Modeling maxima with autoregressive conditional Fréchet model[☆]

Zifeng Zhao [a], Zhengjun Zhang [b,*], Rong Chen [c]

[a] *Department of Information Technology, Analytics and Operations, Mendoza College of Business, University of Notre Dame, United States*
[b] *Department of Statistics, University of Wisconsin-Madison, United States*
[c] *Department of Statistics and Biostatistics, Rutgers University, United States*

ARTICLE INFO

ABSTRACT

This paper introduces a novel dynamic generalized extreme value (GEV) framework for modeling the time-varying behavior of maxima in financial time series. Specifically, an autoregressive conditional Fréchet (AcF) model is proposed in which the maxima are modeled by a Fréchet distribution with time-varying scale parameter (volatility) and shape parameter (tail index) conditioned on past information. The AcF provides a direct and accurate modeling of the time-varying behavior of maxima and furthermore offers a new angle to study the tail risk dynamics in financial markets. Probabilistic properties of AcF are studied, and a maximum likelihood estimator is used for model estimation, with its statistical properties investigated. Simulations show the flexibility of AcF and confirm the reliability of its estimators. Two real data examples on cross-sectional stock returns and high-frequency foreign exchange returns are used to demonstrate the AcF modeling approach, where significant improvement over the static GEV has been observed for market tail risk monitoring and conditional VaR estimation. Empirical result of AcF is consistent with the findings of the dynamic peak-over-threshold (POT) literature that the tail index of financial markets varies through time.

© 2018 Elsevier B.V. All rights reserved.

1. Introduction

The study of extreme events in financial markets is always one of the main foci in risk management. Maximum observations, as the representation of extreme behavior, are of particular interest. For example, mutual fund managers are keen to assess the potential maximum daily loss across all stocks in their managed portfolio; the level of potential intra-day maximum loss is important to high-frequency traders. By Fisher–Tippett–Gnedenko theorem, the generalized extreme value distribution (GEV) can be used to characterize the behavior of maxima, making extreme value theory (EVT) a widely researched and practiced approach for risk management in financial industry (e.g. Embrechts et al., 1999; McNeil and Frey, 2000; Laurini and Tawn, 2009). Besides the Maxima-GEV methodology, the other fundamental methodology of EVT is the peak-over-threshold (POT), which is based on generalized Pareto distribution (GPD). By Pickands–Balkema–de

---

Haan theorem, GPD can be used to approximate the conditional behavior of random variable after it exceeds certain high thresholds (e.g. Balkema and de Haan, 1974; Picklands, 1975; Davison and Smith, 1990). Under EVT framework, Maxima-GEV and POT-GPD are closely related and can often reveal the same information, especially when used to model tail index ξ. See Chapters 4 and 7 in Coles (2001) and Section 2.5 in this paper for more details about the connection between Maxima-GEV and POT-GPD.

As mentioned in Diebold et al. (1998), most applications of EVT focus on modeling extreme events in time series with a static approach under equilibrium distribution. However, the behavior of the underlying time series may change through time. For example, financial time series tends to exhibit structural changes and time-varying dynamics such as volatility clustering. To accommodate the dynamics of extreme events and study the conditional behavior of tail risk in financial markets, there have been several recent studies of dynamic POT-GPD models. For example, Smith and Goodman (2000) and Chavez-Demoulin et al. (2014) use Bayesian method to update the time-varying GPD parameters. Kelly (2014) and Kelly and Jiang (2014) build a dynamic tail model with POT-GPD for panel data. Massacci (2016) and Zhang and Schwaab (2017) employ a generalized autoregressive score[1] type of observation-driven dynamics for the GPD parameters. These studies show strong evidence of the time-varying behavior of extreme events in financial markets, especially for the tail index ξ.

One advantage of Maxima-GEV approach over POT-GPD approach is that it offers a direct modeling of maxima in time series, which can be of particular importance. Unlike the dynamic GPD models, there is little research on dynamic GEV models. Bali and Weinbaum (2007) design a time-varying GEV to estimate the realized volatility in an empirical study of market risk, however, theoretical results are not provided.

In this paper we are mainly interested in modeling time series of maxima $\{Q_t\}$, where $\{Q_t\}$ is a univariate time series of maxima based on a set of underlying financial time series $\{X_{it}\}_{i=1}^{p}$. There are mainly two types of $\{Q_t\}$. The first type is the time series of cross-sectional maxima, where $\{X_{it}\}_{i=1}^{p}$ are a set of panel time series, and we are interested in modeling the cross-sectional maxima $Q_t = \max_{1 \le i \le p} X_{it}$. Such problems arise in many applications, including modeling the maximum daily loss across a group of stocks in a portfolio. The second type is the time series of intra-period maxima, where $\{X_{it}\}_{i=1}^{p}$ denote the p intra-period observations for a univariate time series within period t, and we are interested in modeling the intra-period maxima $Q_t = \max_{1 \le i \le p} X_{it}$. For example, one may be interested in the intra-day maxima of high-frequency trading losses that occur on the same day.

It is worth noting that there is an important difference between "maxima" and "extreme event". "Maxima" has a clear definition of being the maximum of a set of observations, while "extreme event" is a more vague term, typically defined as rare observations over a high threshold. A "maxima" may not necessarily be an "extreme event" though most likely it is. As a time series of maxima, $\{Q_t\}$ is observed at every t. On the other hand, an extreme event may not be observed at each t, or there may be several extreme events within a time period. There is ample research on extremal process that offers stochastic characterization and modeling of "extreme event" in stationary process (see Resnick, 1987; Basrak et al., 2002; Basrak and Segers, 2009; Drees et al., 2015, for more details). However, in this paper the focus is the direct modeling of the maxima $\{Q_t\}$ process and its time-varying behavior.

An important byproduct from dynamic modeling of maxima $\{Q_t\}$ is the tail index ξ_t, which is arguably the most important indicator for financial market tail risk. As shown later, the tail index of maxima $\{Q_t\}$ corresponds to the tail index of the underlying time series $\{X_{it}\}_{i=1}^{p}$. Thus, a better modeling of maxima can help obtain a more accurate assessment of the current market risk level and offer more insight into the potential market extreme movement.

With the aim of modeling the time-varying behavior of maxima and tail risk in the financial market, in this paper we introduce a novel dynamic conditional GEV framework, in which parameters (μ, σ, ξ) of a conditional GEV are allowed to vary through time with a GARCH-like[2] autoregressive mechanism. Due to the heavy-tailed nature of financial data, Fréchet distribution is widely used for modeling maxima originated from financial time series. Thus, we propose an autoregressive conditional Fréchet (AcF) model that allows for an observation-driven time evolution of the scale parameter σ and the tail index ξ of a Fréchet (Type-II GEV) distribution. Since the scale parameter and the tail index play the key role in characterizing the tail behavior of Fréchet distribution, AcF provides a more flexible and applicable model for the time-varying behavior of maxima in financial time series.

The main contributions of this paper are twofold. From a statistical point of view, this paper provides the first complete treatment of a dynamic GEV model. The AcF is a direct approach to modeling the time-varying behavior of maxima in financial time series. Probabilistic properties of the model and statistical properties of its estimator are investigated and developed. They make the paper theoretically sound and help lay the foundation for further development of dynamic models under EVT framework. From an econometric point of view, the newly designed AcF offers a new angle to study the time-varying behavior of tail risk in financial markets and serves as a promising alternative to the dynamic POT-GPD methodology in the literature. Real data applications show that the tail index of financial market is indeed time-varying. Compared to the static GEV, AcF captures the dynamics of maxima more adequately and offers more promising performance in detecting potential market extreme movement and providing more accurate conditional VaR prediction for maxima.

The rest of the paper is organized as follows. In Section 2 we present a detailed description of AcF and investigate its probabilistic properties. A maximum likelihood estimator (MLE) is used for estimation and its statistical properties as an irregular MLE are derived in Section 3. Simulation studies are presented in Section 4 to demonstrate the flexibility and

---

[1] See Harvey and Chakravarty (2008) and Creal et al. (2013) for more details about generalized autoregressive score model.

[2] See Engle (1982) and Bollerslev (1986) for more details about GARCH model.
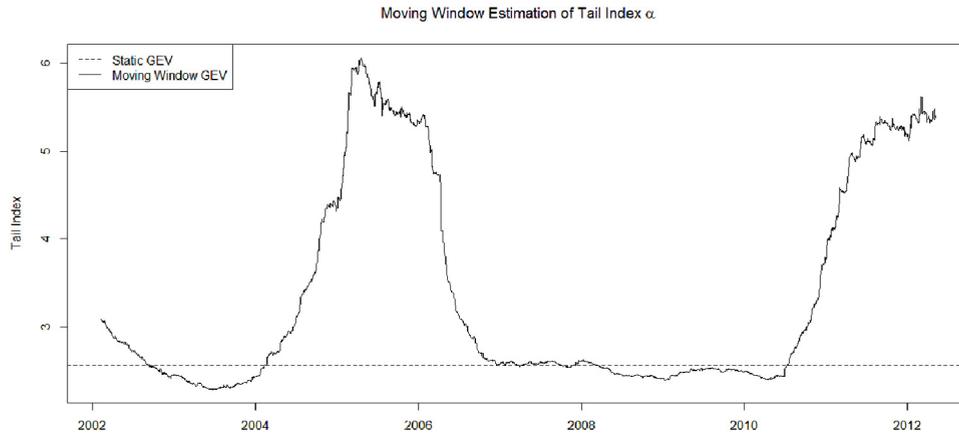
Fig. 1. Tail index $\hat{\alpha}_t$ estimated by moving window of size 1000 (solid curve) v.s. tail index $\hat{\alpha}$ estimated by the static GEV model based on total observations (dashed line).

robustness of AcF and to evaluate the performance of MLE. Section 5 presents two real data applications, one on market tail risk monitoring and tail connectedness based on cross-sectional maximum loss of stock markets and one on conditional VaR estimation of intra-day maximum loss from high-frequency foreign exchange trading. We conclude the paper in Section 6. The Appendix contains the proofs of the theorems and other technical materials.

2. Autoregressive conditional Fréchet model

2.1. Motivation

As a time series of maxima, $\{Q_t\}$ cannot be directly modeled by conventional time series models like ARMA or GARCH. By Fisher–Tippett–Gnedenko theorem, we know that under certain condition, marginally $\{Q_t\}$ can be accurately approximated by a GEV distribution with three parameters (μ, σ, ξ), the location, the scale, and the shape parameter, respectively. The common practice in the literature of modeling maxima $\{Q_t\}$ is to treat it as *i.i.d.* data and model it by a GEV distribution. The obvious limitation is that the time dependency between $\{Q_t\}$ has been completely ignored in this approach, which can potentially cause a huge loss in model efficiency if $\{Q_t\}$ has a strong dependency across time. To overcome this drawback, we propose a dynamic GEV framework, under which a conditional evolution scheme is designed for the parameters (μ_t, σ_t, ξ_t) of GEV, so that time dependency of $\{Q_t\}$ can be captured. Due to the heavy-tailedness of financial data, $\{Q_t\}$ marginally can be well modeled by Fréchet distribution (i.e. Type II GEV distribution), which corresponds to the GEV of which $\xi > 0$. For the rest of the paper, we focus on the case that conditionally Q_t can be modeled by a Fréchet distribution with parameters $(\mu_t, \sigma_t, \alpha_t)$ where $\alpha_t = 1/\xi_t$ as often used for the parametrization of Fréchet distribution:

$$Q_t = \mu_t + \sigma_t Y_t^{1/\alpha_t}, \tag{1}$$

where $\{Y_t\}$ is a sequence of *i.i.d.* unit Fréchet random variables and $(\mu_t, \sigma_t, \alpha_t) \in \mathcal{F}_{-1} = \sigma(Q_{-1}, Q_{-2}, \ldots)$. Here μ_t and σ_t are location-scale parameters and α_t is the shape parameter, also called tail index of Fréchet distribution. Note that the support of Q_t is (μ_t, ∞) since $Y_t > 0$.

The scale parameter σ_t should not be taken exactly as volatility since the conditional variance of Q_t depends on both σ_t and α_t. However, as shown later in Proposition 1, σ_t can be closely related to the volatility process of the underlying time series $\{X_{it}\}_{i=1}^p$ and thus requires a dynamic treatment due to the well-known volatility clustering in financial time series. Moreover, Harvey (2013) observes that if volatility clustering is not accounted for, movements of the tail parameter α_t can be potentially confounded with movements of the scale parameter σ_t.

The tail index α_t is the essential parameter since it governs the underlying tail behavior of $\{Q_t\}$ process and plays the most important role in quantifying the potential tail risk. To demonstrate the necessity of dynamic modeling of α_t, we perform an ad-hoc moving-window GEV analysis on the cross-sectional maxima of negative daily log-returns (i.e. daily losses) of the component stocks in S&P100 index which includes 100 leading U.S. stocks. The observation period is from January 1, 2000 to December 31, 2014 with 3773 trading days. For each trading day t, we record the maximum daily loss across the 100 stocks and denote it by Q_t. Hence the time series $\{Q_t\}$ has 3773 observations. For each t such that $500 \leq t \leq 3273$, a GEV model is fitted using $\{Q_k\}_{k=t-499}^{t+500}$, the observations within a 1000-day (approximately 4 years) local window centered at t. We plot the estimated tail index $\hat{\alpha}_t$ in Fig. 1 along with the estimated tail index $\hat{\alpha}$ obtained by directly fitting the static GEV model with the entire series, treating them as *i.i.d.* observations.

It can be clearly seen that compared to the static estimation (dashed line), the tail index estimated with smaller moving window (solid curve) changes quite drastically throughout the years, indicating an insufficiency of the static GEV model.

A similar finding of varying tail index has been reported in Kelly and Jiang (2014) and Massacci (2016). An interesting phenomenon is that the static GEV seems to give a significantly lower estimation of the tail index than the moving-window GEV. An under-estimation of tail index over-estimates the tail risk, which in turn may result in higher reserve requirements and other expenses for financial institutions.

2.2. Model specification

For parsimony, we set μ_t to be constant, which is the common practice in the extreme value analysis, and concentrate on the dynamics of σ_t and α_t, which are the key parameters of modeling tail behavior. We impose an autoregressive structure on the time-varying parameters (σ_t, α_t) similar to the approach of GARCH model in Bollerslev (1986), autoregressive conditional density model (ACD) in Hansen (1994), and autoregressive conditional duration model in Engle and Russell (1998).

Specifically, the autoregressive conditional Fréchet (AcF) model assumes the form

$$\log \sigma_t = \beta_0 + \beta_1 \log \sigma_{t-1} + \eta_1(Q_{t-1}), \tag{2}$$

$$\log \alpha_t = \gamma_0 + \gamma_1 \log \alpha_{t-1} + \eta_2(Q_{t-1}), \tag{3}$$

where $\beta_1, \gamma_1 \geq 0$ and the two terms $\eta_1(\cdot)$ and $\eta_2(\cdot)$ are the observation-driven factors for $\{\log \sigma_t\}$ and $\{\log \alpha_t\}$. The log transform is used to ensure the positivity of the parameters.

We further assume that $\eta_1(\cdot)$ is a continuous increasing function and $\eta_2(\cdot)$ is a continuous decreasing function of Q_{t-1}. One salient feature of the maxima series $\{Q_t\}$ in many applications, especially in financial time series, is the clustering of extreme events in time. It has been observed that large-valued maxima tend to happen around the same period in many applications. One possible explanation is that an extreme event observed at time $t-1$ (large Q_{t-1}) causes the distribution of Q_t to have larger scale (large σ_t) and heavier tail (small α_t), resulting in a larger tail risk of Q_t. An increasing $\eta_1(\cdot)$ and a decreasing $\eta_2(\cdot)$ ensure that larger Q_{t-1} is followed by larger σ_t and smaller α_t. Together with the autoregressive scheme of $\{\log \sigma_t\}$ and $\{\log \alpha_t\}$ (i.e. $\gamma_1, \beta_1 \geq 0$), this evolution dynamics offers a joint modeling of both volatility clustering for $\{\sigma_t\}$ process and heavy-tail clustering for $\{\alpha_t\}$ process.

There are many choices of the continuous monotone functions $\eta_1(\cdot)$ and $\eta_2(\cdot)$. In this paper we use the simple exponential function $a_0 \exp(-a_1 x)$. It is a simplified version of the widely used logistic function $\frac{L}{1 + a_0 \exp(-a_1 x)}$. Due to its monotonicity, differentiability, and boundedness, the logistic function is employed in many studies of observation-driven time series models (e.g. Hansen, 1994; Lundbergh et al., 2003; Boutahar et al., 2008; Hall et al., 2016). The simplification here is due to $Q_t > \mu$, hence there is no need to have the 1 in the denominator of the logistic function for boundedness. We set $a_1 > 0$ to ensure boundedness and let the sign of a_0 control monotonicity of the exponential function.

For the rest of the paper, we consider the following model:

$$Q_t = \mu + \sigma_t Y_t^{1/\alpha_t}, \tag{4}$$

$$\log \sigma_t = \beta_0 + \beta_1 \log \sigma_{t-1} - \beta_2 \exp(-\beta_3 Q_{t-1}), \tag{5}$$

$$\log \alpha_t = \gamma_0 + \gamma_1 \log \alpha_{t-1} + \gamma_2 \exp(-\gamma_3 Q_{t-1}), \tag{6}$$

where $\{Y_t\}$ is a sequence of *i.i.d.* unit Fréchet random variables, $0 \leq \beta_1 \neq \gamma_1 < 1$, $\beta_2 > 0$, $\beta_3 > 0$, $\gamma_2 > 0$, and $\gamma_3 > 0$.

One alternative for $\eta_1(\cdot)$ and $\eta_2(\cdot)$ is the widely used generalized autoregressive score (GAS) models by Harvey and Chakravarty (2008) and Creal et al. (2013), which has been successfully employed in the literature of dynamic POT-GPD models, see Massacci (2016) and Zhang and Schwaab (2017). However, as shown in A.5, in our dynamic GEV context, the $\eta_1(\cdot)$ and $\eta_2(\cdot)$ implied under the GAS framework are complicated and may not be monotone, and thus may suffer from lack of interpretability.

The exponential function is simple, flexible and at the same time intuitive and interpretable. Although the exponential function implies an upper bound for the $\{\sigma_t\}$ and $\{\alpha_t\}$ process,[3] as demonstrated in Hansen (1994) for logistic functions, the boundedness does not affect the flexibility of the model but facilitates numerical and technical tractability. Moreover, as shown in Section 5, the current model can flexibly capture dynamics of both the scale parameter and the tail index, and offers an accurate modeling of the maxima in financial time series. Extensions of AcF to allowing μ_t to vary more freely can be implemented by imposing an ARMA structure on μ_t with added complexity and potential model instability. A further justification of AcF is given in Section 2.4 under a general factor model setting.

To our best knowledge, this is the first formal presentation of dynamic GEV model that offers a complete dynamic treatment for both the scale parameter σ_t and the tail index α_t. In contrast to the static GEV, AcF is a time series model of the *conditional* maxima. Given all the past information $\mathcal{F}_{t-1}$, the conditional distribution of maxima Q_t is Fréchet$(\mu, \sigma_t, \alpha_t)$, where $(\mu, \sigma_t, \alpha_t) \in \mathcal{F}_{t-1}$.

Remark 1. AcF can be easily extended to include q_1 autoregressive terms of $\log \sigma_t$ and $\log \alpha_t$, and q_2 lagged terms of $\eta(Q_t)$, similar to that of GARCH(q_1, q_2) model. Similar theoretical properties can be derived and similar estimation procedure can be used. Our empirical experience shows that the extension does not necessarily improve the performance of the model,

---

[3] As shown in A.5, the $\eta_1(\cdot)$ and $\eta_2(\cdot)$ implied by GAS also give an upper bound on the $\{\sigma_t\}$ and $\{\alpha_t\}$ process.

but instead induces instability in estimation. A similar phenomenon has been observed in Creal et al. (2013) and Zhang and Schwaab (2017) for POT-GPD model. In this paper we focus on AcF(1,1) model.

Remark 2. The choice of $\eta_1(\cdot)$ and $\eta_2(\cdot)$ may require further consideration when the model is used for other applications, as the exponential function used here is designed to accommodate volatility clustering and heavy tail clustering for financial applications. An increasing $\eta_1(\cdot)$ produces volatility clustering and a decreasing $\eta_2(\cdot)$ produces heavy-tail clustering. However, if it is observed that an extreme event tends to be followed by a period of 'normal' activities, then $\eta_2(\cdot)$ may be assumed to be an increasing function. In general, as long as $\eta_1(\cdot)$ and $\eta_2(\cdot)$ are continuous bounded functions, the probabilistic properties shown below still hold and the same estimation procedure can be applied.

