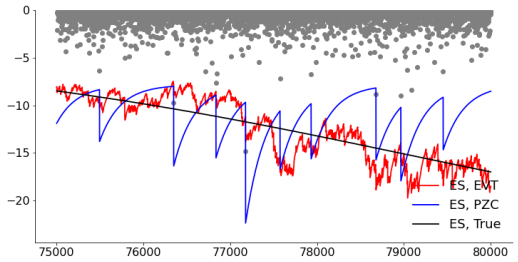
Figure 1: EVT-based versus PZC filtering results at 99.9%

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.

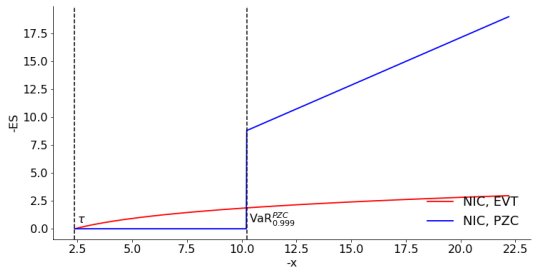
(a) ES dynamics: full sample



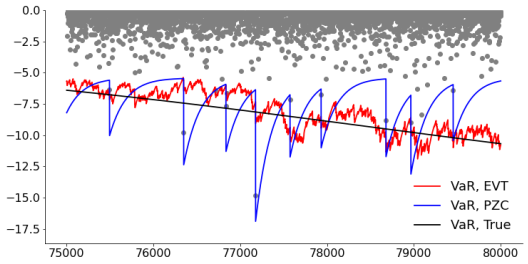
(b) ES dynamics: zoomed in



(c) NIC of ES



(d) VaR dynamics: zoomed in



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 by plotting the reaction of the 99.9% ES to an observation $-X_t$ for a 99.9% VaR level of 10.0 and an EVT 5% tail area threshold level of 2.5. 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 $c = \alpha \log(\kappa/\gamma)$ and for some C that does not depend on X_t; 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 $\tilde{f}_t$ is high. As a result, $\tilde{f}_t$ 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.

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. In our asymptotic framework, we take the thresholds τ_t as given and let n_T, the number of POTs, diverge to infinity. For a recent framework where the number of observations T tends to infinity and n_T is treated as random, see Cavaliere et al. (2025). Sections 4 and 5 discuss how we estimate τ_t in our simulations and empirical applications.

We first define two random variables, namely a standard uniform $u_i \sim U(0,1)$, and a standard unit exponential $\epsilon_i = -\ln(1-u_i) \sim \mathcal{E}xp(1)$. Define $G(y_i \mid f_i) = 1-(1+y_i)^{-1/f_i}$ as the expression for the tail GPD approximation from (1). We let $f_i(\theta_0)$ 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 $f_i(\theta_0)$ is the unique stationary and ergodic limit of its initialized counterpart $\hat{f}_i(\theta_0)$ from (4), initialized at $\hat{f}_1$.

Using these definitions, we obtain

$$G(y_i \mid f_i(\theta_0)) = u_i = 1-(1+y_i)^{-1/f_i(\theta_0)} \iff \frac{1}{f_i(\theta_0)}\ln(1+y_i) = -\ln(1-u_i) = \epsilon_i, \quad (7)$$

and thus

$$f_{i+1}(\theta_0) = \omega_0 + f_i(\theta_0) + \alpha_0\left(\ln(1+y_i) - f_i(\theta_0)\right) = \omega_0 + \left(1+\alpha_0\left(\epsilon_i - 1\right)\right) f_i(\theta_0), \quad (8)$$

for $i \in \mathbb{Z}$. We make the following assumptions.

Assumption 1. *$\{\epsilon_i\}_{i\in\mathbb{Z}}$ is an independent and identically distributed (IID) noise sequence where each ϵ_i 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* $\theta_0 \in int(\Theta)$*.*

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_i. These realizations then feed into the next tail shape parameter via the recursion (8). 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 i. We now obtain the following theorem, which establishes (i) stationarity and ergodicity of the data y_i and of the uninitialized true time-varying parameter $f_i(\theta_0)$; (ii) invertibility of the filter $\hat{f}_i(\theta)$ 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 $f_i(\theta_0)$ *and* y_i *to* (7) *and* (8)*; moreover, there exists some small* $r > 0$ *such that* $\mathbb{E}|f_i(\theta_0)|^r < \infty$ *and* $\mathbb{E}\left|\ln(1+y_i)\right|^r < \infty$*;*

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

(iii) the ratio process $\hat{z}_i^f(\theta) = f_i(\theta_0)/\hat{f}_i(\theta)$ *converges to a unique stationary and ergodic solution* $z_i^f(\theta) = f_i(\theta_0)/f_i(\theta)$*. Moreover,* $z_i^f(\theta)$ *and* $1/z_i^f(\theta)$ *have finite k-th order moment, uniformly over* Θ*, for any* $k > 0$*.*

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* $f_i(\theta_0)$ *satisfies* $f_i(\theta_0) = 0$ *for all* i*;*

(ii) for $\omega_0 > 0$*, the stationary solution for* $f_i(\theta_0)$ *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 $f_i(\theta_0)$. The intuition for this is immediately clear from the recursion (8), 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 $f_i(\theta_0) = 0$, 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 (8) for $\omega_0 > 0$, and using the fact that the scaled score has conditional expectation zero, we have $\mathbb{E}[f_{i+1}(\theta_0)] = \omega_0 + \mathbb{E}[f_i(\theta_0)]$. It then also follows directly that $f_i(\theta_0)$ 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 $\ln(1 + y_i)$ has a small moment r, the filter is invertible at the DGP. Invertibility, however, holds for generic stationary and ergodic y_i with a $\log^+$-log-moment, meaning that it continues to hold if the model is misspecified. Also note that $\hat{f}_i(\theta)$ does not have a finite first unconditional moment at the DGP, as it is driven by the innovation $\alpha \ln(1 + y_i) = \alpha f_i(\theta_0)\epsilon_i$, where ϵ_i has a unit exponential distribution. The first part of the theorem already implied that $f_i(\theta_0)$ does not have a first moment.