Remark 3. We have assumed that the conditional distribution of Q_t is of Fréchet type since the main focus here is on financial applications. It can be extended to other types of GEVs. In some cases a proper transformation can be used. For example, if a random variable X follows Gumbel(μ, σ), then $\exp(X)$ is Fréchet with location parameter 0, scale parameter $\exp(\mu)$ and tail index $1/\sigma$. Hence, if $\{X_{it}\}_{i=1}^{p}$ are in the Domain of Attraction of Gumbel distribution (Type I GEV), an exponential transformation of the data can be modeled with AcF.

2.3. Stationarity and ergodicity

The evolution schemes (5) and (6) can be written as

$$\log \sigma_t = \beta_0 + \beta_1 \log \sigma_{t-1} - \beta_2 \exp(-\beta_3(\mu + \sigma_{t-1} Y_{t-1}^{1/\alpha_{t-1}})),$$
$$\log \alpha_t = \gamma_0 + \gamma_1 \log \alpha_{t-1} + \gamma_2 \exp(-\gamma_3(\mu + \sigma_{t-1} Y_{t-1}^{1/\alpha_{t-1}})),$$

where $\{Y_t\}$ is an *i.i.d.* sequence of unit Fréchet random variables. Hence $\{\sigma_t, \alpha_t\}$ form a homogeneous Markov chain in $\mathbb{R}^2$. The following theorem gives a general sufficient condition under which the process is stationary and ergodic.

Theorem 1. *For an AcF with $\beta_2, \beta_3, \gamma_2, \gamma_3 > 0$, $\beta_0, \gamma_0, \mu \in \mathbb{R}$, and $0 \le \beta_1 \ne \gamma_1 < 1$, the latent process $\{\sigma_t, \alpha_t\}$ is stationary and geometrically ergodic.*

The proof can be found in A.1. Since $\{Q_t\}$ is a coupled process of $\{\sigma_t, \alpha_t\}$ through (4), $\{Q_t\}$ is also stationary and ergodic. Unlike GARCH model, the stationarity of AcF mainly requires the autoregressive coefficient $0 \le \beta_1 \ne \gamma_1 < 1$ with no restriction on the parameters associated with the shock process Q_{t-1}. This is due to the boundedness of $\eta_1(Q_{t-1})$ and $\eta_2(Q_{t-1})$.

2.4. AcF under a factor model setting

In this section, we illustrate that the limiting form of maxima Q_t under a general factor model framework leads to an AcF model. Assume $\{X_{it}\}_{i=1}^{p}$ follow a general factor model,

$$X_{it} = f_i(Z_{1t}, Z_{2t}, \ldots, Z_{dt}) + \sigma_{it} \varepsilon_{it},$$

where $\{X_{it}\}_{i=1}^{p}$ are observed time series at time t, $\{Z_{1t}, Z_{2t}, \ldots, Z_{dt}\}$ consist of observed and unobserved factors, $\{\varepsilon_{it}\}_{i=1}^{p}$ are *i.i.d.* random noises that are independent with the factors $\{Z_{it}\}_{i=1}^{d}$, and $\{\sigma_{it}\}_{i=1}^{p} \in \mathcal{F}_{t-1}$ are the conditional volatilities of $\{X_{it}\}_{i=1}^{p}$. The function $f_i : \mathbb{R}^d \to \mathbb{R}$ is a Borel function.

This general factor model has been widely used for analyzing high dimensional panel data. The (dynamic) factor models of Bai and Ng (2002), Geweke (1977), Stock and Watson (2011), Lam and Yao (2012), and many others assume unobservable factors. Asset pricing models of Sharpe (1964), Mossin (1966), Fama and French (1993), and others use observable factors.

One fundamental characteristic of many financial time series is that they are often heavy-tailed. To incorporate this observation, we make the common assumption that the random noise $\{\varepsilon_{it}\}_{i=1}^{p}$ are *i.i.d.* random variables in the Domain of Attraction of Fréchet distribution (Leadbetter et al., 1983). Specifically, we adopt the following definition:

Definition 1 (*Leadbetter et al., 1983*). A random variable ε is in the Domain of Attraction of Fréchet distribution with tail index α if and only if $x_F = \infty$ and $1 - F_\varepsilon(x) \sim l(x)x^{-\alpha}$, $\alpha > 0$, where F_ε is the cdf of ε, $l(x)$ is a slowly-varying function and $x_F = \sup\{x : F_\varepsilon(x) < 1\}$. Here and after, for two positive functions $m_1(x)$ and $m_2(x)$, $m_1(x) \sim m_2(x)$ means $\frac{m_1(x)}{m_2(x)} \to 1$, as $x \to \infty$.

Distributions in Domain of Attraction of Fréchet distribution include a broad class of random variables such as Cauchy, Lévy, Pareto and t distributions. To facilitate algebraic derivation, we further assume that for ε_{it}, $l_t(x) \to K_t$ as $x \to \infty$, where $K_t \in \mathcal{F}_{t-1}$ is a positive constant. This is a rather weak assumption with all the aforementioned random variables satisfying this condition. Since K_t can be incorporated into each σ_{it}, without loss of generality, we set $K_t = 1$ in the following. Under a dynamic model, we assume that the conditional tail index α_t of ε_{it} evolves through time according to certain dynamics (e.g. (6)) and $\alpha_t \in \mathcal{F}_{t-1}$.

We also assume that

$$\sup_{1 \le p < \infty} \sup_{1 \le i \le p} |f_i(Z_{1t}, Z_{2t}, \ldots, Z_{dt})| < \infty \quad a.s.$$

Notice here the supremum is taken over p with the number of latent factors d fixed. This is a mild assumption and it includes all the commonly encountered factor models. For example, if the factor model takes a linear form, $f_i(Z_{1t}, \ldots, Z_{dt}) = \sum_{j=1}^{d} \beta_j^{(i)} Z_{jt}$, a sufficient condition for the assumption to hold would be $\sup_{1 \le p < \infty} \sup_{1 \le i \le p} \|\beta^{(i)}\| < \infty$.

We further assume that

$$\lim_{p \to \infty} \sum_{i=1}^{p} \sigma_{it}^{\alpha_t} = \infty \quad \text{and} \quad \lim_{p \to \infty} \sup_{1 \le i \le p} \frac{\sigma_{it}^{\alpha_t}}{\sum_{j=1}^{p} \sigma_{jt}^{\alpha_t}} = 0.$$

Intuitively, it means the magnitudes of conditional volatility σ_{it} are comparable to each other and there is no single X_{it} that dominates the total volatility. For example, if $\sigma_{it} = c_i \sigma_t$ and all c_i's are in a compact positive interval, then the assumption holds.

Given the assumptions, the following result gives the asymptotic conditional distribution of maxima $Q_t = \max_{1 \le i \le p} X_{it}$ when p goes to infinity.

Proposition 1. *Given $\mathcal{F}_{t-1}$, denote $a_{pt} = 0$ and $b_{pt} = \left(\sum_{i=1}^{p} \sigma_{it}^{\alpha_t} \right)^{1/\alpha_t}$, we have, as $p \to \infty$,*

$$\frac{Q_t - a_{pt}}{b_{pt}} \xrightarrow{d} \Psi_{\alpha_t}(x), \tag{7}$$

where $\Psi_{\alpha_t}(x)$ is a Fréchet type random variable with tail index α_t and $\Psi_{\alpha_t}(x) = \exp\left(-x^{-\alpha_t} \right)$.

The proof of Proposition 1 can be found in A.2. Proposition 1 shows that under the framework of the general factor model and some mild conditions, the conditional distribution of maxima $Q_t = \max_{1 \le i \le p} X_{it}$ can be well approximated by a Fréchet distribution. In terms of stochastic representation, (7) can be rewritten as $Q_t \approx \sigma_t Y_t^{1/\alpha_t}$, where Y_t is a unit Fréchet random variable and $\sigma_t = b_{pt}$. To be more flexible and accurate in finite samples, a location parameter μ_t can be included. That is,

$$Q_t \approx \mu_t + \sigma_t Y_t^{1/\alpha_t}, \tag{8}$$

where $(\mu_t, \sigma_t, \alpha_t)$ are time-varying parameters. Setting $\mu_t = \mu$ for parsimonious modeling, we obtain the dynamic structure of $\{Q_t\}$ specified in (4).

Remark 4. In the general factor model, the cross-sectional dependence among X_{it}'s, such as tail dependence, can be introduced by the factor structures. Notice that the independence assumption on ε_{it}'s is not essential for Proposition 1. Based on the results of maxima in stationary series in Leadbetter et al. (1983), similar and more elaborate results can be derived if we impose a stationarity assumption or block independence assumption on ε_{it}'s.

Remark 5. Note that Proposition 1 can handle heterogeneous volatilities among $\{X_{it}\}_{i=1}^{p}$ via $\{\sigma_{it}\}_{i=1}^{p}$. The assumption that $\{\varepsilon_{it}\}_{i=1}^{p}$ share the same tail index α_t may seem to be strong, however, van Oordt and Zhou (2016) found it reasonable in financial applications. See Kelly (2014) and Kelly and Jiang (2014) for a similar assumption. See Proposition 3 in the Appendix for a more involved version of Proposition 1, which handles the case that $\{\varepsilon_{it}\}_{i=1}^{p}$ have heterogeneous tail indices.

Remark 6. The general factor model considered in this section is just an example whose limiting form of maxima coincides with our proposed AcF model. In this paper we are not focusing on the general factor model. Instead, we focus on AcF models.

Remark 7. For the intra-day high-frequency returns of a stock, a sensible assumption (e.g. Bali and Weinbaum, 2007) is that on the same day t, the high-frequency returns $\{X_{it}\}_{i=1}^{p}$ follow a stationary time series such as a GARCH process or a Stochastic Volatility model. Under such a stationary assumption, the intra-day maxima Q_t asymptotically follows a Fréchet distribution with the same tail index α_t as the p high-frequency returns $\{X_{it}\}_{i=1}^{p}$ observed on day t (e.g. Davis and Mikosch, 2009a, b).

2.5. Connection between AcF and the dynamic POT-GPD approach

By Proposition 1 and Remark 7 in Section 2.4, the tail index α_t of the maxima Q_t corresponds to the tail index of the underlying time series $\{X_{it}\}_{i=1}^{p}$ under both the cross-sectional panel data setting and the high-frequency univariate time series setting. Thus, the estimated tail index given by AcF can approximate the underlying true tail index process of $\{X_{it}\}_{i=1}^{p}$ and can then be used to study the tail risk for both types of data.

As mentioned before, AcF belongs to the dynamic Maxima-GEV approach, while most existing literature on time-varying tail risk use the dynamic POT-GPD approach to estimate the tail index. For example, see Kelly (2014) and Kelly and Jiang (2014) for the cross-sectional panel data setting and see Zhang and Schwaab (2017) for the high-frequency univariate time series setting.

As is well known (e.g. Coles, 2001), both Maxima-GEV and POT-GPD provide consistent estimation of the true tail index. One advantage of Maxima-GEV is that it offers a direct modeling of the maxima, which may be of interest especially under the high-frequency univariate time series setting. Also, for POT-GPD, the choice of the threshold can be a sensitive tuning

parameter. On the other hand, POT-GPD picks up all "relevant" observations above a high threshold and may make better use of the available information comparing to Maxima-GEV which only focuses on the maxima. In terms of estimation efficiency, the general view is that POT-GPD can be more efficient than Maxima-GEV under many cases since on average more data are available for the estimation of POT-GPD, while POT-GPD and Maxima-GEV often have comparable performances for large sample sizes. In financial applications, the sample size is usually large, thus AcF offers a promising alternative for accurately estimating the time-varying tail index while also providing a direct modeling of the maxima behavior. See Ferreira and de Haan (2015) and references therein for a nice summary and further theoretical comparisons between the two approaches.

Remark 8. The objective of both AcF and the popular dynamic POT-GPD approach is to model the time-varying tail index rather than the full distributional behavior of the underlying financial time series. The tail index is the focus of the tail risk literature. It provides a measure of the tail risk for a given stock/portfolio and is of interest in a wide range of applications, such as global markets connectedness during crisis, asset pricing theory and effectiveness assessment of central bank intervention (see e.g. Kelly and Jiang, 2014; Massacci, 2016; Zhang and Schwaab, 2017).

3. Parameter estimation

We denote all the parameters in the model by $\theta = (\beta_0, \beta_1, \beta_2, \beta_3, \gamma_0, \gamma_1, \gamma_2, \gamma_3, \mu)$ and denote $\Theta_s = \{\theta | \beta_0, \gamma_0, \mu \in \mathbb{R}, 0 \leq \beta_1, \gamma_1 < 1, \beta_2, \beta_3, \gamma_2, \gamma_3 > 0\}$. In the following, we assume that all allowable parameters are in Θ_s and denote the true parameter by $\theta_0 = (\beta_0^0, \beta_1^0, \beta_2^0, \beta_3^0, \gamma_0^0, \gamma_1^0, \gamma_2^0, \gamma_3^0, \mu_0)$.

The conditional p.d.f. of Q_t given $(\mu, \sigma_t, \alpha_t)$ is

$$f_t(\theta) = f(Q_t | \sigma_t, \alpha_t) = \alpha_t \sigma_t^{\alpha_t} (Q_t - \mu)^{-(\alpha_t + 1)} \exp\left\{-\sigma_t^{\alpha_t}(Q_t - \mu)^{-\alpha_t}\right\}.$$

Hence, by conditional independence, the log-likelihood function with observations $\{Q_t\}_{t=1}^n$ is

$$L_n(\theta) = \frac{1}{n}\sum_{t=1}^n l_t(\theta) = \frac{1}{n}\sum_{t=1}^n \left[\log \alpha_t + \alpha_t \log \sigma_t - (\alpha_t + 1)\log(Q_t - \mu) - \sigma_t^{\alpha_t}(Q_t - \mu)^{-\alpha_t}\right],$$

where $\{\sigma_t, \alpha_t\}_{t=1}^n$ can be obtained recursively through (5) and (6), with an initial value (σ_1, α_1).

Notice here the true value of (σ_1, α_1), denoted as (σ_1^0, α_1^0), is unknown since the state variables $\{\sigma_t, \alpha_t\}$ is a hidden processes. Fortunately, with $0 \leq \beta_1, \gamma_1 < 1$, the influence of (σ_1, α_1) on future (σ_t, α_t) decays exponentially as t increases, hence its impact on parameter estimation will be minimum with a sufficiently large sample size. Theorems 2 and 3 show that the consistency and asymptotic normality of MLE do not depend on whether (σ_1^0, α_1^0) is known and the asymptotic distribution does not depend on the initial value (σ_1, α_1). For simplicity, we use the estimated $(\hat{\sigma}, \hat{\alpha})$ from the static GEV as the initial value for (σ_1, α_1).

Denote the log-likelihood function based on an arbitrary $(\tilde{\sigma}_1, \tilde{\alpha}_1)$ as $\tilde{L}_n(\theta)$. Theorems 2 and 3 show that there always exists a sequence $\hat{\theta}_n$, which is a local maximizer of $\tilde{L}_n(\theta)$, such that $\hat{\theta}_n$ is consistent and asymptotically normal, regardless of the initial value $(\tilde{\sigma}_1, \tilde{\alpha}_1)$.

Theorem 2 (Consistency). *Assume the parameter space Θ is a compact set of Θ_s. Suppose the observations $\{Q_t\}_{t=1}^n$ are generated by a stationary and ergodic AcF with true parameter θ_0 and θ_0 is in the interior of Θ, then there exists a sequence $\hat{\theta}_n$ of local maximizer of $\tilde{L}_n(\theta)$ such that $\hat{\theta}_n \to_p \theta_0$ and $\|\hat{\theta}_n - \theta_0\| \leq \tau_n$, where $\tau_n = O_p(n^{-r})$, $0 < r < 1/2$. Hence $\hat{\theta}_n$ is consistent.*

By the differentiability of $\tilde{L}_n(\theta)$ with respect to θ, the sequence $\hat{\theta}_n$ is also the solution to the score function $\frac{\partial \tilde{L}_n}{\partial \theta}(\theta) = 0$. Theorem 2 guarantees the existence of a sequence of consistent MLE $\hat{\theta}_n$ and is a result about the local behavior of the likelihood function $\tilde{L}_n(\theta)$ near the true parameter value θ_0. The uniqueness of MLE remains an open question due to the complication brought by μ. The same difficulty also applies to the MLE of the static GEV as noted in Smith (1985). Proposition 2 gives a partial answer to the asymptotic uniqueness of MLE.

Proposition 2 (Asymptotic Uniqueness). *Denote $V_n = \{\theta \in \Theta | \mu \leq cQ_{n,1} + (1-c)\mu_0\}$ where $Q_{n,1} = \min_{1 \leq t \leq n} Q_t$, under the conditions in Theorem 2, for any fixed $0 < c < 1$, there exists a sequence of $\hat{\theta}_n = \arg\max_{\theta \in V_n} \tilde{L}_n(\theta)$ such that, $\hat{\theta}_n \to_p \theta_0$, $\|\hat{\theta}_n - \theta_0\| \leq \tau_n$, where $\tau_n = O_p(n^{-r})$, $0 < r < 1/2$, and $P(\hat{\theta}_n$ is the unique global maximizer of $\tilde{L}_n(\theta)$ over $V_n) \to 1$.*

Since $\tilde{L}_n(\theta)$ is defined on $Q_t > \mu$, the parameter space for the maximization of $\tilde{L}_n(\theta)$ is actually $\Theta_n = \{\theta \in \Theta | \mu < Q_{n,1}\}$. Note that for any $0 < c < 1$, $V_n \subseteq \Theta_n$ since $\mu_0 < cQ_{n,1} + (1-c)\mu_0 < Q_{n,1}$. Proposition 2 states that there is an asymptotic unique MLE over V_n, where V_n can be made arbitrarily close to Θ_n by the fact that $Q_{n,1} \searrow \mu_0$ a.s. and by setting c close to 1. In practice, we take $\hat{\theta}_n = \arg\max_{\theta \in \Theta_n} \tilde{L}_n(\theta)$. Numerical experiments confirm its good performance under finite sample.

Theorem 3 (Asymptotic Normality). *Under the conditions in Theorem 2, we have $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, M_0^{-1})$, where $\hat{\theta}_n$ is that in Theorem 2 and M_0 is the Fisher Information matrix evaluated at θ_0. Further, the sample variance of plug-in estimated score functions $\{\frac{\partial}{\partial \theta} l_t(\hat{\theta}_n)\}_{t=1}^n$ is a consistent estimator of M_0.*

The empirical CDF of Q and its limiting Fréchet CDF.
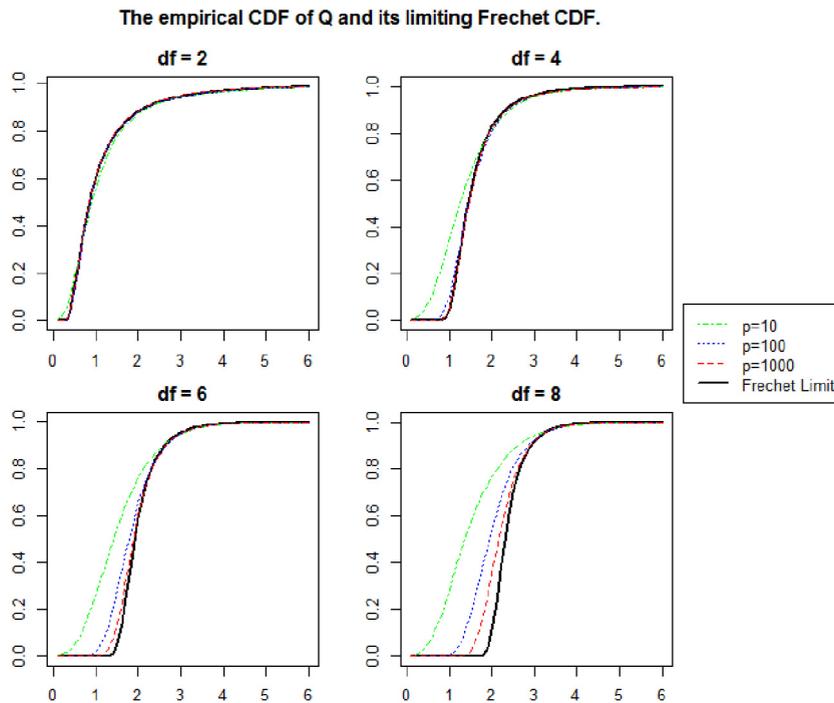
Fig. 2. Finite sample empirical distribution of the maxima Q and its corresponding Fréchet limit, with different combinations of p and degrees of freedom (df) of the t-distribution in the factor model.

The proofs of Theorems 2 and 3 and Proposition 2 can be found in A.3. The main technical difficulty is that the MLE here is irregular in the sense that μ affects the support of the observations Q_t (since $Q_t > \mu$), so regularity conditions of standard MLE are violated. Another technical complication is caused by the fact that the true initial value (σ_1^0, α_1^0) is unknown. In the literature, Smith (1985) proves consistency and asymptotic normality for such irregular MLE for a wide range of distributions, including Fréchet distribution, with *i.i.d.* observations. See Dombry (2015), Bücher and Segers (2016, 2017) for a recent development on this topic. Our proof extends the theoretical result to a dynamic model for dependent time series under stationary and ergodic conditions. We note that this is the first formal treatment for statistical properties of MLE under dynamic EVT framework. The technical tools developed here can also be used for dynamic POT-GPD models.

4. Simulation study

4.1. Convergence of maxima in factor model

In this section, we conduct numerical experiments to investigate the finite sample behavior of Q_t described in Proposition 1. Specifically, we study the convergence of the marginal distribution of Q_t to its Fréchet limit under a one-time period factor model. To simplify notation, we drop the time index t in this section. We simulate data from the following one-factor linear model,

$$X_i = \beta_i Z + \sigma_i \varepsilon_i, \quad i = 1, \dots, p,$$

where $Z \sim N(0, 1)$ is the latent factor, β_i's are *i.i.d.* random coefficients generated from a uniform distribution $U(-2, 2)$ and ε_i's are *i.i.d.* t-distributions with degrees of freedom ν. The σ_i's are *i.i.d.* random variables generated from a mixture of uniform distribution $\frac{1}{2}U(0.5, 1.5) + \frac{1}{2}U(0.75, 1.25)$ such that most σ_i's are moderate in $(0.75, 1.25)$ and the ratio between maximum and minimum σ_i's is 3. This setting roughly matches the pattern of volatilities of different stocks in S&P100 index. For t-distribution, ν corresponds to the tail index α in Definition 1. We set $Q = \max_{1 \le i \le p}(X_i)$.

We compare the finite sample empirical distribution of Q and its corresponding Fréchet limit stated in Proposition 1 under different ν and p. For each (ν, p) combination, 1000 sets of *i.i.d.* $\{X_i\}_{i=1}^p$ are generated, resulting in 1000 sampled $Q = \max_{1 \le i \le p}(X_i)$. Fig. 2 plots the empirical cdf of the normalized Q in (7) along with the corresponding limiting Fréchet distribution. It is clearly seen that as p increases, the empirical distribution of Q approaches its Fréchet limit. A large ν requires larger p for accurate approximation. We also conduct experiments with t-distributed latent factors Z and observe similar results.

Table 1
The performance of AcF on approximation of 1-day conditional VaR for $\{Q_t\}$ process with independent errors ε_{it} and the correlation between the true tail index and the one estimated by AcF.

T_1	$\bar{q}\,(q^0 = 0.1)$	$\bar{q}\,(q^0 = 0.05)$	$\bar{q}\,(q^0 = 0.01)$	mean cor.	median cor.
1000	0.095	0.049	0.012	0.871	0.928
2000	0.096	0.049	0.012	0.909	0.952
5000	0.097	0.051	0.012	0.947	0.973

4.2. AcF estimation for conditional VaR of maxima

In this section, we investigate the temporal approximation ability of AcF to the maxima $\{Q_t\}$ process from a general factor model in terms of 1-day conditional Value at Risk (cVaR). cVaR is the most commonly used measure for tail risk in financial applications. For $0 < q < 1$, cVaR_t^q is defined as the $1 - q$ extreme quantile of Q_t given all past information $\mathcal{F}_{t-1}$, where q is often taken to be 0.1, 0.05 or 0.01. Here, we model the $\{Q_t\}$ process using AcF and calculate the corresponding cVaR_t^q for Q_t using the fitted AcF.

Specifically, we simulate the $\{Q_t\}$ process from a similar one-factor linear model as in Section 4.1,

$$X_{it} = 0.009(\beta_i Z_t + \sigma_i \varepsilon_{it}), \quad i = 1, \ldots, p; \;\; t = 1, \ldots, T,$$

where $Z_t \sim N(0, 1)$ is the latent factor, β_i's are *i.i.d.* random coefficients generated from $U(-2, 2)$, σ_i's are *i.i.d.* random variables generated from a mixture of uniform distribution $\frac{1}{2}U(0.5, 1.5) + \frac{1}{2}U(0.75, 1.25)$ and ε_{it}'s follow *i.i.d.* t-distributions with degrees of freedom ν_t. The multiplier 0.009 is used to control the magnitude of X_{it} to be at the same level of typical stock returns. We fix $p = 100$ and change the observation length T throughout this section. For each day t we obtain $Q_t = \max_{1 \leq i \leq p} X_{it}$. We allow ν_t to evolve, following

$$\log \nu_t = \gamma_0 + \gamma_1 \log \nu_{t-1} + \gamma_2 \exp(-\gamma_3 Q_{t-1}),$$

which resembles the tail index evolution scheme in AcF. Note that the volatility of ε_{it} also evolves implicitly through time due to the dynamics of ν_t. We set the parameters to be $(\gamma_0, \gamma_1, \gamma_2, \gamma_3) = (-0.1, 0.9, 0.3, 5)$ such that the typical range of ν_t is [2, 6]. Note that on day t, $\{X_{it}\}_{i=1}^p$ are dependent and have heterogeneous volatilities.

We use AcF to model the simulated $\{Q_t\}_{t=1}^T$ process and assess the goodness of approximation by AcF's out-sample performance on predicting 1-day cVaR for Q_t. Specifically, we first fit AcF based on the training set $\{Q_t\}_{t=1}^{T_1}$. Then using the fitted AcF, we calculate cVaR_t^q for each Q_t on the test set $\{Q_t\}_{t=T_1+1}^{T_1+T_2}$. The true $\{Q_t\}_{t=T_1+1}^{T_1+T_2}$ are then compared with the $\{\mathrm{cVaR}_t^q\}_{t=T_1+1}^{T_1+T_2}$ and the number of violations is recorded. A violation happens when the observed daily maxima Q_t is larger than the corresponding cVaR_t^q given by AcF. If AcF approximates the tail behavior of $\{Q_t\}$ process well, the expected proportion of violations in the test set should be close to q.

Besides the 1-day cVaR, we also assess the goodness of approximation by calculating the correlation between the true process ν_t and the estimated process $\hat{\alpha}_t$ by AcF in the training set $\{Q_t\}_{t=1}^{T_1}$. Based on the fitted AcF and the observations $\{Q_t\}_{t=1}^{T_1}$, we can recover the estimated tail index $\hat{\alpha}_t$. If AcF is accurate and robust, the correlation between ν_t and the estimated $\hat{\alpha}_t$ is expected to be high, which implies that AcF can detect the true evolution of the tail index ν_t.

We set $T_1 = 1000, 2000, 5000$, $T_2 = 100$ and $q^0 = 0.1, 0.05, 0.01$. For each combination of (T_1, T_2, q^0), we repeat the experiment 500 times. The ith experiment gives a realized violation percentage q_i and we report the average percentage, $\bar{q} = \sum_{i=1}^{500} q_i / 500$, in Table 1.[4] Each experiment also gives a realized correlation between ν_t and $\hat{\alpha}_t$, and we report the mean and median correlation observed in the 500 experiments. As shown in Table 1, the 1-day cVaR given by AcF performs well with the actual violation rate close to the target rate, and the mean and median correlation are reasonably high which indicates that AcF can detect the evolution of the true tail index accurately. Also, a larger training set tends to produce better performance.

Extension 1 — Dependent errors ε_{it}: To further investigate the performance of AcF when ε_{it}'s are dependent, we repeat the experiment for the case where ε_{it}'s are generated from multivariate t-distributions. Specifically, we assume ε_{it}'s are generated from 10 different multivariate t-distributions of size 10. There are 45 pairwise correlations in the correlation matrix of each multivariate t-distribution, 30 of them are generated from $U(0, 0.3)$, 10 are from $U(0.3, 0.4)$ and 5 are from $U(0.4, 0.5)$. For each day t, the 100 ε_{it}'s are generated independently from the 10 multivariate t-distributions with degrees of freedom ν_t and corresponding correlation matrices. Note that marginally each ε_{it} is still a t-distribution with degrees of freedom ν_t. We keep all the other settings unchanged and report the result in Table 2. Again, AcF performs well in terms of 1-day cVaR. Though the mean and median correlation are slightly lower than the ones for independent errors, they are still reasonably high. Thus, it implies that AcF is robust when there are mild dependence among the errors.

Extension 2 — Heterogeneous (in tail indices) errors ε_{it}: To further investigate the performance of AcF when ε_{it}'s have heterogeneous tail indices, we repeat the experiment for the case where ε_{it}'s are generated independently from different

---

[4] We also calculate the p-values for testing $E(q_i) = q^0$ using one-sample Z-test based on $\{q_i\}_{i=1}^{500}$, the result confirms that AcF can approximate the cVaR of Q_t accurately. The result is omitted to conserve space and is available upon request.

Table 2

The performance of AcF on approximation of 1-day conditional VaR for $\{Q_t\}$ process with dependent errors ε_{it} and the correlation between the true tail index and the one estimated by AcF.

T_1	$\bar{q}\,(q^0 = 0.1)$	$\bar{q}\,(q^0 = 0.05)$	$\bar{q}\,(q^0 = 0.01)$	mean cor.	median cor.
1000	0.094	0.048	0.012	0.862	0.921
2000	0.097	0.047	0.011	0.876	0.936
5000	0.096	0.048	0.011	0.918	0.960

Table 3

The performance of AcF on approximation of 1-day conditional VaR for $\{Q_t\}$ process with independent errors ε_{it} having heterogeneous tail indices and the correlation between the true tail index and the one estimated by AcF.

T_1	$\bar{q}\,(q^0 = 0.1)$	$\bar{q}\,(q^0 = 0.05)$	$\bar{q}\,(q^0 = 0.01)$	mean cor.	median cor.
1000	0.098	0.051	0.013	0.864	0.922
2000	0.097	0.050	0.012	0.905	0.953
5000	0.098	0.051	0.012	0.956	0.974

univariate t-distributions. Specifically, we assume ε_{it}'s are generated from 100 different univariate t-distributions with different tail indices $c_i v_t$, where c_i are generated independently from $U(0.8, 1.2)$ for $i = 1, \ldots, 100$. For each day t, the 100 ε_{it}'s are generated independently such that ε_{it} is simulated from the univariate t-distribution with tail index $c_i v_t$. We keep all the other settings unchanged.

Note that though the errors have heterogeneous tail indices $\{c_i v_t\}_{i=1}^{100}$, they share the same time-varying component v_t.[5] As noted by Kelly (2014), v_t is "common to all assets and may therefore be viewed as economy-wide extreme event risk in returns", and it is desirable for AcF to uncover the dynamics of v_t such that the correlation between $\hat{\alpha}_t$ and v_t is close to 1. We report the result in Table 3. AcF performs well in terms of 1-day cVaR and the correlation between $\hat{\alpha}_t$ and v_t is close to 1, indicating the AcF is robust against heterogeneous tail indices and is able to detect the true dynamics of the tail index v_t.

The results in Tables 1 to 3 show that, under various scenarios, the cVaR given by AcF can achieve the desired violation rate q^0 and the tail index estimated by AcF is highly correlated with the true tail index v_t. All together, it indicates that AcF can accurately approximate the tail behavior of the maxima $\{Q_t\}$ process that originates from a general factor model and AcF is robust under misspecification of the scale parameter.

4.3. Performance of the maximum likelihood estimator

To study the finite sample performance of the MLE, we simulate data from an AcF with the following parameters $(\beta_0, \beta_1, \beta_2, \beta_3, \gamma_0, \gamma_1, \gamma_2, \gamma_3, \mu) = (-0.050, 0.96, -0.051, 6.68, -0.068, 0.89, 0.33, 5.33, -0.069)$. This set of parameters is the MLE obtained from an analysis of the S&P100 returns using AcF, shown in Section 5.1. Under this setting, the typical range of α_t is $[2, 8]$ and the typical range of σ_t is $[0.06, 0.21]$.

We investigate the performance of MLE and the corresponding confidence intervals with sample sizes $N = 1000, 5000, 10\,000$. For each sample size, we conduct 500 experiments. Table 4 shows the average of the estimates, the standard deviation from the 500 experiments, and the percentage of estimates that fall into the various confidence intervals based on the asymptotic theory. As can be seen from Table 4, both the bias and variance of the MLE decrease as the sample size N increases, demonstrating the consistency of the MLE under correct model specification. Note that the performance of the MLE is already satisfactory when $N = 1000$. Also, the coverage rate of the asymptotic confidence interval is close to the target rate and improves with the increase of the sample size, validating the asymptotic properties presented in Section 3.

5. Real data applications

In this section, we present two real data applications of AcF, one on the cross-sectional maxima of negative log-returns of stocks in two major U.S. stock indices and one on the intra-day maxima of negative log-returns from high-frequency foreign exchange trading. In both cases, AcF shows its superiority over the static GEV for modeling maxima and its ability to reveal the time-varying nature of the financial market tail risk. Moreover, AcF demonstrates its potential usefulness as a market tail risk measure and an early warning signal for potential extreme movement in the financial market.

5.1. Cross-sectional maxima of the negative daily log-returns of stocks in S&P100 index and DJI30 index

In this section, we consider the cross-sectional maxima of the negative daily log-returns (i.e. daily losses) of component stocks in S&P100 Index (hereafter S&P100) and Dow Jones Industrial Average Index (hereafter DJI30) respectively. For both indices, the time horizon we consider here is from January 1, 2000 to December 31, 2014. The S&P100 index includes 100 leading U.S. stocks and represents about 51% of the market capitalization of the U.S. equity market. The DJI30 index is a

---

[5] See Kelly (2014) and Kelly and Jiang (2014) for similar assumptions on cross-sectional returns.

Table 4
Numerical results for performance of MLE with sample size 1000, 5000, 10 000. Mean and S.D. are the sample mean and standard deviation of the MLE's obtained from 500 simulations. 90% C.I. reports the coverage rate of the 90% C.I. constructed from the estimated Fisher Information matrix; 95% C.I. and 99% C.I. report corresponding coverage rates.

$N = 1000$	γ_0	γ_1	γ_2	γ_3	β_0	β_1	β_2	β_3	μ
True value	−0.068	0.890	0.330	5.33	−0.050	0.960	−0.051	6.68	−0.069
Mean	−0.060	0.884	0.346	6.28	−0.051	0.956	−0.054	5.88	−0.066
S.D.	0.029	0.028	0.058	1.93	0.028	0.019	0.023	3.25	0.011
90% C.I.	81	82	90	91	85	81	75	78	88
95% C.I.	84	88	93	94	87	87	79	80	95
99% C.I.	88	92	97	97	95	94	87	85	98
$N = 5000$	γ_0	γ_1	γ_2	γ_3	β_0	β_1	β_2	β_3	μ
True value	−0.068	0.890	0.330	5.33	−0.050	0.960	−0.051	6.68	−0.069
Mean	−0.066	0.889	0.332	5.52	−0.051	0.959	−0.052	6.53	−0.069
S.D.	0.014	0.012	0.029	0.88	0.012	0.008	0.009	1.83	0.005
90% C.I.	88	87	90	85	88	87	88	87	86
95% C.I.	92	96	93	94	92	91	93	93	94
99% C.I.	95	99	98	99	98	98	97	97	99
$N = 10\,000$	γ_0	γ_1	γ_2	γ_3	β_0	β_1	β_2	β_3	μ
True value	−0.068	0.890	0.330	5.33	−0.050	0.960	−0.051	6.68	−0.069
Mean	−0.067	0.890	0.330	5.44	−0.050	0.960	−0.051	6.55	−0.069
S.D.	0.010	0.007	0.018	0.61	0.007	0.005	0.006	1.37	0.003
90% C.I.	90	88	88	85	89	89	86	89	90
95% C.I.	93	94	94	94	92	94	93	94	98
99% C.I.	98	100	100	99	97	98	98	98	99

Table 5
MLE for cross-sectional maxima of negative daily log-returns for S&P100 (top) and DJI30 (bottom) from January 1, 2000 to December 31, 2014.

S&P100	γ_0	γ_1	γ_2	γ_3	β_0	β_1	β_2	β_3	μ
Mean	−0.068	0.890	0.328	5.33	−0.050	0.961	−0.051	6.68	−0.069
S.D.	0.014	0.013	0.063	1.27	0.006	0.004	0.0072	1.01	0.006
DJI30	γ_0	γ_1	γ_2	γ_3	β_0	β_1	β_2	β_3	μ
Mean	0.023	0.895	0.261	16.32	−0.052	0.964	−0.047	7.38	−0.059
S.D.	0.016	0.013	0.041	3.529	0.005	0.004	0.0066	0.813	0.006

major U.S. stock index, consisting of 30 largest publicly owned companies based in U.S. Both indices are arguably the most important and most-quoted U.S. financial market indicators and provide indications of sentiments in the entire U.S. stock market. To maintain a better management of financial risk, it is essential for financial institutions, especially mutual funds and banks, to understand the cross-sectional tail risks of S&P100 and DJI30.

We present the AcF modeling result for S&P100 in detail. For each trading day t we calculate the negative daily log-returns of each component stock in S&P100 to obtain $\{X_{it}\}_{i=1}^{100}$ and then obtain the daily cross-sectional maxima $Q_t = \max_{1 \leq i \leq 100} X_{it}$. The time series $\{Q_t\}$ has 3773 observations and is shown in the bottom panel of Fig. 3.

The estimation result of AcF is summarized in Table 5. The estimated autoregressive parameter $\hat{\beta}_1$ for the scale parameter $\{\sigma_t\}$ process is 0.96, which suggests a strong persistence of the $\{\hat{\sigma}_t\}$ series. The estimated scale parameter $\{\hat{\sigma}_t\}$ (solid line) is plotted in Fig. 4. For comparison, we also fit a GARCH(1,1) model for each component stock in S&P100 and plot the daily average volatility given by the GARCH models (dashed line) across the 100 stocks in Fig. 4. The two series move very closely with each other with an overall correlation of 0.918. It suggests that AcF's dynamic scale parameter σ_t is an accurate measure of market volatility.

The estimated autoregressive parameter $\hat{\gamma}_1$ for the tail index $\{\alpha_t\}$ process is 0.89, indicating a strong persistence in the tail index process. The estimated tail index $\{\hat{\alpha}_t\}$ is shown in the top panel of Fig. 3. The estimated tail index by AcF is roughly in the range of 2 to 8, which agrees with the empirical finding of Massacci (2016) via a dynamic POT-GPD model. The two periods of persistent small tail index ($\alpha_t < 4$) coincide with the early 2000s U.S. recession and the 2008 financial crisis. Note that the difference between $\alpha_t = 2$ and $\alpha_t = 8$ is very significant. A Fréchet type random variable has its kth moment if and only if $\alpha > k$. It is also noted that almost all $\hat{\alpha}_t$'s are greater than 2, hence the conditional mean and variance of the cross-sectional maxima always exist, which agrees with the existing literature (e.g. Hansen, 1994).