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

$$\begin{aligned}
\hat{\mathcal{Q}}_{n_T}(\theta) &= \hat{L}_{n_T}(\theta) - L_{n_T}(\theta_0) = \sum_{i=1}^{n_T} \hat{Q}_i(\theta) \\
&= \sum_{i=1}^{n_T} -\ln \hat{f}_i(\theta) - (1 + \hat{f}_i(\theta)^{-1}) \ln(1 + y_i) + \ln f_i(\theta_0) + (1 + f_i(\theta_0)^{-1}) \ln(1 + y_i) \\
&= \sum_{i=1}^{n_T} \ln \frac{f_i(\theta_0)}{\hat{f}_i(\theta)} - \frac{f_i(\theta_0)}{\hat{f}_i(\theta)} \epsilon_i + \epsilon_i = \sum_{i=1}^{n_T} \ln \hat{z}_i^f(\theta) \; - \epsilon_i \left(\hat{z}_i^f(\theta) - 1 \right). \qquad (9)
\end{aligned}$$

We also define $\mathcal{Q}_{n_T}(\theta) = L_{n_T}(\theta) - L_{n_T}(\theta_0)$, and define $Q_i(\theta)$ similar to $\hat{Q}_i(\theta)$, but with $\hat{f}_i(\theta)$ replaced by its stationary and ergodic limit $f_i(\theta)$. As the maximizer of $\hat{L}_{n_T}(\theta)$ is the same as that of $\hat{\mathcal{Q}}_{n_T}(\theta)$, the properties of $\hat{z}_i^f(\theta)$ can be used to derive the properties of the MLE. In particular, (9) clarifies that consistency results can be obtained if a first moment exists for $\hat{z}_i^f(\theta)$. This is precisely what the last part of Theorem 1 establishes. Though $f_i(\theta_0)$ does not have a finite first moment for $\omega_0 > 0$, the *normalized* process $\hat{z}_i^f(\theta) = f_i(\theta_0)/\hat{f}_i(\theta)$ 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,* $\hat{\theta}_{n_T} \xrightarrow{a.s.} \theta_0$ *and asymptotically normally distributed with* $\sqrt{n_T}\left(\hat{\theta}_{n_T} - \theta_0\right) \to \mathcal{N}\left(\mathbf{0} \; , \; \boldsymbol{\mathcal{I}}(\theta_0)^{-1}\right)$ *for* $n_T \to \infty$*, where* $\boldsymbol{\mathcal{I}}(\theta_0) = -\mathbb{E}\left[\partial^2 Q_i(\theta_0)/\partial\theta\partial\theta^\top\right]$ *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 $\hat{\tau}_{t_i}$ in the following way. The main arguments in the proof of Theorem 3 continue to hold as long as $F(y_i) = 1 - (1 + y_i)^{-1/f_i(\theta_0)} = 1 - \exp(-\epsilon_i)$ for some *IID* $\{\epsilon_i\}_{i\in\mathbb{Z}}$ that is not necessarily unit exponentially distributed as in the correctly specified case, but still has $\mathbb{E}[\epsilon_i] = 1$ and $\mathbb{E}[\epsilon_i^4] < \infty$. 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 $\boldsymbol{\mathcal{J}}(\theta_0) = \mathbb{E}\left[(\partial Q_i(\theta_0)/\partial\theta)\,(\partial Q_i(\theta_0)/\partial\theta)^\top\right]$ 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 $1-\kappa$, 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 empirical exchange rates. With probability κ, we draw a right tail observation as $X_{t_i} = \tau_{t_i}\exp(f_i\epsilon_i)$, where τ_{t_i} 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_i is the tail shape, and ϵ_i 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 to filter the tail parameters, and the estimated thresholds $\hat{\tau}_t$ based on Patton et al. (2019) in the second experiment. This allows us to study the effect of the estimation of the thresholds on the results.

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

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

where $\{\eta_i\}_{i \in \mathbb{Z}}$ is an *IID* noise sequence where each η_i has a standard Gaussian distribution. The model is thus misspecified, and we investigate whether the model can still reliably track the true time-varying parameter f_i 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 misspecified filtered model matches the unknown DGP as closely as possible (compare Blasques et al., 2015; Beutner et al., 2024).

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. Though setting $\omega_0 \neq 0$ is important in the DGP in order not to get a degenerate solution, setting $\omega = 0$ in the *filter* is less problematic. In fact, the results below show that setting $\omega = 0$ in the filter may even improve the stability of the estimator. This is in line with the fill-in asymptotics results in Beutner et al. (2024), who show that f_i can still be *filtered* consistently using a score-driven filter, even though the filter itself (including the value of ω) is dynamically misspecified, in line with similar earlier results for GARCH volatility filters by Nelson and Foster

(1994, 1995). We consider four different sample sizes: $T \in \{5000, 10000, 25000, 50000\}$ and $\kappa = 10\%$ observations coming from the tail. 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.

4.2 Simulation results

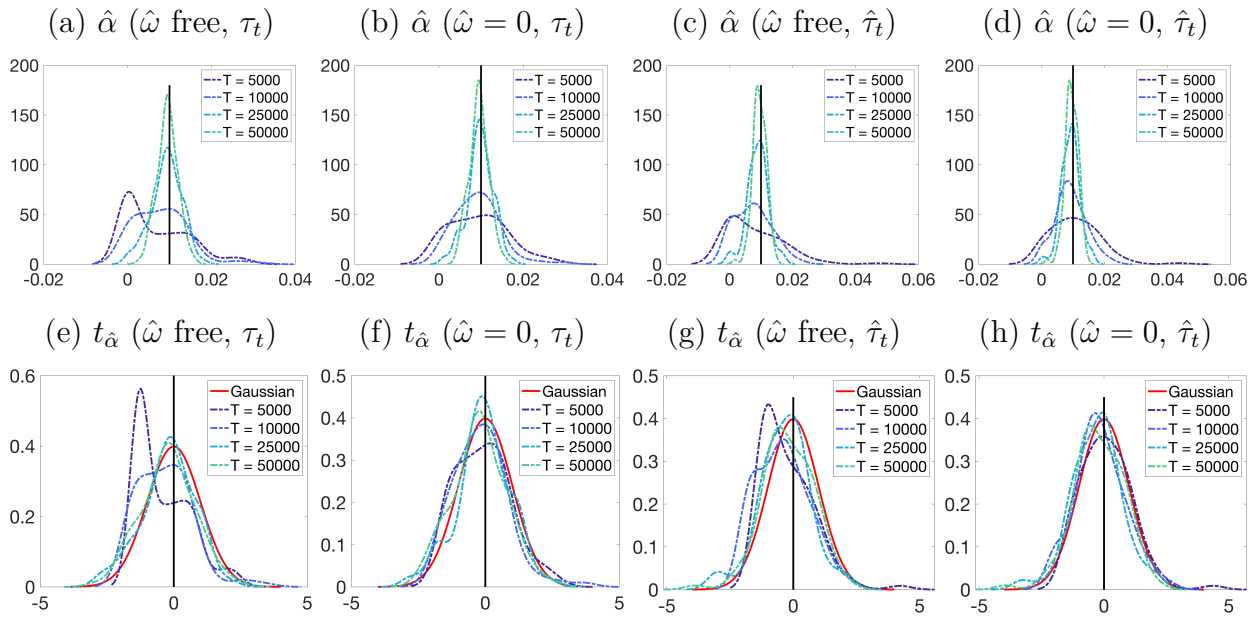
Figure 2 presents the results for the first two sets of experiments. We clearly see that the sample size matters for the results. If the sample size is too small ($\kappa = 10\%$ of $T = 5000$), the parameter $\hat{\alpha}$ is regularly estimated on the edge of the parameter space, i.e., at zero, if 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_i. If the number of observations is too small, it is apparently difficult for the model to distinguish between a trending f_i and an integrated random f_i. 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

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 $\hat{\tau}_t$ and based on Patton et al. (2019). 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.