The stationary mean of $(\alpha_t, \sigma_t, \mu)$ of the estimated AcF is $(5.73, 0.099, -0.069)$. We also fitted the static GEV model to the data, assuming the Q_t's are *i.i.d.* observations. The estimated parameters are $(\hat{\alpha}, \hat{\sigma}, \hat{\mu}) = (2.56, 0.058, -0.025)$. It is seen that the estimated tail index of the static GEV model is suspiciously low (see Figs. 1 and 3). It is clear that the static GEV fails to adequately model the time-varying tail risk. On the other hand, the estimated tail index $\hat{\alpha}_t$ given by AcF matches the general pattern of the estimated tail index obtained by the moving-window GEV in the ad-hoc analysis shown in Fig. 1. As shown in Fig. 3, there is a clear negative association between the daily maxima $\{Q_t\}$ series and the estimated tail index $\{\hat{\alpha}_t\}$, making $\hat{\alpha}_t$ a useful measure of the underlying market tail risk, i.e. a market stability index.
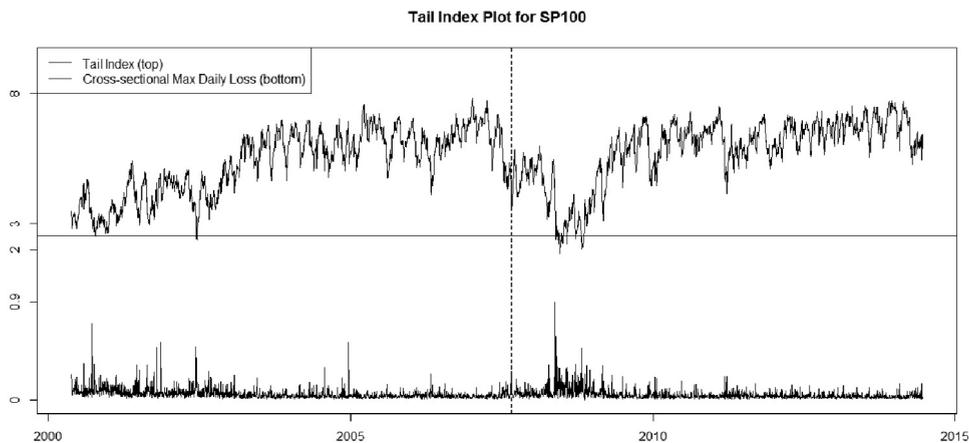
Fig. 3. Estimated tail index $\{\hat{\alpha}_t\}$ (top) and cross-sectional maximum negative daily returns $\{Q_t\}$ (bottom) from January 1, 2000 to December 31, 2014 for S&P100 Index.
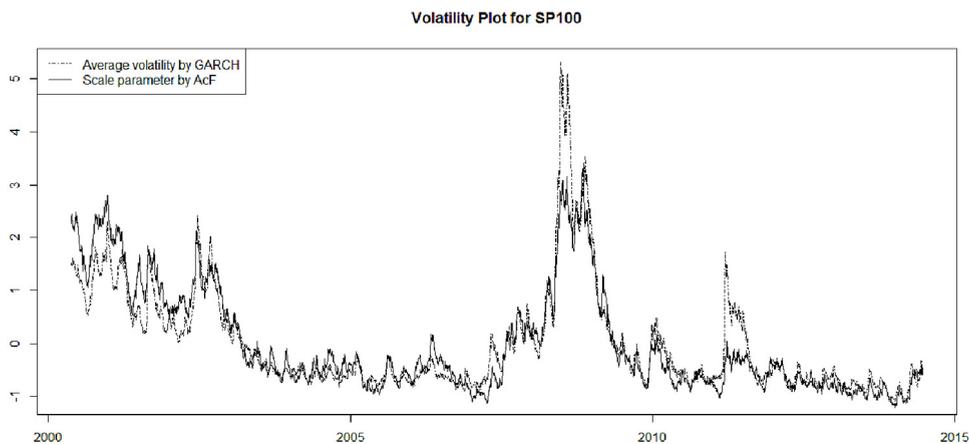
Fig. 4. Estimated scale parameter $\{\hat{\sigma}_t\}$ by AcF (solid line) v.s Estimated average volatility by GARCH (dashed line) from January 1, 2000 to December 31, 2014. Both series are standardized to be zero mean and unit variance for comparison.

We have also applied the same procedure to DJI30. The estimation result of AcF is shown in Table 5 and is similar to the one obtained for S&P100. Due to limited space, we only present its estimated tail index plot (Fig. 5) here. The typical range of tail index for DJI30 is [2.5, 10], with a slight up-shift compared to the one of S&P100, indicating that the tail risk of DJI30 is lower than that of S&P100. This is reasonable considering that the companies in DJI30 are more stable and well-established than those in S&P100. The correlation between the estimated scale parameter $\{\hat{\sigma}_t\}$ from AcF and the average volatility obtained by fitting GARCH model to each stock is 0.909. It again indicates that the evolution scheme of AcF's scale parameter captures the dynamics of the underlying stock market volatility very well. Note that the tail indices of both S&P100 and DJI30 experience some sudden downside movement around late 2007 (marked by vertical dashed lines in the Figures) to reach their lowest level for the past several years. This unusual movement of tail indices may be seen as an early warning signal of the 2008 financial crisis and shows the possibility of using $\hat{\alpha}_t$ as a market stability indicator such as VIX.

The overall correlation between the estimated tail indices of S&P100 and DJI30 is 0.93, suggesting strong common trend between tail risks of the two markets. Based on the estimated tail indices of S&P100 (denoted by $\{\hat{\alpha}_t^S\}$) and DJI30 (denoted by $\{\hat{\alpha}_t^D\}$), we further investigate the tail-connectedness between the two major stock markets following the procedure in Massacci (2016). Let $\hat{\boldsymbol{\alpha}}_t = (\hat{\alpha}_t^S, \hat{\alpha}_t^D)$ be the estimated tail index from the two stock markets and let $\hat{\Sigma}$ be the sample covariance matrix of $\hat{\boldsymbol{\alpha}}_t$ based on a sample $\{\hat{\boldsymbol{\alpha}}_t\}_{t=1}^T$. Using principal components as in Kritzman et al. (2011) and Billio et al. (2012), Massacci (2016) proposes to use the ratio between the maximum eigenvalue of $\hat{\Sigma}$ and the sum of all eigenvalues of $\hat{\Sigma}$ as a measure of tail-connectedness, which "monotonically increases in connectedness among the elements of $\hat{\boldsymbol{\alpha}}_t$ and quantifies the degree of dependence in tail risk among different assets" (in our case, S&P100 and DJI30). Following Billio et al. (2012), we estimate the tail-connectedness measure using a 36-month (approximately 756 days) rolling window and plot the estimated measures in Fig. 6. As expected, the two stock markets have strong connectedness with the maximum eigenvalue explaining between 80% to almost 100% of the variation in tail risk. The two periods where the connectedness
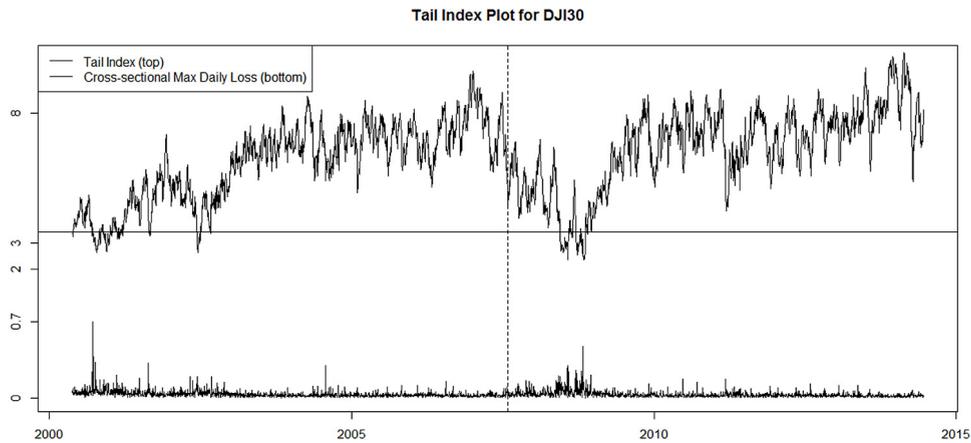
Fig. 5. Estimated tail index $\{\hat{\alpha}_t\}$ (top) and cross-sectional maximum negative daily return $\{Q_t\}$ (bottom) from January 1, 2000 to December 31, 2014 for DJI30 Index.
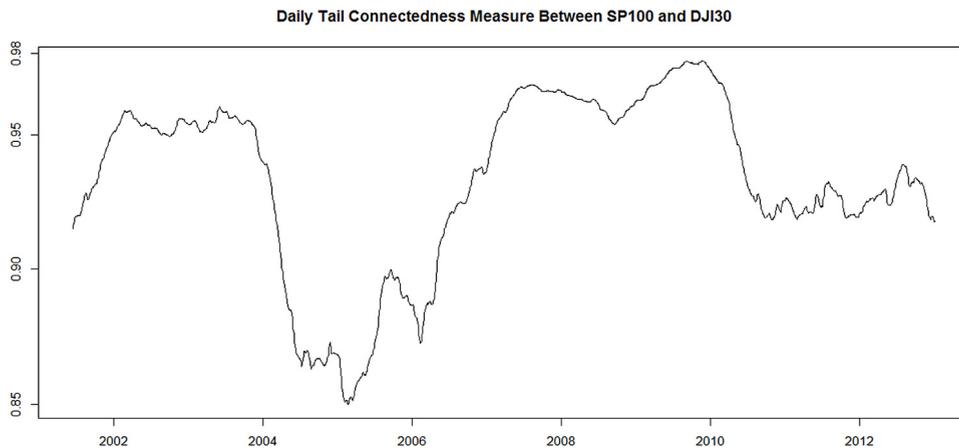
Fig. 6. Estimated tail connectedness measures based on a 36-month rolling window between the tail indices of S&P100 and DJI30 from January 1, 2000 to December 31, 2014.

measure is close to 1 correspond to the early 2000s U.S. recession and the 2008 financial crisis. Our results resemble those in Diebold and Yilmaz (2009) and Massacci (2016), indicating the tail risks of financial assets are more connected during market turmoil.

5.2. Intra-day maxima of 3-minute negative log-returns for USD/JPY foreign exchange rate

In this section, we consider the modeling of intra-day maxima of 3-minute[6] negative log-returns from USD/JPY exchange trading. Specifically, we collect the historical 3-minute intra-day exchange rate of USD/JPY from January 1, 2008 to June 26, 2013. The 3-minute negative log-returns $\{X_{it}\}_{i=1}^{p}$ are obtained and intra-day maxima Q_t are calculated. The total length of the series is 1616. The maxima $\{Q_t\}$ series is shown in Fig. 7(a).

We fit AcF to the intra-day maxima series. Estimated parameters with their standard deviations are shown in Table 6. Similar to the result in the stock market, the estimated autoregressive parameter $\hat{\beta}_1$ for $\{\sigma_t\}$ is 0.89, showing a strong persistence of the $\{\sigma_t\}$ series; while the autoregressive parameter $\hat{\gamma}_1$ for $\{\alpha_t\}$ is 0.59, indicating a less persistent tail index series for the foreign exchange market. The stationary mean of $(\alpha_t, \sigma_t, \mu)$ is $(3.47, 0.167, -0.051)$ under the estimated AcF, while $(\hat{\alpha}, \hat{\sigma}, \hat{\mu}) = (3.25, 0.180, -0.068)$ for the static GEV model. The static GEV model gives a relatively smaller estimated tail index.

---

[6] As pointed out by an anonymous referee, in general it is customary to use 5-minute intervals to avoid the effect of microstructure noise (Diebold and Yilmaz, 2014). Here, we use the 3-minute interval data obtained from Zhang and Zhu (2016) and we believe the result should mainly stay the same if we switch to the 5-minute interval data.

Table 6
MLE for intra-day maxima of 3-minute negative log-returns for USD/JPY from January 1, 2008 to June 26, 2013.

	γ_0	γ_1	γ_2	γ_3	β_0	β_1	β_2	β_3	μ
Mean	0.448	0.587	0.658	20.84	−0.120	0.890	−0.195	6.59	−0.051
S.D.	0.144	0.123	0.203	4.52	0.016	0.012	0.024	0.955	0.010
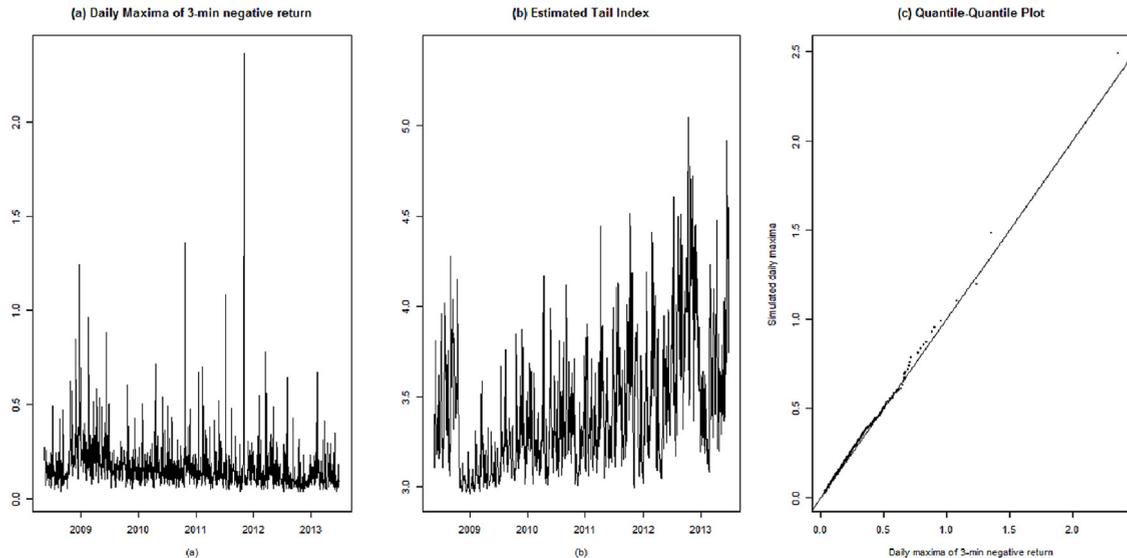
Fig. 7. (a) Daily maxima of 3-minute negative log-returns of USD/JPY from January 1, 2008 to June 26 2013; (b) Estimated tail index $\{\hat{\alpha}_t\}$ from the fitted AcF; (c) Quantile–quantile plot of real data and simulated data from the fitted AcF.

The estimated tail index $\{\hat{\alpha}_t\}$ is shown in Fig. 7(b). It is seen that the tail index is small around 2009, showing a riskier foreign exchange market during the financial crisis. Compared to the tail index of stock market, the tail index series here is also more volatile due to the smaller autoregressive parameter $\hat{\gamma}_1$. The range of the tail index is roughly [3, 5], which suggests that the high-frequency trading of USD/JPY has relatively high risks, as observed by Malinowski et al. (2015). We simulate a $\{Q_t\}$ series of length 10 000 from the estimated AcF and compare its stationary marginal distributions with the observed series using a quantile–quantile plot in Fig. 7(c). It confirms that AcF is a suitable model for the series.

We further test the out-sample performance of AcF for predicting 1-day cVaR^q for the intra-day maxima. First, we fit AcF using the 1000 observations where $1 \leq t \leq 1000$ (roughly 4 years). For the rest 616 observations where $1001 \leq t \leq 1616$, based on the fitted AcF and past information $\mathcal{F}_{t-1}$, we calculate their 1-day cVaR$^{q^0}$ at $q^0 = 0.1, 0.05, 0.01, 0.005, 0.001$. The true daily maxima are then compared with the estimated cVaR and the number of violations is recorded. For comparison, we also fit the static GEV model using the first 1000 observations and calculate the corresponding 1-day cVaR for the rest 616 observations.

Table 7 shows the comparison results. For each q^0, the table presents the number of expected violations ($616q^0$) and the number of actual violations. We also report the p-value of a binomial test[7] for the hypothesis that the actual violation probability and the corresponding q^0 are the same. It is clearly seen that the 1-day cVaR based on AcF performs extremely well, with large p-values for all levels of q^0. On the other hand, the static GEV tends to produce much more conservative cVaR estimates. The comparison clearly demonstrates the time-varying nature of the tail index and the importance of having a dynamic structure such that current market condition is incorporated in the estimation of cVaR.

6. Conclusion

In this paper, we propose a general dynamic GEV framework for the modeling of time series of maxima and the time-varying tail risk. By allowing time-varying scale parameter and tail index of a conditional Fréchet distribution, AcF provides a direct modeling of dynamics of maxima in financial time series and offers a new angle to study the tail risk dynamics in financial markets. Probabilistic properties of AcF are investigated. We implement a maximum likelihood estimator for AcF and investigate its asymptotic properties, using a set of unique technical tools due to the irregularity of the MLE. The real data examples illustrate the efficacy of AcF in practice and its potential broad use in financial risk management and other

---

[7] Under the null hypothesis, the number of violations should follow a binomial distribution with success probability q^0. See Kratz et al. (2016) for more details of the binomial test.

Table 7
Result of 1-day cVaR calculated from AcF and static GEV for intra-day maxima of negative log-returns of USD/JPY exchange rate.

q^0 (%)	Expected Violation	AcF		Static GEV	
		Violation	p-value	Violation	p-value
10	61.6	60	0.89	32	0.00
5.0	30.8	35	0.41	17	0.01
1.0	6.2	8	0.41	2	0.10
0.5	3.1	4	0.56	1	0.39
0.1	0.6	0	1.00	0	1.00

types of tail risk monitoring. A separate empirical paper is being prepared to demonstrate the economic applications of AcF more thoroughly.

The AcF can be extended to incorporate other structures and phenomenon observed in real applications. One potential extension is to assume a parametric dynamic structure for the location parameter μ. An ARMA model is a natural choice and is currently being studied. The choices for function $\eta(\cdot)$ may need further study when it comes to different applications. Another natural extension is to construct multivariate AcF for multivariate maxima. Incorporating the idea of Chen and Fan (2006), an extreme value copula can be used to "link" different univariate AcF models and thus provide a natural model for multivariate time series of maxima. One limitation of AcF is that it cannot model time series with marginal distributions of Weibull type, a distribution used in certain applications. Further investigation is required.

Appendix

A.1. Proof of stationarity and ergodicity

Proof of Theorem 1. The proof of Theorem 1 follows closely the result of Chan and Tong (1994) on non-linear dynamic system. In the following, we assume $\{\sigma_t, \alpha_t\}$ comes from an AcF with parameter $\theta \in \Theta$ as specified in Theorem 1. Without loss of generality, in the following proof, we assume $\mu = 0$. In AcF, $(\log \sigma_t, \log \alpha_t)$ forms a non-linear dynamic system according to the following equation,

$$\log \sigma_t = \beta_0 + \beta_1 \log \sigma_{t-1} - \beta_2 \exp(-\beta_3(\sigma_{t-1}Y_{t-1}^{1/\alpha_{t-1}})),$$

$$\log \alpha_t = \gamma_0 + \gamma_1 \log \alpha_{t-1} + \gamma_2 \exp(-\gamma_3(\sigma_{t-1}Y_{t-1}^{1/\alpha_{t-1}})).$$

To fit $\{\log \sigma_t, \log \alpha_t\}$ into the framework of Chan and Tong (1994), we reparameterize the autoregressive equations as follows:

$$\log \sigma_t = [\beta_0 - z_1 + \beta_1 \log \sigma_{t-1}] + \left[z_1 - \beta_2 \exp(-\beta_3(\sigma_{t-1}Y_{t-1}^{1/\alpha_{t-1}}))\right],$$

$$\log \alpha_t = [\gamma_0 + z_2 + \gamma_1 \log \alpha_{t-1}] + \left[\gamma_2 \exp(-\gamma_3(\sigma_{t-1}Y_{t-1}^{1/\alpha_{t-1}})) - z_2\right],$$

where z_1 is a positive constant such that $0 < z_1 < \beta_2$ (e.g. we can set $z_1 = \beta_2/2$) and z_2 is a positive constant such that $0 < z_2 = \gamma_2 \exp(\frac{\gamma_3}{\beta_3} \log(\frac{z_1}{\beta_2})) < \gamma_2$. The reason for defining z_1, z_2 as above will be made more clear in the proof of Lemma 2. Let $\mathbf{X}_t = (\log \sigma_t, \log \alpha_t)$ and

$$T(\mathbf{X}_{t-1}) = [\beta_0 - z_1 + \beta_1 \log \sigma_{t-1}, \gamma_0 + z_2 + \gamma_1 \log \alpha_{t-1}],$$

$$S(\mathbf{X}_{t-1}, Y_{t-1}) = \left[z_1 - \beta_2 \exp(-\beta_3(\sigma_{t-1}Y_{t-1}^{1/\alpha_{t-1}})), \gamma_2 \exp(-\gamma_3(\sigma_{t-1}Y_{t-1}^{1/\alpha_{t-1}})) - z_2\right],$$

we can rewrite the nonlinear dynamic system of $(\log \sigma_t, \log \alpha_t)$ as

$$\mathbf{X}_t = T(\mathbf{X}_{t-1}) + S(\mathbf{X}_{t-1}, Y_{t-1}),$$

where $\{Y_t\}$ is a sequence of *i.i.d.* Fréchet innovations.