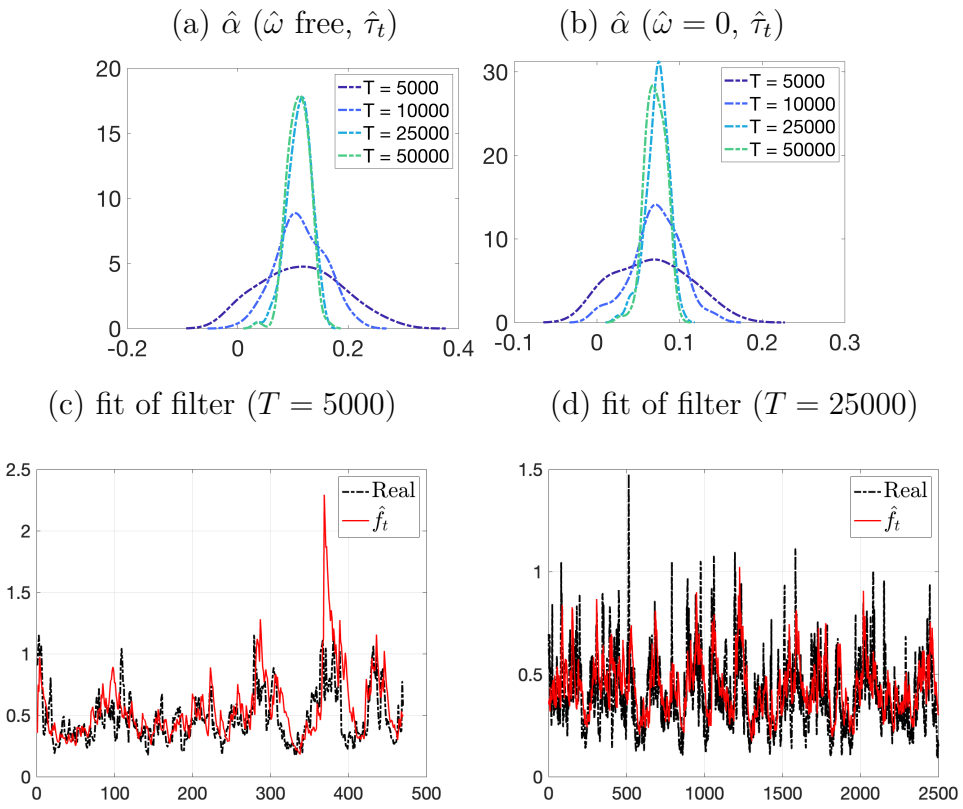


effect on the distribution of $\hat{\alpha}$. For small sample sizes, the effect of $\hat{\alpha}$ collapsing to the edge of the parameter space are more severe when the thresholds are estimated. As in the case of true thresholds τ_t, however, these degenerate cases disappear quickly as the sample size 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 3 presents the results for scenario 3 where the dynamics for the tail shape are fully misspecified in the model compared to the DGP. The results are consistent with the previous findings. For small sample sizes, the model sometimes has 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

Figure 3: 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 $\log(f_i)$, such that the model's tail shape dynamics are misspecified), 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 $\hat{f}_i(\hat{\theta})$ to the true $f_i(\theta_0)$ in a typical simulation run for two of the sample sizes. Note that the number of POTs is about $\kappa = 10\%$ of the sample size given the mixture setup of the DGP.



sample sizes, the problem disappears. The problem is less pronounced if $\hat{\omega}$ is fixed at zero rather than estimated, similar to scenarios 1 and 2. The bottom panels in Figure 3 give the fit of the filter to the true time-varying tail shape parameter $f_i(\theta_0)$ and show that the model fits the true, unobserved process $f_i(\theta_0)$ quite well, despite mis-specification. Note that the filtered parameter $\hat{f}_i(\hat{\theta})$ 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 $\log(f_i(\theta_0))$. Despite this severe mis-specification, the filter still tracks the salient dynamics of the true $f_i(\theta_0)$. The 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, possibly at the cost of a slight bias.

5 Empirical illustration

5.1 In-sample analysis

We obtain hourly prices for Bitcoin (BTC) and Ether (ETH) in USD from Binfinex via Cryptodatadownload.com. The series range from May 15, 2018 to August 31, 2025, yielding 63,000+ observations. We remove days for which we do not observe all trading hours and then transform prices into *negative* log-returns $X_t = -100 \times (\ln s_t - \ln s_{t-1})$ to concentrate on the original series' left tail, where s_t is the price of either BTC or ETH in USD. The left tail is the economically relevant tail for a U.S. investor holding cryptocurrencies. We model the time-varying thresholds τ_t using Patton et al. (2019) with tail probability κ, where we vary κ across 10%, 5%, and 2.5%. We initialize the dynamic tail shape parameter f_i by its static maximum likelihood estimate $\hat{f}_1$ based on the first 50 POTs and estimate $\theta = (\omega, \alpha)'$ by the maximum likelihood estimator given in (3).

Table 1 presents the results. The estimates of α are all small and positive.[1] The values increase if we move farther out into the tail by decreasing the tail probability κ that is used to identify the dynamic thresholds τ_t. This is not surprising. A larger α indicates that f_i reacts

[1]Note that we cannot simply test whether α is equal to zero as $\alpha = 0$ is on the boundary of the parameter space and ω and β are not jointly identified if $\alpha = 0$; see Assumption 2 and Newey and McFadden (1994).

Table 1: Parameter estimates

Parameter estimates for the dynamic EVT model. The estimation sample of negative hourly cryptocurrency log-returns ranges from May 15, 2018 to August 31, 2025. Sandwich standard errors are in parentheses. Dynamic thresholds τ_t are estimated using Patton et al. (2019) with a tail probability of κ, resulting in n_T POTs to estimate the dynamic EVT model. The tail probability is chosen as $\gamma = \kappa/10$.

	$\kappa = 10\%$			$\kappa = 5\%$			$\kappa = 2.5\%$		
	α	$\omega \times 10^4$	n_T	α	$\omega \times 10^4$	n_T	α	$\omega \times 10^4$	n_T
BTC/USD	0.0034	0.0022	6919	0.0053	0.0007	3462	0.0114	0.0106	1689
	(0.0011)	(0.0030)		(0.0017)	(0.0007)		(0.0036)	(0.0475)	
ETH/USD	0.0014	0.0008	6391	0.0077	0.0181	3298	0.0130	0.0776	1648
	(0.0288)	(0.0225)		(0.0033)	(0.0446)		(0.0050)	(0.1158)	

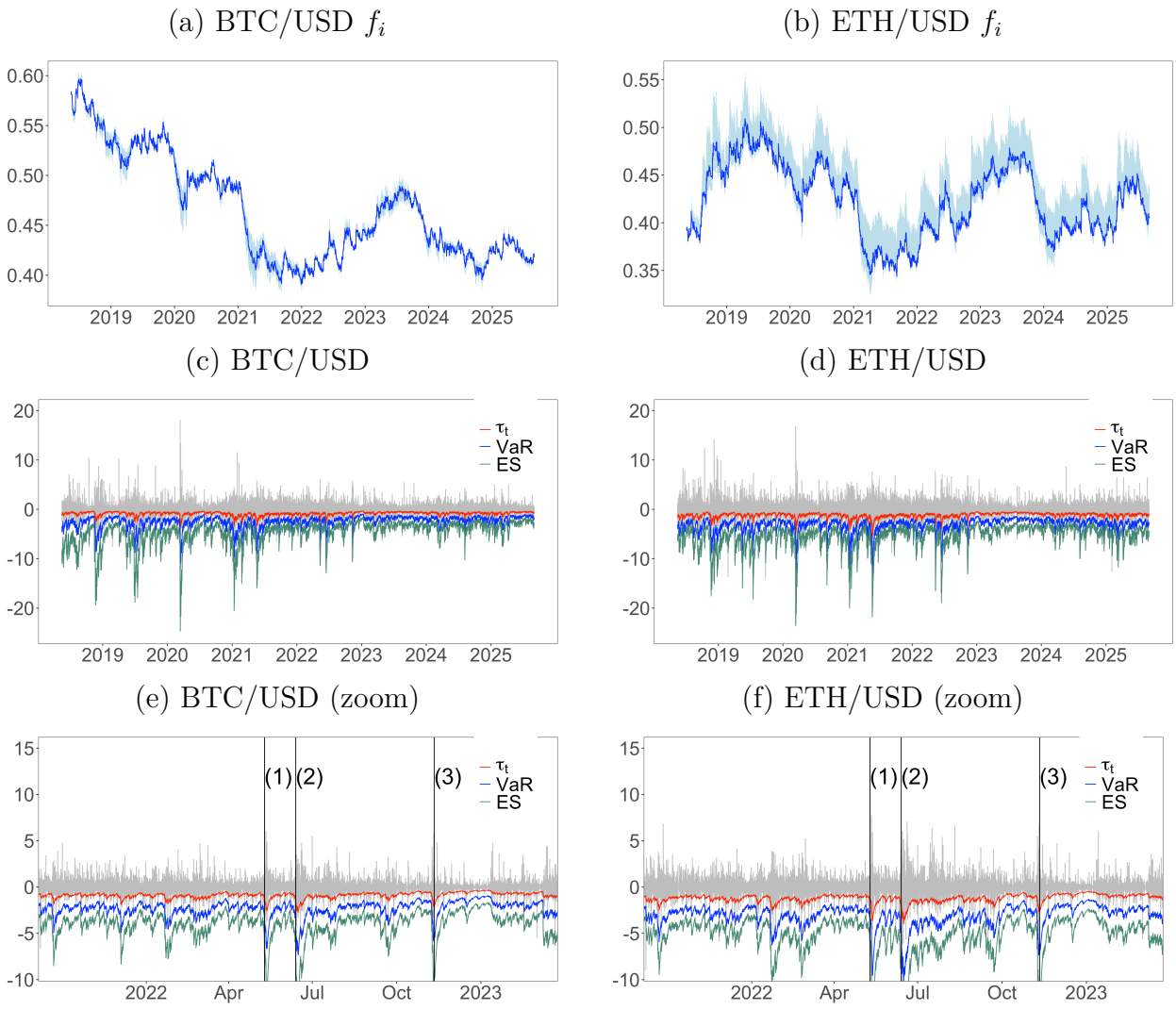
more strongly to new information using the score transition equation (2). As for a smaller κ this information also arrives more sporadically, a stronger reaction is indeed warranted when such tail information finally comes in. Smaller κs also automatically reduce the number of POTs (n_T) available to estimate the model's static parameters α and ω, thus increasing standard errors. Finally, we note that the estimates of ω are very close to zero, given that they are multiplied by 10^4 in Table 1. In fact, setting ω to zero in the estimation hardly changes the results, in line with our earlier remarks and the simulation results of Section 4.

Figure 4 gives a graphical presentation of the in-sample estimation results. Figure 4a and 4b present the filtered estimates of $\tilde{f}_t$ along with its asymmetric interquartile range confidence bands. The tail shape parameter varies between approximately 0.4–0.6 for BTC and 0.35–0.50 for ETH. This implies that the existence of a second or third moment is already problematic for BTC and ETH, though not always. The confidence bands around f_t suggest that the tail shape parameter 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 restrictions $\omega > 0$ and $0 < \alpha < 1$, which we have imposed during parameter estimation through non-linear logarithmic and logistic link functions, respectively. Note that the reported confidence bands are conditional on the estimated thresholds $\hat{\tau}_t$.

Panels 4c–4f present BTC's and ETH's log-returns and the 95% confidence level thresholds

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

Top panels: filtered tail shape parameter $\hat{\hat{f}}_t$ with asymmetric interquartile range confidence band. Middle panels: Bitcoin/USD (left) and Ether/USD (right) hourly log-returns ($\times$ 100). Thresholds τ_t are reported at a 95% confidence level. VaR and ES are plotted at an extreme 99.5% confidence level. The thresholds τ_t, VaR, and ES or all for the left tail of the log-returns. Bottom panels: Zoomed-in extreme risks with key events.



τ_t. We also plot the 99.5% VaR and ES, noting that VaR and ES for even more extreme confidence percentages could easily be computed via (5)–(6). There is clear time variation in extreme market risks. 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 99.5% confidence level are particularly

volatile during 2022. Both market risk measures respond strongly to the collapse of the Terra/Luna cryptocurrency on May 10, 2022 (first vertical line), see e.g. Uhlig (2022); the collapse of FTX, a major cryptocurrency intermediary and shadow bank on June 13, 2022 (second vertical line); and the collapse of Celsius, another cryptocurrency intermediary and shadow bank on Nov 11, 2022 (third vertical line). The 99.5% ES approximately tripled around each of these events, before reverting to more "normal" levels later on. Panels 4c and 4d also show the more recent decline in tail fatness between mid-2023 and 2025, in line with the pattern of $\tilde{f}_t$ as shown in the top panels.

If we consider the VaR violation rates, we obtain a values of 0.89%/0.44%/0.24% for BTC, and 0.77%/0.45%/0.24% for ETH, which are all quite close to the nominal levels of 1%/0.5%/0.25%. Indeed, given that we have more than 63,000 observations and testing at a significance level of 1 per cent, we only reject the unconditional coverage test of Kupiec (2000) for $\kappa = 10\%$, i.e., when the limiting approximation of de Haan and Ferreira (2006) for $\tau_t \to \infty$ might not yet be sufficiently accurate. All these test are in-sample, however. In the next subsection, we therefore perform an out-of-sample evaluation of the new model in relation to other dynamic tail-shape models.