Following the terminology in Chan and Tong (1994), $T(\cdot)$ admits a compact attractor $\Lambda = \left(\frac{\beta_0 - z_1}{1 - \beta_1}, \frac{\gamma_0 + z_2}{1 - \gamma_1}\right)$, which is a singleton in $\mathbb{R}^2$, and the domain of attraction for Λ is $\mathbb{R}^2$. In other words, for any $x \in \mathbb{R}^2$, we have that the iterates $T^n(x) \to \Lambda$ as $n \to \infty$. We further define $G = \left(\frac{\beta_0 - \beta_2}{1 - \beta_1}, \frac{\beta_0}{1 - \beta_1}\right) \times \left(\frac{\gamma_0}{1 - \gamma_1}, \frac{\gamma_0 + \gamma_2}{1 - \gamma_1}\right)$, which is an open rectangle in $\mathbb{R}^2$.

Lemma 1. G is absorbing for $\mathbf{X}_t$.

Proof. We only prove the result for $\log \alpha_t$, the proof for $\log \sigma_t$ is the same. Suppose $\log \alpha_t > \frac{\gamma_0}{1 - \gamma_1}$, then $\log \alpha_{t+1} = \gamma_0 + \gamma_1 \log \alpha_t + \gamma_2 \exp(-\gamma_3 Q_t) > \gamma_0 + \gamma_1 \frac{\gamma_0}{1 - \gamma_1} = \frac{\gamma_0}{1 - \gamma_1}$. Similarly, we can show that $\log \alpha_{t+1} < \frac{\gamma_0 + \gamma_2}{1 - \gamma_1}$ if $\log \alpha_t < \frac{\gamma_0 + \gamma_2}{1 - \gamma_1}$. □

To prove the geometric ergodicity of $\mathbf{X}_t$, we only need to verify conditions (a)–(e) of Theorem 1 in Chan and Tong (1994). The verification of conditions (a), (b), (c) and (e) is trivial and thus is omitted. We verify condition (d) here.

Lemma 2. *For any $x \in G$, 0 is in the support of $|S(x, Y_{t-1})|$ where $|\cdot|$ is the norm of the vector. And there exists a continuous and positive function $r(x)$ for $x \in G$, such that the second step transition probability for $\mathbf{X}_t$, $P^2(x, dy)$, has an absolutely continuous component whose probability density function is positive over $B(T^2(x), r(x))$ where $B(x, \delta)$ denotes the open ball in G with center at x and radius equal to δ.*

Proof. Since $\sigma_{t-1}, \alpha_{t-1} > 0$ and $0 < Y_{t-1} < \infty$, it is easy to see that for any $\mathbf{X}_{t-1}$, there always exists a unique Y_{t-1}^* depending on $\mathbf{X}_{t-1}$ such that $Q_{t-1}^* = \sigma_{t-1}(Y_{t-1}^*)^{1/\alpha_{t-1}} = -\frac{1}{\beta_3} \log(\frac{z_1}{\beta_2})$. By the definition of z_1 and z_2 in $|S(\mathbf{X}_{t-1}, Y_{t-1})|$, it can be verified that given $\mathbf{X}_{t-1}$, Y_{t-1}^* is the unique value that makes $S(\mathbf{X}_{t-1}, Y_{t-1}^*) = 0$. Hence for any $x \in G$, 0 is in the support of $|S(x, Y_{t-1})|$. In the following, we denote $Q^* = -\frac{1}{\beta_3} \log(\frac{z_1}{\beta_2})$.

We now show that there exists a positive constant $r(x) = C$ such that $P^2(x, dy)$ has an absolutely continuous component whose probability density function is positive over $B(T^2(x), C)$. Given $\mathbf{X}_{t-1}$, we have for $\mathbf{X}_{t+1} = (\log \sigma_{t+1}, \log \alpha_{t+1})$,

$$\log \sigma_{t+1} = T^2(\mathbf{X}_{t-1})[1] + [z_1 - \beta_2 \exp(-\beta_3 Q_t)] + \beta_1 [z_1 - \beta_2 \exp(-\beta_3 Q_{t-1})],$$
$$\log \alpha_{t+1} = T^2(\mathbf{X}_{t-1})[2] + [\gamma_2 \exp(-\gamma_3 Q_t) - z_2] + \gamma_1 [\gamma_2 \exp(-\gamma_3 Q_{t-1}) - z_2],$$

where $T^2(\mathbf{X}_{t-1})[1]$ and $T^2(\mathbf{X}_{t-1})[2]$ denote the first and second components of $T^2(\mathbf{X}_{t-1})$ respectively. Given $\mathbf{X}_{t-1}$, $\mathbf{X}_{t+1}$ is a vector function of (Q_{t-1}, Q_t), thus we denote $\mathbf{X}_{t+1} = \mathbf{F}_{\mathbf{X}_{t-1}}(Q_{t-1}, Q_t)$. At $(Q_{t-1}^*, Q_t^*) = (Q^*, Q^*)$, we have $\mathbf{X}_{t+1} = \mathbf{F}_{\mathbf{X}_{t-1}}(Q^*, Q^*) = T^2(\mathbf{X}_{t-1})$. It is easy to verify that the determinant of the Jacobian matrix of $\mathbf{X}_{t+1} = \mathbf{F}_{\mathbf{X}_{t-1}}(Q^*, Q^*)$ at (Q^*, Q^*) is $\exp(-(\beta_3 + \gamma_3)Q^*)\beta_2\beta_3\gamma_2\gamma_3(\beta_1 - \gamma_1)$, which is not zero since $\theta \in \Theta$.

By the Inverse Function Theorem, we know that an inverse function to $\mathbf{F}_{\mathbf{X}_{t-1}}(\cdot)$ exists in an open neighborhood of $\mathbf{X}_{t+1} = \mathbf{F}_{\mathbf{X}_{t-1}}(Q^*, Q^*) = T^2(\mathbf{X}_{t-1})$. By the nature of the vector function $\mathbf{F}_{\mathbf{X}_{t-1}}(\cdot)$, $\mathbf{X}_{t-1}$ does not affect the size of the open neighborhood. Thus for all $\mathbf{X}_{t-1} \in G$, we can find a uniform C such that $B(T^2(\mathbf{X}_{t-1}), C)$ is a subset of the open neighborhood. The rest of the proof simply follows from the fact that (Y_{t-1}, Y_t) are *i.i.d.* unit Fréchet random variables and there is a one-to-one relationship between (Y_{t-1}, Y_t) and (Q_{t-1}, Q_t) given $\mathbf{X}_{t-1}$. $\square$

Now we have verified all five conditions of Theorem 1 in Chan and Tong (1994). Hence $\{\log \sigma_t, \log \alpha_t\}$, as a Markov chain on $G \in \mathbb{R}^2$, is stationary and geometrically ergodic. $\square$

A.2. Proof of conditional distribution of Q_t in general factor model

Proof of Proposition 1. In the following, we drop the time index t for notation simplicity. The conditioning on $\mathcal{F}_{t-1}$ is implicit here. The proof follows standard procedure in the extreme value literature by deriving the cdf of $(Q_p - a_p)/b_p$ directly. Here, $Q_p = \max_{1 \le i \le p} X_i$, $a_p = 0$ and $b_p = (\sum_{i=1}^p \sigma_i^\alpha)^{1/\alpha}$. We have

$$
\begin{aligned}
P\left(\frac{Q_p - a_p}{b_p} \le x\right) &= P\left(\frac{\max_{1 \le i \le p} X_i - a_p}{b_p} \le x\right) = P\left(\max_{1 \le i \le p} X_i \le a_p + b_p x\right) \\
&= P\left(f_i(Z_1, Z_2, \ldots, Z_d) + \sigma_i \varepsilon_i \le a_p + b_p x, \text{ for all } 1 \le i \le p\right) \\
&= P\left(\varepsilon_i \le b_p x/\sigma_i - f_i(Z_1, Z_2, \ldots, Z_d)/\sigma_i, \text{ for all } 1 \le i \le p\right) \\
&= E\left[P\left(\varepsilon_i \le b_p x/\sigma_i - f_i(Z_1, Z_2, \ldots, Z_d)/\sigma_i, \text{ for all } 1 \le i \le p | Z_1, \ldots, Z_d\right)\right] \\
&= E\left(\prod_{i=1}^p F_\varepsilon\left(b_p x/\sigma_i - f_i(Z_1, Z_2, \ldots, Z_d)/\sigma_i\right)\right),
\end{aligned}
$$

where the last equality follows from the independence between ε_i's and the latent factors Z_i's. By the assumption that

$$\sup_{1 \le p < \infty} \sup_{1 \le i \le p} |f_i(Z_1, Z_2, \ldots, Z_d)| < \infty \ a.s.$$

and

$$\lim_{p \to \infty} \sum_{i=1}^p \sigma_i^\alpha = \infty \text{ and } \lim_{p \to \infty} \sup_{1 \le i \le p} \frac{\sigma_i^\alpha}{\sum_{j=1}^p \sigma_j^\alpha} = 0,$$

it is easy to see that, for any fixed $x > 0$,

$$\lim_{p \to \infty} \inf_{1 \le i \le p} \left(b_p x/\sigma_i - f_i(Z_1, Z_2, \ldots, Z_d)/\sigma_i\right) = \infty \ a.s.$$

Together with the assumption that F_ε is in the Domain of Attraction of Fréchet distribution, we have uniformly for all i,

$$F_\varepsilon(b_p x/\sigma_i - f_i(Z_1, Z_2, \ldots, Z_d)/\sigma_i) \sim$$
$$1 - l(b_p x/\sigma_i - f_i(Z_1, Z_2, \ldots, Z_d)/\sigma_i)(b_p x/\sigma_i - f_i(Z_1, Z_2, \ldots, Z_d)/\sigma_i)^{-\alpha} \ a.s.,$$

where $\sim$ has the same meaning as in Definition 1. Together with the fact that $\lim_{x\to\infty} l(x) = 1$, we have

$$\sum_{i=1}^{p} l(b_p x/\sigma_i - f_i(Z_1, Z_2, \ldots, Z_d)/\sigma_i)(b_p x/\sigma_i - f_i(Z_1, Z_2, \ldots, Z_d)/\sigma_i)^{-\alpha}$$

$$= \frac{1}{b_p^{\alpha}} \sum_{i=1}^{p} l(b_p x/\sigma_i - f_i(Z_1, Z_2, \ldots, Z_d)/\sigma_i)(x/\sigma_i - b_p^{-1} f_i(Z_1, Z_2, \ldots, Z_d)/\sigma_i)^{-\alpha} \to x^{-\alpha} \quad a.s.,$$

where the last equality follows from the fact that $b_p \to \infty$ and $\sup_{1 \le p < \infty} \sup_{1 \le i \le p} |f_i(Z_1, Z_2, \ldots, Z_d)| < \infty$. By the bounded convergence theorem, we have for any fixed $x > 0$,

$$P\left(\frac{Q_p - a_p}{b_p} \le x\right) = P\left(\max_{1 \le i \le p} X_i \le a_p + b_p x\right) \to \exp\left(-x^{-\alpha}\right), \quad \text{as } p \to \infty. \quad \square$$

In the following, we give Proposition 3, which handles the case when $\{\varepsilon_{it}\}_{i=1}^{p}$ have heterogeneous tail indices. Under the general setting of Proposition 1, we assume that $\{\varepsilon_{it}\}_{i=1}^{p}$ are independently but not identically distributed. Specifically, we assume that $\{\varepsilon_{it}\}_{i=1}^{p}$ form K groups of *i.i.d.* errors, indexed by $G_1, G_2, \ldots, G_K$, where the distributions of ε_{it} in the same group are identical and the distribution of ε_{it} for each group is in the Domain of Attraction of a Fréchet distribution with a different tail index $\alpha_t(k), k = 1, \ldots, K$. Without loss of generality, assume that $\alpha_t(1) = \min_{1 \le k \le K} \alpha_t(k)$, i.e. Group 1 has the smallest tail index. Note that for every $1 \le i \le p$, there exists one and only one $1 \le k \le K$ such that $i \in G_k$ and $\alpha_{it} = \alpha_t(k)$.

As in Proposition 1, we assume that

$$\lim_{p\to\infty} \sum_{i=1}^{p} \sigma_{it}^{\alpha_{it}} = \infty \quad \text{and} \quad \lim_{p\to\infty} \sup_{1 \le i \le p} \frac{\sigma_{it}^{\alpha_{it}}}{\sum_{j=1}^{p} \sigma_{jt}^{\alpha_{jt}}} = 0.$$

In addition, we further assume that

$$\lim_{p\to\infty} \frac{\sum_{i \in G_1} \sigma_{it}^{\alpha_{it}}}{\sum_{j=1}^{p} \sigma_{jt}^{\alpha_{jt}}} = \lim_{p\to\infty} \frac{\sum_{i \in G_1} \sigma_{it}^{\alpha_t(1)}}{\sum_{j=1}^{p} \sigma_{jt}^{\alpha_{jt}}} = \pi > 0.$$

Intuitively, it means that the volatility of errors in Group 1 (with the smallest tail index $\alpha_t(1)$) is not ignorable compared to the ones of the other Groups.

Proposition 3. *Given $\mathcal{F}_{t-1}$, denote $a_{pt} = 0$ and $b_{pt} = \left(\pi \cdot \sum_{i=1}^{p} \sigma_{it}^{\alpha_{it}}\right)^{1/\alpha_t(1)}$, we have, as $p \to \infty$,*

$$\frac{Q_t - a_{pt}}{b_{pt}} \xrightarrow{d} \Psi_{\alpha_t(1)}(x), \tag{9}$$

where $\Psi_{\alpha_t(1)}(x)$ is a Fréchet type random variable with tail index $\alpha_t(1)$ and $\Psi_{\alpha_t(1)}(x) = \exp\left(-x^{-\alpha_t(1)}\right)$.

The proof of Proposition 3 follows the same line as the one of Proposition 1 and thus is omitted. Proposition 3 states that under the case that the errors $\{\varepsilon_{it}\}_{i=1}^{p}$ have heterogeneous tail indices, the conditional distribution of maxima $Q_t = \max_{1 \le i \le p} X_{it}$ can still be approximated by a Fréchet distribution, however, the tail index of Q_t takes the value of the smallest tail index $\alpha_t(1)$ among the K different tail indices.

Remark: The setting of Proposition 3 is natural for cross-sectional multivariate stock returns, where stocks can usually be clustered into different groups according to certain criteria such as industrial sector and it is expected that stocks within the same group k share similar behavior such as tail index $\alpha_t(k)$. For the cross-sectional setting, a common and sensible assumption for heterogeneity of tail indices $\alpha_t(k)$'s (e.g. see Kelly (2014), Kelly and Jiang (2014)) is that $\alpha_t(k) = c_k \alpha_t$, where $c_k > 0$. In other words, the tail indices $\alpha_t(k)$'s for different groups are different but they share the same time-varying component α_t. By Proposition 3, asymptotically the tail index of Q_t then takes the value $(\min_{1 \le k \le K} c_k) \cdot \alpha_t$, which is perfectly correlated with the dynamics of the true α_t. Thus the estimated $\hat{\alpha}_t$ by AcF should be highly correlated with the dynamics of the true tail index α_t.

A.3. Proof of consistency and asymptotic normality

To facilitate the proof of Theorems 2, 3 and Proposition 2, we first give several technical lemmas (Lemmas 3 to 15). As mentioned in Section 3, the main technical difficulty is that the location parameter μ_0 is unknown and the support of Q_t depends on μ_0, so that the standard argument for MLE cannot be directly applied. Also, the true initial value (σ_1^0, α_1^0) is unknown. New uniform convergence results about the log-likelihood function $\tilde{L}_n(\theta)$, its first and second order derivatives need to be established. The main result on uniform convergence is stated in Lemma 14. Part of the proof follows that in Francq and Zakoian (2004) for MLE of GARCH model.

In the following, we assume the conditions in Theorem 2 hold, i.e. the parameter space Θ is a compact set of Θ_s and the observations $\{Q_t\}_{t=1}^{n}$ come from a stationary and ergodic AcF with true parameter θ_0 where θ_0 is in the interior of Θ. We

use $Y_{n,k}$ and $Q_{n,k}$ to denote the kth order statistics of $\{Y_t\}_{t=1}^n$ and $\{Q_t\}_{t=1}^n$. In the following, $\tau_n \sim n^{-r}$ means $\tau_n/n^{-r} \to 1$ as $n \to \infty$. We denote the upper bound of β_1, γ_1 in Θ by $C_b < 1$ and use C to denote a generic positive constant.

We first prove the identifiability of AcF in Lemma 3, which states that each parameter value θ defines a unique AcF.

Lemma 3 (*Identifiability*). *If $Q_t(\theta) = Q_t(\theta_0)$ a.s. for all t, then $\theta = \theta_0$. Here a.s. is for the infinite product space generated by $\{\dots, Y_{-1}, Y_0, Y_1, Y_2, \dots\}$, where Y_i's are i.i.d. unit Fréchet random variables.*

Proof. We denote $\sigma_t = \sigma_t(\theta), \alpha_t = \alpha_t(\theta)$ and $\sigma_t^0 = \sigma_t(\theta_0), \alpha_t^0 = \alpha_t(\theta_0)$. Suppose there exist θ and θ_0 such that $Q_t(\theta) = Q_t(\theta_0)$ a.s., then

$$\mu_0 + \sigma_t^0 Y_t^{1/\alpha_t^0} = \mu + \sigma_t Y_t^{1/\alpha_t}, \ a.s.$$

Since $Y_{n,1} \searrow 0$ a.s., by the boundedness of (σ_t, α_t) and (σ_t^0, α_t^0), we have $\mu = \mu_0$. After rearrangement,

$$Y_t^{1/\alpha_t^0 - 1/\alpha_t} = \sigma_t/\sigma_t^0, \ a.s.$$

Denote $\mathcal{F}_t = \sigma(Y_t, Y_{t-1}, \dots)$, we know that $Y_t \perp \mathcal{F}_{t-1}$ and $\alpha_t, \alpha_t^0 \in \mathcal{F}_{t-1}$, so the above equation holds if and only if $\sigma_t(\theta) = \sigma_t(\theta_0)$ and $\alpha_t(\theta) = \alpha_t(\theta_0)$ a.s. From the autoregressive equation of $\log \alpha_t$, we know that if $\alpha_t(\theta) = \alpha_t(\theta_0)$ a.s., we have

$$\gamma_0^0 + \gamma_1^0 \log \alpha_{t-1} + \gamma_2^0 \exp(-\gamma_3^0 Q_{t-1}) = \gamma_0 + \gamma_1 \log \alpha_{t-1} + \gamma_2 \exp(-\gamma_3 Q_{t-1}).$$

After rearrangement, we have

$$\gamma_0^0 - \gamma_0 + (\gamma_1^0 - \gamma_1) \log \alpha_{t-1} = \gamma_2 \exp(-\gamma_3 Q_{t-1}) - \gamma_2^0 \exp(-\gamma_3^0 Q_{t-1}).$$

By the same argument as above, since $\alpha_{t-1} \in \mathcal{F}_{t-2}$ and $Q_{t-1} \notin \mathcal{F}_{t-2}$, we must have $\gamma_0 = \gamma_0^0, \gamma_1 = \gamma_1^0, \gamma_2 = \gamma_2^0$ and $\gamma_3 = \gamma_3^0$. Similarly, we can prove that $\beta_0 = \beta_0^0, \beta_1 = \beta_1^0, \beta_2 = \beta_2^0$ and $\beta_3 = \beta_3^0$. □

Given parameter θ and an initial value $(\sigma_1, \alpha_1), \{\sigma_t, \alpha_t\}_{t=1}^n$ can be recovered recursively by their autoregressive equations. In the following, we use $\sigma_t(\theta), \alpha_t(\theta)$ (or σ_t, α_t for simplicity) to denote the scale parameter series and the tail index series based on θ and true initial (σ_1^0, α_1^0), and use $\tilde{\sigma}_t(\theta), \tilde{\alpha}_t(\theta)$ (or $\tilde{\sigma}_t, \tilde{\alpha}_t$ for simplicity) to denote the ones based on θ and an arbitrary initial value $(\tilde{\sigma}_1, \tilde{\alpha}_1)$. We denote the unobserved true hidden process by $\sigma_t(\theta_0), \alpha_t(\theta_0)$ (or σ_t^0, α_t^0 for simplicity).