5.2 Out-of-sample evaluation

In this section, we provide an out-of-sample evaluation where we compare our single time-varying tail shape model (EVT) from Section 2 with several benchmarks. We first include two alternative time-varying tail shape methods, namely the two-parameter time-varying generalized Pareto model (TVGPD) of D'Innocenzo et al. (2024), and the dynamic VaR-ES model of Patton et al. (2019, referred to as PZC). For the τ_t thresholds required for the EVT and TVGPD methods, we use Patton et al. (2019) at a tail-probability level $\kappa > \gamma$, where the VaR and ES are computed at a $1 - \gamma$ confidence level. We also compute the Patton et al. (2019) VaR and ES for the extreme tail percentage γ rather than the more moderate

κ. We use $\kappa \in \{10\%, 5\%, 1\%\}$ and consider $\gamma = \kappa/10$ for all methodologies, as in Section 5.1. We might expect the TVGPD to generally perform somewhat better than the EVT method, given its additional freedom with the scale parameter. Whether such an advantage persists out-of-sample is, however, uncertain. In addition, we might also expect any such difference to diminish as κ becomes smaller (and thus τ_t becomes larger) and the limiting result of de Haan and Ferreira (2006) becomes effective. For completeness, we also include a simple standard Gaussian GARCH(1,1) model as used in the simulations in Section 4.

We use Nolde and Ziegel (2017)'s approach to compare the different methodologies in terms of VaR and ES forecasts, which in turn is based on strictly consistent scoring rules and Diebold-Mariano (DM) tests applied to loss differentials. In particular, we use their Equations (2.19) and (2.23). The former evaluates the VaR, whereas the latter is based on the double VaR-ES elicitation criterion function as also used in Patton et al. (2019). The results are very similar for the other scoring rules in their paper. We use the observations from 2018–2021 to compute the first in-sample estimates. We then forecast the VaR and ES for the subsequent year, after which we (recursively) update the parameter estimates of all models. All DM comparisons are done vis-à-vis the new EVT model. Negative DM statistics indicate that the EVT model has a lower loss than its competitor and thus performs better in terms of VaR or VaR–ES.

Table 2 presents the out-of-sample results. In line with the 'traffic light' approach in Nolde and Ziegel (2017), we color cells green if the DM statistic is significantly negative at a 5% level, and red if it is significantly positive. The table yields three main takeaways. First, we see that differences in the Nolde and Ziegel (2017) criteria are generally limited for all models, despite the sizable number of out-of-sample observations: the absolute values of the DM statistics in the table often fall below 2 and have a maximum of slightly over 6. All models thus seem to perform reasonably well and to be of approximately comparable quality.

Second, we see that the new model and the TVGPD perform similarly, particularly if we look at the more extreme tail. It is here that the limiting result of de Haan and Ferreira

Table 2: Out-of-sample comparison of tail risk measures

Comparison of the EVT model against the PZC model of Patton et al. (2019), the TVGPD model of D'Innocenzo et al. (2024), and a standard GARCH(1,1) model. The two loss functions are taken from Nolde and Ziegel (2017) and relate to the VaR (2.19) or to VaR and EL jointly (2.23). Negative values indicate that the loss of the EVT model is lower. Green (red) values indicate the EVT loss is significantly lower (higher) at 5%. The tail probability κ is used when estimating the time-varying thresholds τ_t. The VaR and ES for all models are estimated using a more extreme tail probability $\gamma = \kappa/10$, and thus a high confidence level of $1-\gamma$.

Methods	$\kappa = 10\%$		$\kappa = 5\%$		$\kappa = 2.5\%$	
	Nolde and Ziegel (2017) loss function					
	(2.19)	(2.23)	(2.19)	(2.23)	(2.19)	(2.23)
			BTC/USD			
EVT↔PZC	-1.924	3.307	-3.031	-2.973	-1.997	-2.264
	(0.054)	(0.001)	(0.002)	(0.003)	(0.046)	(0.024)
EVT↔TVGPD	-1.364	4.779	-1.289	-0.195	0.928	0.984
	(0.172)	(0.000)	(0.198)	(0.846)	(0.353)	(0.325)
EVT↔GARCH	1.818	1.858	-3.504	-4.370	-5.170	-6.019
	(0.069)	(0.063)	(0.000)	(0.000)	(0.000)	(0.000)
			ETH/USD			
EVT↔PZC	-3.256	0.162	-4.466	-3.951	-3.468	-3.979
	(0.001)	(0.872)	(0.000)	(0.000)	(0.001)	(0.000)
EVT↔TVGPD	-2.392	1.941	-1.662	-0.717	-1.353	-1.599
	(0.017)	(0.052)	(0.097)	(0.473)	(0.176)	(0.110)
EVT↔GARCH	-1.305	-0.546	-4.549	-4.782	-5.418	-6.000
	(0.192)	(0.585)	(0.000)	(0.000)	(0.000)	(0.000)

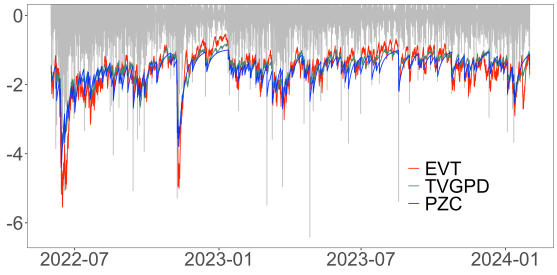
(2006) becomes particularly applicable, and the extra tail scale parameter of the TVGPD model adds less value. We also see that less far out in the tail (e.g., $\kappa = 10$), this limiting result is less accurate and the new model sometimes underperforms compared to the other models, and the additional flexibility of the TVGPD with a time-varying tail scale parameter adds value. This is regularly (insignificantly) reversed for $\kappa = 5\%, 2.5\%$.

Third, we confirm our earlier result that PZC and EVT methodologies nicely complement each other: whereas PZC performs typically better less far out in the tails, the EVT method appears to do a somewhat better job in the extreme tails, with DM statistics that are regularly significant. The EVT model also outperforms the standard Gaussian GARCH model. Given the mutual ranking of PZC, EVT, and TVGPD, the new single time-varying parameter EVT

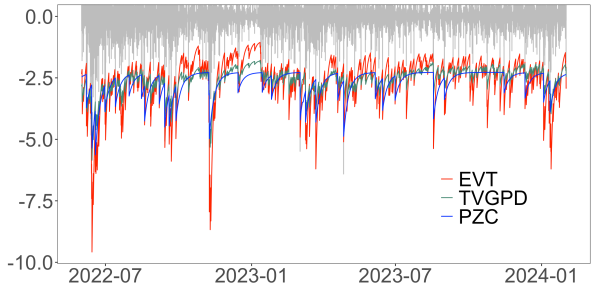
Figure 5: VaR forecasts for BTC/USD

Out-of-sample forecasts of extreme VaR at $\kappa = 10\%$ ($\gamma = 1\%$) and $\kappa = 2.5\%$ ($\gamma = 0.25\%$) for BTC/USD. The EVT and TVGPD methods use PZC estimates of the tail threshold τ_t at a tail probability κ. The EVT, TVGPD, and PZC VaRs in the figure all relate to an extreme tail probability $\gamma = \kappa/10$. Hourly BTC/USD VaR forecasts start with an in-sample period of four years and forecast the VaRs for a year, after which the model is recursively updated annually.