By the compactness of Θ and the boundedness of $-\beta_2 \exp(-\beta_3 Q_{t-1}), \gamma_2 \exp(-\gamma_3 Q_{t-1})$, there exist uniform lower bound and the upper bound of $\{\sigma_t, \alpha_t\}$ and $\{\tilde{\sigma}_t, \tilde{\alpha}_t\}$ for all $\theta \in \Theta$. We denote the lower bound by (σ_L, α_L) and upper bound by (σ_U, α_U). The uniform boundedness plays a key role in the following proof.

Given (σ_t, α_t), the conditional log-likelihood function $l_t(\theta)$ of Q_t is,

$$l_t(\theta) = \log \alpha_t + \alpha_t \log \sigma_t - (\alpha_t + 1) \log(Q_t - \mu) - \left(\frac{Q_t - \mu}{\sigma_t}\right)^{-\alpha_t}.$$

By conditional independence, the log-likelihood function

$$L_n(\theta) = \frac{1}{n} \sum_{t=1}^n l_t(\theta) = \frac{1}{n} \sum_{t=1}^n \log \alpha_t + \alpha_t \log \sigma_t - (\alpha_t + 1) \log(Q_t - \mu) - \left(\frac{Q_t - \mu}{\sigma_t}\right)^{-\alpha_t}.$$

We use $\tilde{l}_t(\theta)$ and $\tilde{L}_n(\theta)$ to denote the corresponding log-likelihood functions when $(\tilde{\sigma}_t, \tilde{\alpha}_t)$ are used.

Lemma 4 gives the result about the behavior of score function and Fisher information matrix at the true parameter θ_0 given true initial value (σ_1^0, α_1^0).

Lemma 4. *Under the conditions in Theorem 2, $E_{\theta_0}(\frac{\partial}{\partial \theta} l_t(\theta_0)) = 0$ and for M_0, the Fisher information matrix at θ_0, we have $M_0 = \text{Var}_{\theta_0}(\frac{\partial}{\partial \theta} l_t(\theta_0)) = -E_{\theta_0}(\frac{\partial^2}{\partial \theta \partial \theta^T} l_t(\theta_0))$ and M_0 is well defined and positive definite.*

Proof. At $\theta = \theta_0$, all the first order partial derivatives $\frac{\partial}{\partial \theta} l_t(\theta_0)$ and second order partial derivatives $\frac{\partial^2}{\partial \theta_i \partial \theta_j} l_t(\theta_0)$ can be simplified and written as functions of $\{Y_t, \sigma_t^0, \alpha_t^0\}_{t=1}^n$. The formulas for $\frac{\partial}{\partial \theta} l_t(\theta)$ and $\frac{\partial^2}{\partial \theta_i \partial \theta_j} l_t(\theta)$ are postponed to Appendix A.4 due to their complexity.

By the fact that (σ_t^0, α_t^0) is bounded between $[\sigma_L, \sigma_U] \times [\alpha_L, \alpha_U]$, $Y_t \perp \mathcal{F}_{t-1} = \sigma(Y_s, s \le t-1)$ and $(\sigma_t^0, \alpha_t^0) \in \mathcal{F}_{t-1}$, it is easy to prove that $E_{\theta_0}(\frac{\partial}{\partial \theta} l_t(\theta_0)) = 0, M_0 = \text{Var}_{\theta_0}\left(\frac{\partial}{\partial \theta} l_t(\theta_0)\right) = -E_{\theta_0}(\frac{\partial^2}{\partial \theta \partial \theta^T} l_t(\theta_0))$ and M_0 is well defined, i.e. $M_0 < \infty$.

To prove that M_0 is positive definite, notice that $M_0 = \text{Var}_{\theta_0}(\frac{\partial}{\partial \theta} l_t(\theta_0))$, so we only need to show that there does not exist a $c \in \mathbb{R}^9$ such that $c^T \frac{\partial}{\partial \theta} l_t(\theta_0) = 0$ a.s. The argument is the same as the one used in Lemma 3, where the essential idea is that $Y_t \perp \mathcal{F}_{t-1}$. □

The result of Lemma 4 is standard and expected, since θ_0 is the true parameter and we assume the data come from a correctly specified model. Lemma 5 gives moment conditions for functions of $\{Q_t\}_{t=1}^n$, which serve as building blocks for the proof of latter lemmas.

Lemma 5. *Under the conditions in Theorem 2, we have (a) for any $\alpha > 0$, $\frac{1}{n}\sum_{t=1}^{n}(Q_t - \mu_0)^{-\alpha} \to_p E_{\theta_0}(Q_1 - \mu_0)^{-\alpha} < \infty$, (b) for any positive integer k, $\frac{1}{n}\sum_{t=1}^{n}[\log(Q_t - \mu_0)]^k \to_p E_{\theta_0}[\log(Q_1 - \mu_0)]^k < \infty$.*

Proof. By the boundedness of scale parameter $\{\sigma_t^0\}$ and tail index $\{\alpha_t^0\}$, we have $Q_t - \mu_0 > \sigma_L \min(Y_t^{1/\alpha_L}, Y_t^{1/\alpha_U})$, so $E_{\theta_0}(Q_t - \mu_0)^{-\alpha} < \infty$ for any $\alpha > 0$ since Y_t^{-1} follows exponential distribution. The result of (a) follows from the ergodicity of AcF and the Law of Large Numbers.

For (b), we have $|\log(Q_t - \mu_0)|^k = |\log\sigma_t + 1/\alpha_t \log Y_t|^k \leq 2^k(C + 1/\alpha_L^k|\log Y_t|^k)$. It is known that $\log Y_t$ follows a Gumbel distribution thus $E_{\theta_0}(|\log Y_t|^k) < \infty$ for any positive integer k. The result of (b) follows from the ergodicity of AcF and the Law of Large Numbers. $\square$

As mentioned above, the main technical difficulty is that the support of Q_t depends on the unknown location parameter μ_0. Lemma 6 to Lemma 14 aim to solve this difficulty by establishing uniform convergence between $\frac{1}{n}\sum_{t=1}^{n}h(Q_t - \mu_n)$ and $\frac{1}{n}\sum_{t=1}^{n}h(Q_t - \mu_0)$ for μ_n within a neighborhood of μ_0, where $h(\cdot)$ denotes some generic function that appears in the first and second order derivatives of $\tilde{L}_n(\theta)$. The main result is stated in Lemma 14.

Lemma 6 gives an asymptotic bound on the distance between $Q_{n,1}$ and μ_0, stating that $Q_{n,1}$ converges to μ_0 at a rate that is slower than polynomial.

Lemma 6. *Under the conditions in Theorem 2, $Q_{n,1} - \mu_0 \geq O_p((\log n)^{-1/\alpha_L})$.*

Proof. Notice that when $Y_t < 1$, we have $Q_t - \mu_0 = \sigma_t Y_t^{1/\alpha_t} \geq \sigma_L Y_t^{1/\alpha_L}$. Since $Y_{n,1} < 1$ *a.s.* as $n \to \infty$, it is obvious that $Q_{n,1} - \mu_0 \geq \sigma_L Y_{n,1}^{1/\alpha_L}$ *a.s.* as $n \to \infty$. The result follows from the fact that $(\log n)Y_{n,1} \to_p 1$. $\square$

Lemma 7 gives the foundation for the uniform convergence result of first and second order derivatives of $L_n(\theta)$ given in Lemma 11, Lemma 14.

Lemma 7. *Denote $S_n^{\alpha}(\mu) = n^{-1}\sum_{k=1}^{n}(Q_{n,k} - \mu)^{-\alpha}$, $\alpha > 0$ or $S_n^{\alpha}(\mu) = n^{-1}\sum_{k=1}^{n}\log(Q_{n,k} - \mu)$ or $S_n^{\alpha}(\mu) = n^{-1}\sum_{k=1}^{n}(Q_{n,k} - \mu)^{-\alpha}[\log(Q_{n,k} - \mu)]^m$ for $m = 1, 2, 3$. Under the conditions in Theorem 2, given positive sequence τ_n, s.t. $\tau_n \sim n^{-r}, r > 0$, the following result holds uniformly over $|\mu_n - \mu_0| < \tau_n$,*

$$\left|S_n^{\alpha}(\mu_n) - S_n^{\alpha}(\mu_0)\right| \leq O_p(\tau_n).$$

Proof. We prove the result for (a) $S_n^{\alpha}(\mu) = n^{-1}\sum_{k=1}^{n}(Q_{n,k} - \mu)^{-\alpha}$, (b) $S_n^{\alpha}(\mu) = n^{-1}\sum_{k=1}^{n}\log(Q_{n,k} - \mu)$ and (c) $S_n^{\alpha}(\mu) = n^{-1}\sum_{k=1}^{n}(Q_{n,k} - \mu)^{-\alpha}\log(Q_{n,k} - \mu)$, a similar argument with more involved calculus can be used for the proof of others. By Lemma 6, we know that $Q_{n,1} - \mu_0 \geq O_p((\log n)^{-1/\alpha_L})$, so $(Q_t - \mu_n)^{-\alpha}$ and $\log(Q_t - \mu_n)$ are asymptotically well defined for $|\mu_n - \mu_0| < \tau_n$.

(a) For $S_n^{\alpha}(\mu) = n^{-1}\sum_{k=1}^{n}(Q_{n,k} - \mu)^{-\alpha}$, assume that $\mu_n > \mu_0$, we have

$$|S_n^{\alpha}(\mu_n) - S_n^{\alpha}(\mu_0)| \leq \frac{1}{n}\sum_{k=1}^{n}|(Q_{n,k} - \mu_n)^{-\alpha} - (Q_{n,k} - \mu_0)^{-\alpha}| \leq \frac{1}{n}\sum_{k=1}^{n}\frac{(\alpha+1)|\mu_n - \mu_0|}{\min\{Q_{n,k} - \mu_n, Q_{n,k} - \mu_0\}^{\alpha+1}}$$

$$\leq \frac{\tau_n}{n}\sum_{k=1}^{n}\frac{\alpha+1}{(Q_{n,k} - \mu_n)^{\alpha+1}} = \frac{\tau_n}{n}\sum_{k=1}^{n}\frac{\alpha+1}{(Q_{n,k} - \mu_0 + \mu_0 - \mu_n)^{\alpha+1}} \leq \frac{\tau_n}{n}\sum_{k=1}^{n}\frac{\alpha+1}{(Q_{n,k} - \mu_0 - \tau_n)^{\alpha+1}},$$

where the second inequality follows from the fact that $a - a^{\alpha+1} \leq (\alpha + 1)(1 - a)$ for all $\alpha > 0$ and $0 < a < 1$.

Since $Q_{n,1} - \mu_0 \geq O_p((\log n)^{-1/\alpha_L})$, for any fixed $0 < \rho < 1$, we have $P(\rho(Q_{n,1} - \mu_0) > \tau_n) \to 1$, so $P(\rho(Q_{n,k} - \mu_0) > \tau_n$, for all $1 \leq k \leq n) \to 1$. With probability goes to 1, we have

$$\frac{\tau_n}{n}\sum_{k=1}^{n}\frac{\alpha+1}{(Q_{n,k} - \mu_0 - \tau_n)^{\alpha+1}} \leq \frac{\tau_n}{n}\sum_{k=1}^{n}\frac{\alpha+1}{[(Q_{n,k} - \mu_0)(1 - \rho)]^{\alpha+1}} = O_p(\tau_n),$$

which follows from Lemma 5(a). For $\mu_n < \mu_0$, the proof is similar but easier.

(b) For $S_n^{\alpha}(\mu) = n^{-1}\sum_{k=1}^{n}\log(Q_{n,k} - \mu)$, assume that $\mu_n > \mu_0$, we have

$$|S_n^{\alpha}(\mu_n) - S_n^{\alpha}(\mu_0)| \leq \frac{1}{n}\sum_{k=1}^{n}|\log(Q_{n,k} - \mu_n) - \log(Q_{n,k} - \mu_0)|$$

$$= \frac{1}{n}\sum_{k=1}^{n}\log\left(1 + \frac{\mu_n - \mu_0}{Q_{n,k} - \mu_n}\right) \leq \frac{\tau_n}{n}\sum_{k=1}^{n}\frac{1}{Q_{n,k} - \mu_n} = O_p(\tau_n),$$

where the last inequality follows from the fact that $\log(1 + x) < x$ when $x > 0$ and the last equality follows from the result for $S_n^{\alpha}(\mu) = n^{-1}\sum_{k=1}^{n}(Q_{n,k} - \mu)^{-\alpha}$. For $\mu_n < \mu_0$, the proof is similar but easier.

(c) For $S_n^\alpha(\mu) = n^{-1}\sum_{k=1}^n (Q_{n,k} - \mu)^{-\alpha} \log(Q_{n,k} - \mu)$, assume that $\mu_n > \mu_0$, we have

$$|S_n^\alpha(\mu_n) - S_n^\alpha(\mu_0)| \leq \frac{1}{n}\sum_{k=1}^n (Q_{n,k} - \mu_n)^{-\alpha} |\log(Q_{n,k} - \mu_n) - \log(Q_{n,k} - \mu_0)|$$

$$+ \frac{1}{n}\sum_{k=1}^n \left|(Q_{n,k} - \mu_n)^{-\alpha} - (Q_{n,k} - \mu_0)^{-\alpha}\right| |\log(Q_{n,k} - \mu_0)|.$$

For the first term in the sum,

$$\frac{1}{n}\sum_{k=1}^n (Q_{n,k} - \mu_n)^{-\alpha} |\log(Q_{n,k} - \mu_n) - \log(Q_{n,k} - \mu_0)|$$

$$= \frac{1}{n}\sum_{k=1}^n (Q_{n,k} - \mu_n)^{-\alpha} \log\left(1 + \frac{\mu_n - \mu_0}{Q_{n,k} - \mu_n}\right)$$

$$\leq \frac{\tau_n}{n}\sum_{k=1}^n (Q_{n,k} - \mu_n)^{-(\alpha+1)} = O_p(\tau_n),$$

where the last equality follows from the result for $S_n^\alpha(\mu) = n^{-1}\sum_{k=1}^n (Q_{n,k} - \mu)^{-\alpha}$. For the second term in the sum,

$$\frac{1}{n}\sum_{k=1}^n \left|(Q_{n,k} - \mu_n)^{-\alpha} - (Q_{n,k} - \mu_0)^{-\alpha}\right| |\log(Q_{n,k} - \mu_0)|$$

$$\leq \frac{\tau_n}{n}\sum_{k=1}^n \frac{\alpha+1}{(Q_{n,k} - \mu_n)^{\alpha+1}} |\log(Q_{n,k} - \mu_0)| \leq \tau_n \left(\frac{1}{n}\sum_{k=1}^n \frac{(\alpha+1)^2}{(Q_{n,k} - \mu_n)^{2\alpha+2}}\right)^{1/2} \left(\frac{1}{n}\sum_{k=1}^n |\log(Q_{n,k} - \mu_0)|^2\right)^{1/2}$$

$$= O_p(\tau_n),$$

where the last inequality follows from the Cauchy–Schwarz inequality and the last equality follows from Lemma 5 and the result for $S_n^\alpha(\mu) = n^{-1}\sum_{k=1}^n (Q_{n,k} - \mu)^{-\alpha}$. For $\mu_n < \mu_0$, the proof is similar but easier. $\square$

Lemmas 8 and 9 provide the impact of parameter difference $|\theta - \theta_0|$ on $|\alpha_t - \alpha_t^0|$ and $|\sigma_t - \sigma_t^0|$ uniformly over t.

Lemma 8. *Denote $\Phi = (\gamma_0, \gamma_1, \gamma_2, \gamma_3)$ and $\Phi_0 = (\gamma_0^0, \gamma_1^0, \gamma_2^0, \gamma_3^0)$, if $\|\Phi - \Phi_0\| < \tau_n$ and $\tau_n \searrow 0$, under the conditions in Theorem 2, we have*

$$(a)\ \sup_{1\leq t\leq n} \left|\alpha_t - \alpha_t^0\right| = O(\tau_n), \quad (b)\ \sup_{1\leq t\leq n} \left|\frac{\partial \alpha_t}{\partial \Phi} - \frac{\partial \alpha_t^0}{\partial \Phi}\right| = O(\tau_n), \quad (c)\ \sup_{1\leq t\leq n} \left|\frac{\partial^2 \alpha_t}{\partial \Phi_i \partial \Phi_j} - \frac{\partial^2 \alpha_t^0}{\partial \Phi_i \partial \Phi_j}\right| = O(\tau_n),$$

uniformly over $\|\Phi - \Phi_0\| < \tau_n$.

Proof. We only prove (a), the proof for others is similar but more involved. Using the fact that a continuously differentiable function is Lipschitz continuous on a compact set, we only need to prove that $\sup_{1\leq t\leq n}\left|\log\alpha_t - \log\alpha_t^0\right| = O(\tau_n)$. By repeatedly applying the autoregressive relation, we can get

$$\log\alpha_t = \gamma_0 \sum_{k=1}^{t-1} \gamma_1^{k-1} + \gamma_2 \sum_{k=1}^{t-1} \gamma_1^{k-1} \exp(-\gamma_3 Q_{t-k}) + \gamma_1^{t-1} \log\alpha_1^0.$$

We have

$$|\log\alpha_t - \log\alpha_t^0| \leq \left|\gamma_0 \sum_{k=1}^{t-1} \gamma_1^{k-1} - \gamma_0^0 \sum_{k=1}^{t-1} (\gamma_1^0)^{k-1}\right| + \left|\gamma_1^{t-1} \log\alpha_1^0 - (\gamma_1^0)^{t-1} \log\alpha_1^0\right|$$

$$+ \left|\gamma_2 \sum_{k=1}^{t-1} \gamma_1^{k-1} \exp(-\gamma_3 Q_{t-k}) - \gamma_2^0 \sum_{k=1}^{t-1} (\gamma_1^0)^{k-1} \exp(-\gamma_3^0 Q_{t-k})\right|.$$

By the fact that $\sum_{k=1}^t \gamma_1^{k-1} < 1/(1-\gamma_1) \leq 1/(1-C_b)$ and $\left|\sum_{k=1}^{t-1} (\gamma_1^0)^{k-1} - \sum_{k=1}^{t-1} (\gamma_1)^{k-1}\right| \leq \left|\frac{1}{1-\gamma_1^0} - \frac{1}{1-\gamma_1}\right| \leq \frac{\tau_n}{(1-C_b)^2} = O(\tau_n)$, it is easy to see that the first two terms of the sum are $O(\tau_n)$ for any $1 \leq t \leq n$. For the third term, we have

$$\left|\gamma_2 \sum_{k=1}^{t-1} \gamma_1^{k-1} \exp(-\gamma_3 Q_{t-k}) - \gamma_2^0 \sum_{k=1}^{t-1} (\gamma_1^0)^{k-1} \exp(-\gamma_3^0 Q_{t-k})\right|$$

$$\le |\gamma_2 - \gamma_2^0| \sum_{k=1}^{t-1} \gamma_1^{k-1} \exp(-\gamma_3 Q_{t-k}) + \gamma_2^0 \sum_{k=1}^{t-1} \left|\gamma_1^{k-1} - (\gamma_1^0)^{k-1}\right| \exp(-\gamma_3 Q_{t-k})$$

$$+ \gamma_2^0 \sum_{k=1}^{t-1} (\gamma_1^0)^{k-1} \left|\exp(-\gamma_3 Q_{t-k}) - \exp(-\gamma_3^0 Q_{t-k})\right|.$$

The first two terms of the sum are $O(\tau_n)$ for any $1 \le t \le n$ by the boundedness of $\exp(-\gamma_3 Q_{t-k})$. For the third term we have,

$$\gamma_2^0 \sum_{k=1}^{t-1} (\gamma_1^0)^{k-1} \left|\exp(-\gamma_3 Q_{t-k}) - \exp(-\gamma_3^0 Q_{t-k})\right|$$

$$= \gamma_2^0 \sum_{k=1}^{t-1} (\gamma_1^0)^{k-1} Q_{t-k} \exp(-\gamma_{3k}' Q_{t-k}) \left|\gamma_3 - \gamma_3^0\right| = O(\tau_n), \text{ for any } 1 \le t \le n,$$

where $\gamma_{3k}' > 0$ is a number between γ_3 and γ_3^0 depending on Q_{t-k}, and $\gamma_{3k}' \to \gamma_3^0$ uniformly over all $k \ge 1$. By the compactness of Θ, $\gamma_{3k}' \ge C > 0$ for all $k \ge 1$. Mean value theorem is used to get the first equality and the uniform boundedness of $Q_{t-k} \exp(-\gamma_{3k}' Q_{t-k})$ is used to get the second equality. □

Lemma 9. *Denote $\Psi = (\beta_0, \beta_1, \beta_2, \beta_3)$ and $\Psi_0 = (\beta_0^0, \beta_1^0, \beta_2^0, \beta_3^0)$, if $\|\Psi - \Psi_0\| < \tau_n$ and $\tau_n \searrow 0$, under the conditions in* [Theorem 2](), *we have*

$$(a) \sup_{1 \le t \le n} \left|\sigma_t - \sigma_t^0\right| = O(\tau_n), \quad (b) \sup_{1 \le t \le n} \left|\frac{\partial \sigma_t}{\partial \Psi} - \frac{\partial \sigma_t^0}{\partial \Psi}\right| = O(\tau_n), \quad (c) \sup_{1 \le t \le n} \left|\frac{\partial^2 \sigma_t}{\partial \Psi_i \partial \Psi_j} - \frac{\partial^2 \sigma_t^0}{\partial \Psi_i \partial \Psi_j}\right| = O(\tau_n),$$

uniformly over $\|\Psi - \Psi_0\| < \tau_n$.