(a) $\gamma = 1\%$ ($\kappa = 10\%$)



(b) $\gamma = 0.25\%$ ($\kappa = 2.5\%$)



model is clearly among the most competitive EVT-based time-varying tail shape models, particularly in extreme tails.

The loss comparisons in Table 2 only provide a partial picture of the full results. To see this, we plot the out-of-sample VaR forecasts for BTC/USD for κ values of 10% and 2.5%, i.e., a moderate and a more extreme tail area. Again, we set $\gamma = \kappa/10$ and plot the VaR over a shorter period to better visualize the results. Figure 5 presents the results. There are two main takeaways. First, we see that all three tail-based models do reasonably well in following the secular dynamics of risk. The TVGPD and EVT behave and thus perform very similarly. Second, we see that farther out into the tails ($\gamma = 0.25\%$), the PZC methodology receives fewer and fewer POTs to estimate the risk dynamics from. This results in incidental spikes, followed by rapid (within one month) reversals to a long-term constant VaR value that we see in the blue line in Figure 5b, and that we have also seen in Figure 1 for simulated data. Such differences are not visible from the scoring rule results alone, but can be important for deciding which model to use for setting adequate capital buffer levels, either from an institutional or supervisory perspective. It also underlines again that both methods nicely complement each other, with the EVT methods complementing the semi-parametric PZC

methodology farther out into the tails.

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, this paper also contributes to 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. 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. The ease of the method's applicability and its implications, particularly in the extreme tail, were illustrated by studying the time variation in the tails of two cryp-

tocurrency exchange rate returns over both quiet and turbulent times. The methodology works particularly well in conjunction with the approach of Patton et al. (2019) to set the tail thresholds, where the EVT approach allows us to extrapolate the risk dynamics further into the tail.

Finally, our out-of-sample evaluation exercise suggests an avenue for future work: combining risk models to obtain an even more comprehensive picture of extreme risk dynamics than any single model can provide in isolation. This idea could involve the combination of extreme VaR and ES forecasts, with weights guided by each model's recent performance under a Nolde and Ziegel (2017) scoring rule. We leave such extensions to future research.

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.

Beutner, E., Y. Lin, and A. Lucas (2024). Consistency, distributional convergence, and optimality of score-driven filters. Technical Report 23-051/III, Tinbergen Institute Discussion Paper.

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

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 (2024). Maximum likelihood estimation for non-stationary location models with mixture of normal distributions. *Journal of Econometrics 238*(1), 105575.

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

Cavaliere, G., T. Mikosch, A. Rahbek, and F. Vilandt (2025). A Comment on: "Autoregressive conditional duration: A new model for irregularly spaced transaction data". *Econometrica 93*(2), 719–729.

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.

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.

Francq, C. and J.-M. Zakoïan (2019). *GARCH Models, 2nd Edition*. John Wiley & Sons.

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.

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.

Nelson, D. B. and D. P. Foster (1994). Asymptotic filtering theory for univariate ARCH models. *Econometrica 62*(1), 1–41.

Nelson, D. B. and D. P. Foster (1995). Filtering and forecasting with misspecified ARCH models II: Making the right forecast with the wrong model. *Journal of Econometrics 67*(2), 303–335.

Newey, W. K. and D. McFadden (1994). Large sample estimation and hypothesis testing. In R. F. Engle and D. L. McFadden (Eds.), *Handbook of Econometrics*, Volume 4, pp. 2111–2245. Amsterdam: Elsevier.

Nolde, N. and J. F. Ziegel (2017). Elicitability and backtesting: Perspectives for banking regulation. *The Annals of Applied Statistics 11*(4), 1833–1874.

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"

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

[a]: *University of Bologna*

[b]: *Vrije Universiteit Amsterdam* and *Tinbergen Institute*

[c]: *European Central Bank, Financial Research*

[d]: *Sveriges Riksbank*

Contents

A Proofs

NIC for the EVT-based VaR

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

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

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

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 $\alpha \ln(\kappa/\gamma) < 1$. 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

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

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 $\ln(1+y_i) = f_i(\theta_0)\epsilon_i$, we can write the score-driven filter as

$$\begin{aligned}
\hat{f}_{i+1}(\theta) =& \omega + \hat{f}_i(\theta) + \alpha\left(\ln(1+y_i) - \hat{f}_i(\theta)\right) \\
=& \omega + \hat{f}_i(\theta) + \alpha\left(f_i(\theta_0)\epsilon_i - \hat{f}_i(\theta)\right) \\
=& \omega + (1-\alpha)\,\hat{f}_i(\theta) + \alpha\epsilon_i f_i(\theta_0).
\end{aligned}$$

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

$$f_{i+1}(\theta_0) = \omega_0 + (1-\alpha_0+\alpha_0\epsilon_i)\,f_i(\theta_0).$$

Note that both $\hat{f}_i(\theta)$ and $f_i(\theta_0)$ are embedded in the stochastic recurrence equations (SREs) of the form $\hat{f}_{i+1}(\theta) = \hat{\phi}(\hat{f}_i(\theta), y_i, \theta)$ and $f_{i+1}(\theta_0) = \phi(f_i(\theta_0), \epsilon_i, \theta_0)$, respectively.

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

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

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

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

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 $f_i(\theta_0)$ exists and that any initialized sequence converges exponentially fast almost surely (*e.a.s.*) to this unique SE limit. Given $y_i=\exp(f_i(\theta_0)\epsilon_i)-1$, it follows immediately that y_i 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 $f_i(\theta_0)$ equals

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

Note that $\mathbb{E}\left[(1+\alpha_0(\epsilon_i-1))^q\right]<\infty$ for any finite $q>0$ given that ϵ_i 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 $\mathbb{E}[(1+\alpha_0(\epsilon_i-1))^r]=\eta$ and thus

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

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

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

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

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

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 $\ln(1+y_i)$ and is automatically satisfied via part (i) of the theorem if the model is correctly specified.

To establish the contraction property, note that

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

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

This establishes the second part of the theorem.

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

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

It then follows directly from Lemma 2.1 of Straumann and Mikosch (2006) that $\hat{z}_i^f(\theta) \xrightarrow{e.a.s.} z_i^f(\theta)$ uniformly over $\theta \in \Theta$ if $\mathbb{E}[\ln^+ f_i(\theta_0)] < \infty$. 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 $1 - \alpha + \alpha\epsilon_i$ for *IID* unit exponential ϵ_i, 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 $\{\hat{\mathcal{Q}}_{n_T}(\theta)\}_{n\in\mathbb{N}}$ as defined in (9). Specifically: (i) The parameter space Θ is compact; (ii) $\{\hat{\mathcal{Q}}_{n_T}(\theta)\}_{n\in\mathbb{N}}$ is a sequence of random functions continuous on Θ almost surely; (iii) $\hat{\mathcal{Q}}_{n_T}(\theta) = n^{-1}\sum_{i=1}^{n_T}\hat{Q}_i(\theta) \to \bar{\mathcal{Q}}(\theta) := \mathbb{E}\left[Q_i(\theta)\right]$ as $n_T \to \infty$ 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 $\{\hat{z}_i^f(\theta)\}_{i\in\mathbb{Z}}$ and $\{z_i^f(\theta)\}_{i\in\mathbb{Z}}$. Furthermore, from Theorem 1 we obtain that $\mathbb{E}\left[\sup_{\theta\in\Theta}\left|\hat{Q}_i(\theta)\right|\right] < \infty$ and $\mathbb{E}\left[\sup_{\theta\in\Theta}\left|Q_i(\theta)\right|\right] < \infty$. Theorem 1 also ensures that the process $\{\hat{z}_i^f(\theta)\}_{i\in\mathbb{Z}}$ converges e.a.s. to its stationary and ergodic limit $\{z_i^f(\theta)\}_{i\in\mathbb{Z}}$. We thus have

$$\sup_{\theta\in\Theta}\left|\hat{Q}_i(\theta)-Q_i(\theta)\right| \leq \sup_{\theta\in\Theta}\left|\ln\hat{z}_i^f(\theta)-\ln z_i^f(\theta)\right| - \epsilon_i \cdot \sup_{\theta\in\Theta}\left|\hat{z}_i^f(\theta)-z_i^f(\theta)\right|.$$

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

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

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

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

It thus follows that

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

with $\mathbb{E}\left[\sup_{\theta\in\Theta}\left|Q_i(\theta)\right|\right]<\infty$. Let $\mathcal{Q}_{n_T}(\theta)=\sum_{i=1}^{n_T}\ln z_i^f(\theta)-\epsilon_i\left(z_i^f(\theta)-1\right)$ be the SE limit of $\hat{\mathcal{Q}}_{n_T}(\theta)$. Now, from the triangle inequality

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

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

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

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

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

$\log(z) - z + 1 \leq 0$ for any $z \in \mathbb{R}^+$, with equality only for $z = 1$. Note that $z_i^f(\theta_0) = 1$. This in turn implies that $\bar{\mathcal{Q}}(\theta_0) = 0$. We conclude the consistency proof by showing that if $z_i^f(\theta) = z_i^f(\theta_0) = 1$ almost surely for every i, 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 $f_i(\theta) = f_i(\theta_0)$. Therefore, let $f_i(\theta) = f_i(\theta_0)$ almost surely for every i. We then have that

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

almost surely. Obviously, from Assumption 1, ϵ_i is an $\mathcal{F}_i$-measurable random variable with a non-degenerate distribution, and from Theorem 1 $f_i(\theta_0)$ 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 $\hat{\theta}_{n_T}$ 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 $\hat{\theta}_{n_T}$, we obtain that, for large enough n_T the following Taylor expansion is allowed:

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

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

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

To prove the asymptotic normality of the MLE $\hat{\theta}_{n_T}$ 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 $\hat{\mathcal{Q}}_{n_T}(\theta) : \Omega \times \Theta \mapsto \mathbb{R}$ is continuously differentiable of order 2 on Θ almost surely, (iii) The MLE $\hat{\theta}_{n_T} : \Omega \mapsto \Theta$ is $\mathcal{F}$-measurable and strongly consistent, i.e. $\hat{\theta}_{n_T} \overset{a.s.}{\to} \theta_0$ where $\theta_0 \in int(\Theta)$; (iv) the score vector satisfies $n_T^{-1/2}\nabla^{\theta}\mathcal{Q}_{n_T}(\theta_0) \Rightarrow \mathcal{N}\left(\mathbf{0}_2, \mathbb{E}\left[\nabla^{\theta}Q_i(\theta_0)\nabla^{\theta}Q_i(\theta_0)^{\top}\right]\right)$; (v) the uniform stochastic convergence of the Hessian matrix, that is, $\sup_{\theta\in\Theta}\left\|\frac{1}{n_T}\nabla^{\theta\theta}\hat{\mathcal{Q}}_{n_T}(\theta) - \nabla^{\theta\theta}\bar{\mathcal{Q}}(\theta)\right\| \overset{a.s.}{\to} 0$, where $\nabla^{\theta\theta}\bar{\mathcal{Q}}(\theta) = \mathbb{E}\left[\nabla^{\theta\theta}Q_i(\theta)\right]$ is finite; (vi) the limit $\nabla^{\theta\theta}\bar{\mathcal{Q}}(\theta)$ evaluated at the true parameter vector θ_0 satisfies $-\bar{\mathcal{Q}}(\theta_0) = -\mathbb{E}\left[\nabla^{\theta\theta}Q_i(\theta_0)\right] = \boldsymbol{\mathcal{I}}(\theta_0)$, 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 $\{\nabla^{\theta}Q_i(\theta_0)\}_{i\in\mathbb{N}}$ is a stationary and ergodic zero-mean martingale difference process with respect to the filtration $\{\mathcal{F}_i\}_{i\in\mathbb{N}}$ with $\mathcal{F}_i = \sigma\{\epsilon_i, \epsilon_{i-1}, \epsilon_{i-2}, \dots\}$. In fact, note that

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$\begin{aligned}\nabla^{\theta}z_{i+1}^f(\theta) &= \nabla^{\theta}\left(\frac{f_{i+1}(\theta_0)}{f_{i+1}(\theta)}\right) = -\frac{f_{i+1}(\theta_0)}{f_{i+1}(\theta)^2}\nabla^{\theta}f_{i+1}(\theta)\\ &= -\frac{f_{i+1}(\theta_0)}{f_{i+1}(\theta)^2}\left(\nabla^{\theta}\omega+\left(\nabla^{\theta}\alpha\right)\left(f_i(\theta_0)\epsilon_i-f_i(\theta)\right)+(1-\alpha)\nabla^{\theta}f_i(\theta)\right)\\ &= \left(\frac{f_{i+1}(\theta_0)}{f_{i+1}(\theta)}\right)^2\left(\frac{f_i(\theta_0)}{f_{i+1}(\theta_0)}\right)\left(w_t(\theta)+(1-\alpha)\nabla^{\theta}z_i^f(\theta)\right), \qquad \text{(A.8)}\\ w_t(\theta) &= \left(-1/f_i(\theta_0)\quad,\quad(1/z_i^f(\theta))-\epsilon_i\right)^\top,\end{aligned}$$