Proof. The proof is the same as the one for [Lemma 8]() and thus omitted. □

[Lemma 10]() is used for the proof of [Lemma 11]().

Lemma 10. *Suppose $\tau_n \sim n^{-r}$, $r > 0$ and $\sup_{1 \le t \le n} |\alpha_t - \alpha_t'| = O(\tau_n)$ where $\{\alpha_t\}$ and $\{\alpha_t'\}$ represent two different series of tail index. Under the conditions in* [Theorem 2](), *we have*

$$\frac{1}{n} \sum_{t=1}^n \left|(Q_t - \mu_n)^{-\alpha_t} - (Q_t - \mu_n)^{-\alpha_t'}\right| = O_p(\tau_n),$$

uniformly over $|\mu_n - \mu_0| < \tau_n$. The same result holds for $\frac{1}{n} \sum_{t=1}^n \left|(Q_t - \mu_n)^{-\alpha_t} - (Q_t - \mu_n)^{-\alpha_t'}\right| [\log(Q_t - \mu_n)]^k$, $k = 1, 2$.

Proof. We only prove the result for $\frac{1}{n} \sum_{t=1}^n \left|(Q_t - \mu_n)^{-\alpha_t} - (Q_t - \mu_n)^{-\alpha_t'}\right|$, the proof for others is the same. Assume $\alpha_t' > \alpha_t$, the proof for the other direction is the same. By mean value theorem,

$$\frac{1}{n} \sum_{t=1}^n \left|(Q_t - \mu_n)^{-\alpha_t} - (Q_t - \mu_n)^{-\alpha_t'}\right| \le \frac{C}{n} \sum_{t=1}^n (Q_t - \mu_n)^{-\alpha_t^*} |\log(Q_t - \mu_n)| \tau_n$$

$$\le \frac{\tau_n C}{n} \sum_{t=1}^n \left((Q_t - \mu_n)^{-\alpha_L} + (Q_t - \mu_n)^{-\alpha_U}\right) |\log(Q_t - \mu_n)| = O_p(\tau_n),$$

where $\alpha_t^* \in (\alpha_t, \alpha_t')$. The last equality follows from [Lemma 7](). □

[Lemma 11]() gives the uniform convergence result of the second order derivatives of $L_n(\theta)$ over a neighborhood of θ_0, which is used in the proof of [Lemma 14]()(a). In the following, we denote $m_{\theta_i \theta_j}(\theta_0) = -E_{\theta_0}\left(\frac{\partial^2}{\partial \theta_i \partial \theta_j} l_1(\theta_0)\right)$.

Lemma 11. *Under the conditions in* [Theorem 2](), *for all second order derivatives of $L_n(\theta_n)$, we have $\frac{\partial^2}{\partial \theta_i \partial \theta_j} L_n(\theta_n) \to_p -m_{\theta_i \theta_j}(\theta_0)$, uniformly over $\|\theta_n - \theta_0\| < \tau_n$, where $\tau_n \sim n^{-r}$, $r > 0$.*

Proof. We only prove the case for $\frac{\partial^2}{\partial \mu^2} L_n(\theta_n)$, the proof for others is similar but more involved. By the Law of Large Numbers, we know that $\frac{\partial^2}{\partial \mu^2} L_n(\theta_0) \to_p m_{\mu\mu}(\theta_0)$, so we only need to prove that $\frac{\partial^2}{\partial \mu^2} L_n(\theta_n) - \frac{\partial^2}{\partial \mu^2} L_n(\theta_0) \to_p 0$ uniformly over the claimed

region.

$$\frac{\partial^2}{\partial \mu^2} L_n(\theta_n) - \frac{\partial^2}{\partial \mu^2} L_n(\theta_0) = \frac{1}{n} \sum_{t=1}^{n} \left[(\alpha_t + 1)(Q_t - \mu_n)^{-2} - (\alpha_t^0 + 1)(Q_t - \mu_0)^{-2} \right]$$

$$- \frac{1}{n} \sum_{t=1}^{n} \left[\alpha_t(\alpha_t + 1)\sigma_t^{\alpha_t}(Q_t - \mu_n)^{-(\alpha_t+2)} - \alpha_t^0(\alpha_t^0 + 1)(\sigma_t^0)^{\alpha_t^0}(Q_t - \mu_0)^{-(\alpha_t^0+2)} \right].$$

We now analyze the difference term by term. For the first term,

$$\left| \frac{1}{n} \sum_{t=1}^{n} \left[(\alpha_t + 1)(Q_t - \mu_n)^{-2} - (\alpha_t^0 + 1)(Q_t - \mu_0)^{-2} \right] \right|$$

$$\leq \frac{1}{n} \sum_{t=1}^{n} \left| (\alpha_t + 1)[(Q_t - \mu_n)^{-2} - (Q_t - \mu_0)^{-2}] \right| + \frac{C\tau_n}{n} \sum_{t=1}^{n} (Q_t - \mu_0)^{-2} = O_p(\tau_n) \to 0,$$

where the inequality comes from the fact that $|\alpha_t - \alpha_t^0| = O(\tau_n)$ uniformly for all $1 \leq t \leq n$ by Lemma 8(a), and the equality comes from Lemma 7 and boundedness of $\{\alpha_t\}$. For the second term,

$$\left| \frac{1}{n} \sum_{t=1}^{n} \alpha_t(\alpha_t + 1)\sigma_t^{\alpha_t}(Q_t - \mu_n)^{-(\alpha_t+2)} - \frac{1}{n} \sum_{t=1}^{n} \alpha_t^0(\alpha_t^0 + 1)(\sigma_t^0)^{\alpha_t^0}(Q_t - \mu_0)^{-(\alpha_t^0+2)} \right|$$

$$\leq \frac{1}{n} \sum_{t=1}^{n} \alpha_t(\alpha_t + 1)\sigma_t^{\alpha_t} \left| (Q_t - \mu_n)^{-(\alpha_t+2)} - (Q_t - \mu_0)^{-(\alpha_t+2)} \right|$$

$$+ \frac{1}{n} \sum_{t=1}^{n} \alpha_t(\alpha_t + 1)\sigma_t^{\alpha_t} \left| (Q_t - \mu_0)^{-(\alpha_t+2)} - (Q_t - \mu_0)^{-(\alpha_t^0+2)} \right|$$

$$+ \frac{1}{n} \sum_{t=1}^{n} \left| \alpha_t(\alpha_t + 1)\sigma_t^{\alpha_t} - \alpha_t^0(\alpha_t^0 + 1)(\sigma_t^0)^{\alpha_t^0} \right| (Q_t - \mu_0)^{-(\alpha_t^0+2)}.$$

By Lemma 8(a), we know that $\sup_{1 \leq t \leq n} |\alpha_t - \alpha_t^0| = O(\tau_n)$. The first two terms go to zero by Lemmas 7 and 10 respectively, and the last term goes to zero by the boundedness of $\{\sigma_t, \alpha_t\}$, the differentiable continuity of $\alpha_t(\alpha_t + 1)\sigma_t^{\alpha_t}$ w.r.t. σ_t, α_t and Lemma 8(a), Lemma 9(a). ☐

Note that our ultimate goal is to establish uniform convergence result about $\tilde{L}_n(\theta)$. Lemmas 12 and 13 state that the impact of arbitrary initial value $(\tilde{\sigma}_1, \tilde{\alpha}_1)$ on the behavior of $\tilde{L}_n(\theta)$ is asymptotically negligible over a neighborhood of μ_0.

Lemma 12. *Under the conditions in Theorem 2, there exists a positive constant C such that for all $\theta \in \Theta$ and $t \geq 1$,*

$$(a) \ |\alpha_t - \tilde{\alpha}_t| \leq C \cdot C_b^{t-1}, \quad (b) \ \left| \frac{\partial \alpha_t}{\partial \Phi} - \frac{\partial \tilde{\alpha}_t}{\partial \Phi} \right| \leq C \cdot t C_b^{t-1}, \quad (c) \ \left| \frac{\partial^2 \alpha_t}{\partial \Phi_i \partial \Phi_j} - \frac{\partial^2 \tilde{\alpha}_t}{\partial \Phi_i \partial \Phi_j} \right| \leq C \cdot t^2 C_b^{t-1},$$

$$(d) \ |\sigma_t - \tilde{\sigma}_t| \leq C \cdot C_b^{t-1}, \quad (e) \ \left| \frac{\partial \sigma_t}{\partial \Phi} - \frac{\partial \tilde{\sigma}_t}{\partial \Phi} \right| \leq C \cdot t C_b^{t-1}, \quad (f) \ \left| \frac{\partial^2 \sigma_t}{\partial \Phi_i \partial \Phi_j} - \frac{\partial^2 \tilde{\sigma}_t}{\partial \Phi_i \partial \Phi_j} \right| \leq C \cdot t^2 C_b^{t-1}.$$

Proof. We skip the proof since it is obvious. ☐

Lemma 13. *Under the conditions in Theorem 2, we have $\frac{1}{n} \sum_{t=1}^{n} \left| (Q_t - \mu_n)^{-\alpha_t} - (Q_t - \mu_n)^{-\tilde{\alpha}_t} \right| \to_p 0$, uniformly over $|\mu_n - \mu_0| < \tau_n$, where $\tau_n \sim n^{-r}, r > 0$. The same result holds for*

$$\frac{1}{n} \sum_{t=1}^{n} \left| (Q_t - \mu_n)^{-\alpha_t} - (Q_t - \mu_n)^{-\tilde{\alpha}_t} \right| [\log(Q_t - \mu_n)]^k, \ k = 1, 2.$$

Proof. We only prove the result for $\frac{1}{n} \sum_{t=1}^{n} \left| (Q_t - \mu_n)^{-\alpha_t} - (Q_t - \mu_n)^{-\tilde{\alpha}_t} \right|$, the proof for others is the same. By Lemma 12(a), we have $|\alpha_t - \tilde{\alpha}_t| \leq C \cdot C_b^{t-1}$. Assume $\tilde{\alpha}_t > \alpha_t$, the proof for the other direction is the same. By mean value theorem,

$$\frac{1}{n} \sum_{t=1}^{n} \left| (Q_t - \mu_n)^{-\alpha_t} - (Q_t - \mu_n)^{-\tilde{\alpha}_t} \right| \leq \frac{C}{n} \sum_{t=1}^{n} (Q_t - \mu_n)^{-\alpha_t^*} \left| \log(Q_t - \mu_n) \right| C_b^{t-1}$$

$$\leq \frac{C}{n} \sum_{t=1}^{n} \left((Q_t - \mu_n)^{-\alpha_L} + (Q_t - \mu_n)^{-\alpha_U} \right) \left| \log(Q_t - \mu_n) \right| C_b^{t-1} \to_p 0$$

where $\alpha_t^* \in (\alpha_t, \tilde{\alpha}_t)$. The result follows from Lemma 7 and that

$$E_{\theta_0}\Big[\sum_{t=1}^{\infty}\big((Q_t - \mu_0)^{-\alpha_L} + (Q_t - \mu_0)^{-\alpha_U}\big)\,|\log(Q_t - \mu_0)|\,C_b^{t-1}\Big] < \infty. \quad \square$$

Lemma 14 states the main uniform convergence result used in the proof of Theorems 2 and 3.

Lemma 14. *Under the conditions in Theorem 2, (a) for all second order derivatives of $\tilde{L}_n(\theta)$, we have $\frac{\partial^2}{\partial\theta_i\partial\theta_j}\tilde{L}_n(\theta) \to_p -m_{\theta_i\theta_j}(\theta_0)$, uniformly over $\|\theta - \theta_0\| < \tau_n$, where $\tau_n \sim n^{-r}, r > 0$ (b) for the score function of $\tilde{L}_n(\theta)$, we have $(\tau_n^*)^{-1}\big(\frac{\partial}{\partial\theta}\tilde{L}_n(\theta_0) - \frac{\partial}{\partial\theta}L_n(\theta_0)\big) \to_p 0$ if $\tau_n^* n \to \infty$, e.g. $\tau_n^* = 1/\sqrt{n}$.*

Proof (a). is a direct result of Lemma 11 and the fact that $\frac{\partial^2}{\partial\theta_i\partial\theta_j}\tilde{L}_n(\theta) - \frac{\partial^2}{\partial\theta_i\partial\theta_j}L_n(\theta) \to_p 0$ uniformly over $\|\theta - \theta_0\| < \tau_n$. The proof of $\frac{\partial^2}{\partial\theta_i\partial\theta_j}\tilde{L}_n(\theta) - \frac{\partial^2}{\partial\theta_i\partial\theta_j}L_n(\theta) \to_p 0$ uniformly is based on Lemmas 12 and 13. The argument is the same as that in the proof of Lemma 11, thus we skip it.

We prove (b) for $\frac{\partial}{\partial\mu}\tilde{L}_n(\theta_0)$, the proof for other first order partial derivatives is similar. Let $g(\sigma_t, \alpha_t) = \alpha_t\sigma_t^{\alpha_t}$, by the fact that $|\sigma_t - \tilde{\sigma}_t| \le C \cdot C_b^{t-1}$, $|\alpha_t - \tilde{\alpha}_t| \le C \cdot C_b^{t-1}$, we have $|g(\sigma_t, \alpha_t) - g(\tilde{\sigma}_t, \tilde{\alpha}_t)| \le C \cdot C_b^{t-1}$.

$$\frac{1}{\tau_n^*}\left(\frac{\partial}{\partial\mu}\tilde{L}_n(\theta_0) - \frac{\partial}{\partial\mu}L_n(\theta_0)\right) = \frac{1}{n\tau_n^*}\sum_{t=1}^{n}\left(\frac{\alpha_t - \tilde{\alpha}_t}{Q_t - \mu_0} - \frac{g(\sigma_t, \alpha_t)}{(Q_t - \mu_0)^{\alpha_t+1}} + \frac{g(\tilde{\sigma}_t, \tilde{\alpha}_t)}{(Q_t - \mu_0)^{\tilde{\alpha}_t+1}}\right)$$

$$= \frac{1}{n\tau_n^*}\sum_{t=1}^{n}\left(\frac{\alpha_t - \tilde{\alpha}_t}{Q_t - \mu_0} - \frac{g(\sigma_t, \alpha_t) - g(\tilde{\sigma}_t, \tilde{\alpha}_t)}{(Q_t - \mu_0)^{\alpha_t+1}} + g(\tilde{\sigma}_t, \tilde{\alpha}_t)[(Q_t - \mu_0)^{-(\tilde{\alpha}_t+1)} - (Q_t - \mu_0)^{-(\alpha_t+1)}]\right).$$

The first term is bounded by $\frac{C}{n\tau_n^*}\sum_{t=1}^{n}\frac{C_b^{t-1}}{Q_t - \mu_0}$ and the second term by $\frac{C}{n\tau_n^*}\sum_{t=1}^{n}\frac{C_b^{t-1}}{(Q_t - \mu_0)^{\alpha_t+1}}$. Both terms go to zero in probability since $n\tau_n^* \to \infty$ and $E_{\theta_0}\big[\sum_{t=1}^{\infty}C_b^{t-1}(Q_t - \mu_0)^{-\alpha}\big] < \infty$ for all $\alpha > 0$. The same argument applies to the third term after applying mean value theorem. $\quad \square$

Lemma 15 gives the standard Martingale CLT.

Lemma 15. *Under the conditions in Theorem 2,*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial l_t(\theta_0)}{\partial\theta} \Rightarrow N(0, M_0^{-1}),$$

where M is the Fisher Information matrix at θ_0.

Proof. We prove this result by using CLT for martingale difference (Billingsley, 1961). It is easy to verify that,

$$E_{\theta_0}\left(\frac{\partial l_t(\theta_0)}{\partial\theta}\Big|\mathcal{F}_{t-1}\right) = 0 \text{ and } \text{Var}_{\theta_0}\left(\frac{\partial l_t(\theta_0)}{\partial\theta}\right) = M_0 < \infty.$$

So for any $\lambda \in \mathbb{R}^9$, $\{\lambda'\frac{\partial l_t(\theta_0)}{\partial\theta}, \mathcal{F}_t\}_t$ is a square-integrable stationary martingale difference. By CLT of Billingsley (1961) and Wold–Cramér device, Lemma 15 is true. $\quad \square$

Proof of Theorem 2. The proof mainly uses Taylor expansion. Let $\{\tau_n\}$ be any sequence s.t. $\tau_n \sim n^{-r}$ and $n^{1/2}\tau_n \to \infty$ (i.e. $0 < r < 1/2$), let $t \in \mathbb{R}$, $y \in \mathbb{R}^8$ and define $f_n(t, y) = \tau_n^{-2}\tilde{L}_n(\mu_0 + \tau_n t, \phi^0 + \tau_n y)$, where we denote $\phi^0 = (\beta_0^0, \beta_1^0, \beta_2^0, \beta_3^0, \gamma_0^0, \gamma_1^0, \gamma_2^0, \gamma_3^0)$.

By Taylor Expansion we have,

$$\frac{\partial}{\partial t}f_n(t, y) = \tau_n^{-1}\frac{\partial\tilde{L}_n(\mu_0 + \tau_n t, \phi^0 + \tau_n y)}{\partial\mu} = \tau_n^{-1}\frac{\partial\tilde{L}_n(\mu_0, \phi^0)}{\partial\mu} + \frac{\partial^2\tilde{L}_n(\mu^*, \phi^*)}{\partial\mu^2}t + \sum_{i=1}^{8}\frac{\partial^2\tilde{L}_n(\mu^*, \phi^*)}{\partial\mu\partial\phi_i}y_i$$

$$= \tau_n^{-1}\left(\frac{\partial\tilde{L}_n(\mu_0, \phi^0)}{\partial\mu} - \frac{\partial L_n(\mu_0, \phi^0)}{\partial t}\right) + \tau_n^{-1}\left(\frac{\partial L_n(\mu_0, \phi^0)}{\partial\mu}\right) + \frac{\partial^2\tilde{L}_n(\mu^*, \phi^*)}{\partial\mu^2}t + \sum_{i=1}^{8}\frac{\partial^2\tilde{L}_n(\mu^*, \phi^*)}{\partial\mu\partial\phi_i}y_i,$$

where the second equality comes from a Taylor expansion of $\frac{\partial\tilde{L}_n(\mu_0+\tau_n t, \phi^0+\tau_n y)}{\partial t}$ at (μ_0, ϕ^0), and we have $|\mu^* - \mu_0| < \tau_n t$ and $\|\phi^* - \phi^0\| < \tau_n\|y\|$. The first term goes to 0 by Lemma 14(b) and the second term goes to 0 by Lemma 15 and the fact that $\tau_n\sqrt{n} \to \infty$. By Lemma 14(a), the last two terms converge uniformly over $t^2 + \|y\|^2 \le 1$, i.e.,

$$\frac{\partial^2\tilde{L}_n(\mu^*, \phi^*)}{\partial\mu^2}t + \sum_i\frac{\partial^2\tilde{L}_n(\mu^*, \phi^*)}{\partial\mu\partial\phi_i}y_i \to_p -m_{\mu\mu}(\theta_0)t - \sum_{i=1}^{8}m_{\mu\phi_i}(\theta_0)y_i.$$

So together we have $\frac{\partial}{\partial t} f_n(t, y) = -m_{\mu\mu}(\theta_0)t - \sum_{i=1}^{8} m_{\mu\phi_i}(\theta_0) y_i + o_p(1)$. Similarly, we have $\frac{\partial}{\partial y_k} f_n(t, y) = -m_{\phi_k\mu}(\theta_0)t - \sum_{i=1}^{8} m_{\phi_k\phi_i}(\theta_0) y_i + o_p(1)$, for $k = 1, \ldots, 8$, where $o_p(1)$'s are uniformly decaying over $t^2 + \|y\|^2 \leq 1$. Let $t^2 + \|y\|^2 = 1$, we have

$$t \frac{\partial f_n}{\partial t}(t, y) + \sum_i y_i \frac{\partial f_n}{\partial y_i}(t, y) = -t^2 m_{\mu\mu}(\theta_0) - 2t \sum_{i=1}^{8} y_i m_{\mu\phi_i}(\theta_0) - \sum_{i=1}^{8} \sum_{j=1}^{8} y_i y_j m_{\phi_i\phi_j}(\theta_0) + o_p(1) < 0,$$

where the negative sign follows from the fact that the Fisher Information matrix M_0 is positive definite.