where $f_i(\theta_0)/f_{i+1}(\theta_0) = 1/(1+\alpha_0(\epsilon_i-1)+\omega_0/f_i(\theta_0))$. Since $\{\nabla^{\theta}z_i(\theta_0)\}_{i\in\mathbb{Z}}$ 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 $\boldsymbol{\lambda}^\top\nabla^{\theta}z_i^f(\theta_0)=\mathbf{0}_2$ almost surely $\forall i\in\mathbb{N}$. This is obviously ruled out by the functional form of (A.8) and the unit exponential distributional form of ϵ_i and, therefore, it must be that $\boldsymbol{\lambda}^\top\nabla^{\theta}z_i^f(\theta_0)=\mathbf{0}_2 \iff \boldsymbol{\lambda}=\mathbf{0}_2$ and thus, $\nabla^{\theta\theta}\bar{\mathcal{Q}}(\theta_0)$ is non-singular. In conclusion, we note that the Fisher's information equality $\mathbb{E}\left[\nabla^{\theta}Q_i(\theta_0)\nabla^{\theta}Q_i(\theta_0)^\top\right]=-\mathbb{E}\left[\nabla^{\theta\theta}Q_i(\theta_0)\right]=\boldsymbol{\mathcal{I}}(\theta_0)$ follows by standard arguments.

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 $\nabla^{\theta} Q_i(\theta) = \left(\nabla^{\omega} Q_i(\theta), \nabla^{\alpha} Q_i(\theta)\right)^\top \in \mathbb{R}^2$, and the Hessian matrix

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

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

$$\nabla^{\theta} Q_i(\theta) = \left(z_i^f(\theta)^{-1} - \epsilon_i\right) \nabla^{\theta} z_i^f(\theta) = \left(\epsilon_i - z_i^{f,\text{inv}}(\theta)\right) \frac{\nabla^{\theta} z_i^{f,\text{inv}}(\theta)}{z_i^{f,\text{inv}}(\theta)^2}, \tag{B.1}$$

and

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

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

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

and

$$\nabla^{\theta} f_{i+1}(\theta) = \phi_i^{\theta}\left(f_i(\theta), \nabla^{\theta} f_i(\theta), \epsilon_i, \theta\right) = \begin{pmatrix} 1 \\ f_i(\theta_0)\epsilon_i - f_i(\theta) \end{pmatrix} + (1-\alpha)\,\nabla^{\theta} f_i(\theta). \tag{B.4}$$

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

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

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

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

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

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

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

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

Proof of Lemma B.1

We note that $\nabla^\theta \hat{f}_{i+1}(\theta)$ is a function of both the filter $\hat{f}_i(\theta)$ and its derivative $\nabla^\theta \hat{f}_i(\theta)$. 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 $\phi_i^\theta\left(\hat{f}_i(\theta), \nabla^\theta \hat{f}_i(\theta), \epsilon_i, \theta\right)$ 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

$$\sup_{\theta\in\Theta}\left\|\phi_i^\theta\left(\hat{f}_i(\theta), \nabla^\theta \bar{f}_1(\theta), \epsilon_i, \theta\right) - \phi_i^\theta\left(f_i(\theta), \nabla^\theta \bar{f}_1(\theta), \epsilon_i, \theta\right)\right\| \xrightarrow{e.a.s.} 0,$$

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

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

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

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

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

for the derivatives of the ratio process $\{z_i^{f,\text{inv}}(\theta)\}_{i\in\mathbb{Z}}$, i.e., $\{\nabla^\theta z_i^{f,\text{inv}}(\theta)\}_{i\in\mathbb{Z}}$ and $\{\nabla^{\theta\theta} z_i^{f,\text{inv}}(\theta)\}_{i\in\mathbb{Z}}$. 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 $\{\nabla^\theta z_i^{f,\text{inv}}(\theta)\}_{i\in\mathbb{Z}}$ and $\{\nabla^{\theta\theta} z_i^{f,\text{inv}}(\theta)\}_{i\in\mathbb{Z}}$ have k uniformly bounded moments $\forall k > 0$, that is

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

Proof of Lemma B.2

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

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

so that, for i sufficiently large, we get

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

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

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

strict stationarity of $\{f_i(\theta)\}_{i\in\mathbb{Z}}$, we have

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

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

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

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

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

and, using $\epsilon_i f_i(\theta_0) \leq \alpha_0^{-1} f_{i+1}(\theta_0)$,

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

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 $\{\nabla^{\theta\theta} z_i^{f,\text{inv}}(\theta)\}_{i\in\mathbb{Z}}$. ∎

Lemma B.3. Under the conditions of Theorem 1, the Hessian processes $\{\nabla^{\theta\theta} z_i^{f,\text{inv}}(\theta)\}_{i\in\mathbb{Z}}$

has a uniformly bounded moment, that is

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

Proof of Lemma B.3

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

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

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

C Derivation of market risk measures

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

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

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

$$\begin{aligned}&\overline{G}(X_{t_i}) = \overline{G}(\tau_{t_i})\overline{F}(y_i) = \gamma \\ \iff &\frac{n_{t_i}}{t_i}(1 + y_i)^{-1/f_i} = \gamma \\ \iff &1 + \tau_{t_i}^{-1}(q_{t_i}^{1-\gamma}(X_{t_i}) - \tau_{t_i}) = \left(\frac{\gamma}{n_{t_i}/t_i}\right)^{-f_i} \\ \iff &q_{t_i}^{1-\gamma}(X_{t_i}) = \tau_{t_i}\left(\frac{\gamma}{n_{t_i}/t_i}\right)^{-f_i},\end{aligned} \tag{C.1}$$

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 $\mathrm{ES}^{1-\gamma}(X_t)$ is given by

$$\begin{aligned}\mathrm{ES}^{1-\gamma}(X_t) &= \frac{1}{\gamma}\int_{1-\gamma}^{1} q_t^s(X_t)\mathrm{d}s \\ &= \frac{\mathrm{VaR}^{1-\gamma}(X_t \mid \mathcal{F}_{t-1}, \theta)}{1 - f_i},\end{aligned} \tag{C.2}$$

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

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

for $f_i < 1$.

For completeness, we note that in D'Innocenzo et al. (2024) the market risk measures are given by

$$VaR^{1-\gamma}(X_{t_i}) = q_{t_i}^{1-\gamma}(X_{t_i}) = \tau_{t_i} + \frac{\delta_{t_i}}{f_i}\left[\left(\frac{\gamma}{n_{t_i}/t_i}\right)^{-f_i} - 1\right], \tag{C.3}$$

$$ES^{1-\gamma}(X_{t_i}) = \frac{\mathrm{VaR}^{1-\gamma}(X_{t_i} \mid \mathcal{F}_{t-1}, \theta)}{1 - f_i} + \frac{\delta_{t_i} - f_i\tau_{t_i}}{1 - f_i}, \tag{C.4}$$

for a tail-scale parameter δ_t. It is easily verified that these expressions collapse to (C.1) and (C.2) if we set $\delta_{t_i} = f_i\tau_{t_i}$ in line with the limiting result of de Haan and Ferreira (2006).

References in Online Appendix

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

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.