By the above argument and Lemma 5 in Smith (1985), we have that with probability going to 1, f_n has a local maximum over the open set $t^2 + \|y\|^2 < 1$, so there exists a sequence of local maximizer $\hat{\theta}_n$ of $\tilde{L}_n(\theta)$ such that $\hat{\theta}_n \to_p \theta_0$ and $\|\hat{\theta}_n - \theta_0\| \leq \tau_n$, where $\tau_n \sim n^{-r}, 0 < r < 1/2$. $\square$

Proof of Theorem 3. Theorem 2 shows the existence of $\hat{\theta}_n$ with $P(\|\hat{\theta}_n - \theta_0\| \leq \tau_n) \to 1$, where $\tau_n \sim n^{-r}, 0 < r < 1/2$. By Lemma 14(a), we have that the second derivatives of $\tilde{L}_n$ are asymptotically constant in this region. The result therefore follows by standard Taylor expansion argument, Lemma 14(b) and Lemma 15. $\square$

Proof of Proposition 2. The arguments used in the proof of Proposition 2 are similar to the ones used in the proof of Theorems 2 and 3, thus we only give an outline of the proof since the actual argument is very tedious. In the following, we use δ to denote a generic small positive value and denote $\phi = (\beta_0, \beta_1, \beta_2, \beta_3, \gamma_0, \gamma_1, \gamma_2, \gamma_3)$. As in Proposition 2, $V_n = \{\theta \in \Theta | \mu \leq cQ_{n,1} + (1-c)\mu_0\}$. Note that for any $0 < c < 1$, we have $\mu_0 < cQ_{n,1} + (1-c)\mu_0 < Q_{n,1}$ and $cQ_{n,1} + (1-c)\mu_0 \searrow \mu_0$ a.s.

Denote $\Theta_n^{\delta} = \{\theta \in V_n | \|\theta - \theta_0\| \geq \delta\}$, $\Theta_n^{\mu} = \{\theta \in V_n | \|\theta - \theta_0\| \geq \delta, \mu > \mu_0\}$ and $\Theta^{\delta} = \{\theta \in V_n | \|\theta - \theta_0\| \geq \delta, \mu \leq \mu_0\}$. Note that $\Theta_n^{\delta} = \Theta_n^{\mu} \cup \Theta^{\delta}$. We first prove that,

(I) for any $\delta > 0, P\left(\sup_{\Theta_n^{\delta}} \tilde{L}_n(\theta) \geq \tilde{L}_n(\theta_0)\right) \to 0$, as $n \to \infty$.

By the same argument in Lemmas 7 and 13, it can be proved that $\sup_{\Theta_n^{\delta}} \left| \tilde{L}_n(\theta) - L_n(\theta) \right| \to_p 0$, as $n \to \infty$. By the same argument in Lemma 7, we can further prove $\sup_{\Theta_n^{\mu}} |L_n(\mu, \phi) - L_n(\mu_0, \phi)| \to_p 0$, as $n \to \infty$. Together, we have $\sup_{\Theta_n^{\delta}} \tilde{L}_n(\theta) = \sup_{\Theta_n^{\delta}} L_n(\theta) + o_p(1) = \max\left(\sup_{\Theta^{\delta}} L_n(\theta), \sup_{\Theta_n^{\mu}} L_n(\theta)\right) + o_p(1) = \max\left(\sup_{\Theta^{\delta}} L_n(\theta), \sup_{\Theta_n^{\mu}} L_n(\mu_0, \phi)\right) + o_p(1) \leq \sup_{\Theta^{\delta/2}} L_n(\theta) + o_p(1)$. The last inequality comes from the fact that $Q_{n,1} \searrow \mu_0$ a.s., so with probability going to 1, we have $\{\phi | \phi \in \Theta_n^{\mu}\} \subseteq \{\phi | \phi \in \Theta^{\delta/2}\}$. It is also easy to prove that $\tilde{L}_n(\theta_0) = L_n(\theta_0) + o_p(1) \to_p E_{\theta_0}(l_1(\theta_0))$. The rest of the proof for (I) follows from the proof of Proposition 2 in Dombry (2015), which is based on the standard compactness argument.

Denote $\Theta_n^{\delta c} = \{\theta \in V_n | \|\theta - \theta_0\| < \delta\}$, $\Theta_n^{\mu c} = \{\theta \in V_n | \|\theta - \theta_0\| < \delta, \mu > \mu_0\}$ and $\Theta^{\delta c} = \{\theta \in V_n | \|\theta - \theta_0\| < \delta, \mu \leq \mu_0\}$. Note that $\Theta_n^{\delta c} = \Theta_n^{\mu c} \cup \Theta^{\delta c}$. We now prove that there exists a $\delta^* > 0$ such that

(II) $P\left(\text{All Hessian matrices } \frac{\partial^2}{\partial \theta \partial \theta^T} \tilde{L}_n(\theta) \text{ over } \theta \in \Theta_n^{\delta^* c} \text{ is negative definite}\right) \to 1$, as $n \to \infty$.

By the same argument in Lemmas 7 and 13, we can prove that

$$\sup_{\Theta_n^{\delta c}} \left| \frac{\partial^2}{\partial \theta_i \partial \theta_j} \tilde{L}_n(\theta) - \frac{\partial^2}{\partial \theta_i \partial \theta_j} L_n(\theta) \right| \to_p 0, \text{ as } n \to \infty,$$

and

$$\sup_{\Theta_n^{\mu c}} \left| \frac{\partial^2}{\partial \theta_i \partial \theta_j} L_n(\mu, \phi) - \frac{\partial^2}{\partial \theta_i \partial \theta_j} L_n(\mu_0, \phi) \right| \to_p 0, \text{ as } n \to \infty.$$

Since $\mu \leq \mu_0$ over $\Theta^{\delta c}$, it can be proved that $\sup_{\Theta^{\delta c}} \left| \frac{\partial^2}{\partial \theta_i \partial \theta_j} l_t(\theta) \right|$ is integrable. By ergodicity of AcF and Uniform Law of Large Numbers, we have

$$\sup_{\Theta^{\delta c}} \left| \frac{\partial^2}{\partial \theta_i \partial \theta_j} L_n(\mu, \phi) - E_{\theta_0}\left(\frac{\partial^2}{\partial \theta_i \partial \theta_j} l_1(\mu, \phi)\right) \right| \to_p 0, \text{ as } n \to \infty.$$

By Lemma 4, $E_{\theta_0}\left(\frac{\partial^2}{\partial \theta \partial \theta^T} l_1(\theta_0)\right) = -M_0$ is negative definite. By the continuity of $E_{\theta_0}\left(\frac{\partial^2}{\partial \theta \partial \theta^T} l_1(\theta)\right)$ w.r.t. θ over $\Theta^{\delta c}$, we can find a $\delta^* > 0$ such that $E_{\theta_0}\left(\frac{\partial^2}{\partial \theta \partial \theta^T} l_1(\theta)\right)$ is negative definite for all $\theta \in \Theta^{\delta^* c}$. Together with the above argument, we can prove (II).

By (I), with probability going to 1, the global maximizer of $\tilde{L}_n(\theta)$ over V_n is located within $\Theta_n^{\delta^* c}$. By Theorem 2, there exists a sequence $\hat{\theta}_n$ of local maximizer of $\tilde{L}_n(\theta)$ such that $\|\hat{\theta}_n - \theta_0\| \leq \tau_n$, where $\tau_n = O_p(n^{-r}), 0 < r < 1/2$. Thus $P(\hat{\theta}_n \in \Theta_n^{\delta^* c}) \to 1$. Also, we know that $\frac{\partial \tilde{L}_n}{\partial \theta}(\hat{\theta}_n) = 0$. Together with (II) and Theorem 2.6 in Makelainen et al. (1981), we can prove Proposition 2. $\square$

A.4. First and second order partial derivative of $l_t(\theta)$

In this section, we give the formula for $\frac{\partial l_t(\theta)}{\partial \theta}$ and $\frac{\partial^2 l_t(\theta)}{\partial \theta \partial \theta^T}$. Denote $\Phi = (\gamma_0, \gamma_1, \gamma_2, \gamma_3)$, i.e. we use Φ as a generic symbol for $(\gamma_0, \gamma_1, \gamma_2, \gamma_3)$. Similarly, we set $\Psi = (\beta_0, \beta_1, \beta_2, \beta_3)$.

For the first order partial derivative, we have

$$\frac{\partial l_t(\theta)}{\partial \mu} = \frac{\alpha_t + 1}{Q_t - \mu} - \frac{\alpha_t}{\sigma_t}\left(\frac{Q_t - \mu}{\sigma_t}\right)^{-(\alpha_t+1)},$$

$$\frac{\partial l_t(\theta)}{\partial \Phi} = \left[\frac{1}{\alpha_t} - \log\left(\frac{Q_t - \mu}{\sigma_t}\right) + \left(\frac{Q_t - \mu}{\sigma_t}\right)^{-\alpha_t} \log\left(\frac{Q_t - \mu}{\sigma_t}\right)\right]\frac{\partial \alpha_t}{\partial \Phi},$$

$$\frac{\partial l_t(\theta)}{\partial \Psi} = \left[\frac{\alpha_t}{\sigma_t} - \frac{\alpha_t}{\sigma_t}\left(\frac{Q_t - \mu}{\sigma_t}\right)^{-\alpha_t}\right]\frac{\partial \sigma_t}{\partial \Psi}.$$

For the second order partial derivative, we have

$$\frac{\partial^2 l_t(\theta)}{\partial \mu^2} = (\alpha_t + 1)(Q_t - \mu)^{-2} - \alpha_t(\alpha_t + 1)\sigma_t^{\alpha_t}(Q_t - \mu)^{-(\alpha_t+2)},$$

$$\frac{\partial^2 l_t(\theta)}{\partial \mu \partial \Phi} = \left[\frac{1}{Q_t - \mu} - \sigma_t^{\alpha_t}(Q_t - \mu)^{-(\alpha_t+1)} + \alpha_t\sigma_t^{\alpha_t}(Q_t - \mu)^{-(\alpha_t+1)}\log\left(\frac{Q_t - \mu}{\sigma_t}\right)\right]\frac{\partial \alpha_t}{\partial \Psi},$$

$$\frac{\partial^2 l_t(\theta)}{\partial \mu \partial \Psi} = -\frac{\alpha_t^2}{\sigma_t^2}\left(\frac{Q_t - \mu}{\sigma_t}\right)^{-(\alpha_t+1)}\frac{\partial \sigma_t}{\partial \Psi},$$

$$\frac{\partial^2 l_t(\theta)}{\partial \Phi \partial \Psi} = \left[\frac{1}{\sigma_t} - \frac{1}{\sigma_t}\left(\frac{Q_t - \mu}{\sigma_t}\right)^{-\alpha_t} + \frac{\alpha_t}{\sigma_t}\left(\frac{Q_t - \mu}{\sigma_t}\right)^{-\alpha_t}\log\left(\frac{Q_t - \mu}{\sigma_t}\right)\right]\frac{\partial \sigma_t}{\partial \Psi}\frac{\partial \alpha_t}{\partial \Phi},$$

$$\frac{\partial^2 l_t(\theta)}{\partial \Psi_i \partial \Psi_j} = \left[\frac{\alpha_t}{\sigma_t} - \frac{\alpha_t}{\sigma_t}\left(\frac{Q_t - \mu}{\sigma_t}\right)^{-\alpha_t}\right]\frac{\partial^2 \sigma_t}{\partial \Psi_i \partial \Psi_j} + \left[-\frac{\alpha_t}{\sigma_t^2} - \frac{\alpha_t(\alpha_t - 1)}{\sigma_t^2}\left(\frac{Q_t - \mu}{\sigma_t}\right)^{-\alpha_t}\right]\frac{\partial \sigma_t}{\partial \Psi_i}\frac{\partial \sigma_t}{\partial \Psi_j},$$

$$\frac{\partial^2 l_t(\theta)}{\partial \Phi_i \partial \Phi_j} = \left[\frac{1}{\alpha_t} - \log\left(\frac{Q_t - \mu}{\sigma_t}\right) + \left(\frac{Q_t - \mu}{\sigma_t}\right)^{-\alpha_t}\log\left(\frac{Q_t - \mu}{\sigma_t}\right)\right]\frac{\partial^2 \alpha_t}{\partial \Phi_i \partial \Phi_j}$$
$$+ \left[-\frac{1}{\alpha_t^2} + \left(\frac{Q_t - \mu}{\sigma_t}\right)^{-\alpha_t}\left[\log\left(\frac{Q_t - \mu}{\sigma_t}\right)\right]^2\right]\frac{\partial \alpha_t}{\partial \Phi_i}\frac{\partial \alpha_t}{\partial \Phi_j}.$$

A.5. Observation-driven functions $\eta_1(\cdot)$, $\eta_2(\cdot)$ implied by GAS

Under the GAS framework described in Creal et al. (2013), we give the formulas of $\eta_1(\cdot)$ and $\eta_2(\cdot)$ implied by GAS in the dynamic GEV context. Set $\tau_t = \log \alpha_t$ and $\zeta_t = \log \sigma_t$ to ensure the positivity of parameters. Given (ζ_t, τ_t), the conditional distribution of Q_t is Fréchet$(\mu, \exp(\zeta_t), \exp(\tau_t))$. The log-likelihood function $l_t(\cdot)$ of Q_t is

$$l_t(Q_t|\mu, \zeta_t, \tau_t) = \tau_t + \exp(\tau_t)\zeta_t - (\exp(\tau_t) + 1)\log(Q_t - \mu) - \left(\frac{Q_t - \mu}{\exp(\zeta_t)}\right)^{-\exp(\tau_t)}.$$

To derive GAS, we need to obtain the score function of $l_t(Q_t|\mu, \zeta_t, \tau_t)$ w.r.t. (ζ_t, τ_t), which is

$$\frac{\partial l_t}{\partial \zeta_t} = \exp(\tau_t) - \exp(\tau_t)\left(\frac{Q_t - \mu}{\exp(\zeta_t)}\right)^{-\exp(\tau_t)},$$

$$\frac{\partial l_t}{\partial \tau_t} = 1 - \exp(\tau_t)\log\left(\frac{Q_t - \mu}{\exp(\zeta_t)}\right) + \exp(\tau_t)\left(\frac{Q_t - \mu}{\exp(\zeta_t)}\right)^{-\exp(\tau_t)}\log\left(\frac{Q_t - \mu}{\exp(\zeta_t)}\right).$$

Following the recommendation in Creal et al. (2013), we take the scaling matrix function $S_t = I$ and obtain the following GAS model,

$$Q_t = \mu + \sigma_t Y_t^{1/\alpha_t}$$

$$\log \sigma_t = \beta_0 + \beta_1 \log \sigma_{t-1} + \beta_2 \frac{\partial l_{t-1}}{\partial \zeta_{t-1}}$$

$$\log \alpha_t = \gamma_0 + \gamma_1 \log \alpha_{t-1} + \gamma_2 \frac{\partial l_{t-1}}{\partial \tau_{t-1}},$$

where we assume $0 \leq \beta_1, \gamma_1 < 1$ and $\beta_2, \gamma_2 > 0$. Notice that the evolution scheme given by GAS is complicated. It is easy to see that $\eta_1(Q_{t-1}) = \beta_2 \frac{\partial l_{t-1}}{\partial \zeta_{t-1}}$ is a monotone increasing function of Q_{t-1}, which is expected for volatility clustering. It is also easy to prove that $\eta_2(Q_{t-1}) = \gamma_2 \frac{\partial l_{t-1}}{\partial \tau_{t-1}}$ is an increasing function of Q_{t-1} when $\frac{Q_{t-1}-\mu}{\sigma_{t-1}} < 1$ and decreasing when $\frac{Q_{t-1}-\mu}{\sigma_{t-1}} > 1$, which is quite counter-intuitive in terms of econometric meaning. On the contrary, for AcF, a larger Q_{t-1} always gives a smaller α_t.

By the fact that $\frac{\partial l_t}{\partial \zeta_t} \leq \exp(\tau_t)$ and $\frac{\partial l_t}{\partial \tau_t} \leq 1$, it is easy to see that (σ_t, α_t) of the GAS model is upper bounded.

References

Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. Econometrica 70 (1), 191–221.
Bali, T., Weinbaum, D., 2007. A conditional extreme value volatility estimator based on high-frequency returns. J. Econom. Dynam. Control 31 (2), 361–397.
Balkema, A., de Haan, L., 1974. Residual life time at great age. Ann. Probab. 2, 792–804.
Basrak, B., Davis, R.A., Mikosch, T., 2002. A characterization of multivariate regular variation. Ann. Appl. Probab. 12, 908–920.
Basrak, B., Segers, J., 2009. Regularly varying multivariate time series. Stochastic Process. Appl. 119 (4), 1055–1080.
Billingsley, P., 1961. The Lindeberg-Levy theorem for martingales. Proc. Amer. Math. Soc. 12, 788–792.
Billio, M., Getmansky, M., Lo, A.W., Pelizzon, L., 2012. Econometric measures of connectedness and systemic risk in the finance and insurance sectors. J. Financ. Econ. 104, 535–559.
Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. J. Econometrics 31, 307–327.
Boutahar, M., Dufrénot, G., Péguin-Feissolle, A., 2008. A simple fractionally integrated model with a time-varying long memory parameter d_t. Comput. Econ. 31, 225–241.
Bücher, A., Segers, J., 2016. On the maximum likelihood estimator for the generalized extreme-value distribution, arxiv.
Bücher, A., Segers, J., 2017. Maximum likelihood estimation for the Fréchet distribution based on block maxima extracted from a time series. Bernoulli.
Chan, K., Tong, H., 1994. A note on noisy chaos. J. R. Stat. Soc. Ser. B Stat. Methodol. 56 (2), 301–311.
Chavez-Demoulin, V., Embrechts, P., Sardy, S., 2014. Extreme-quantile tracking for financial time series. J. Econometrics 188 (1), 44–52.
Chen, X., Fan, Y., 2006. Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification. J. Econometrics 135 (12), 125–154.
Coles, S., 2001. In: Bickel, P., Diggle, P., Fienberg, S., Gather, U., Zeger, S. (Eds.), An Introduction to Statistical Modeling of Extreme Values. In: Springer Series in Statistics, Springer.
Creal, D., Koopman, S., Lucas, A., 2013. Generalized autoregressive score models with applications. J. Appl. Econometrics 28 (5), 777–795.
Davis, R.A., Mikosch, T., 2009a. Extreme value theory for GARCH processes. In: Mikosch, T., Krei, J.P., Davis, R., Andersen, T. (Eds.), Handbook of Financial Time Series. Springer, Berlin, Heidelberg.
Davis, R.A., Mikosch, T., 2009b. Extremes of stochastic volatility models. In: Mikosch, T., Krei, J.P., Davis, R., Andersen, T. (Eds.), Handbook of Financial Time Series. Springer, Berlin, Heidelberg.
Davison, A., Smith, R., 1990. Models for exceedances over high thresholds. J. R. Stat. Soc. Ser. B Stat. Methodol. 52 (3), 393–442.
Diebold, F., Schuermann, T., Stroughair, J., 1998. Pitfalls and opportunities in the use of extreme value theory in risk management. In: Refenes, A., Moody, J., Burgess, A. (Eds.), Decision Technologies for Computational Finance. In: Advances in Computational Management Science, vol. 2, Springer US, pp. 3–12.
Diebold, F.X., Yilmaz, K., 2009. Measuring financial asset return and volatility spillovers, with application to global equity markets. Econ. J. 119, 158–171.
Diebold, F., Yilmaz, K., 2014. On the network topology of variance decompositions: Measuring the connectedness of financial firms. J. Econometrics 182 (1), 119–134.
Dombry, C., 2015. Existence and consistency of the maximum likelihood estimators for the extreme value index within the block maxima framework. Bernoulli 21 (1), 420–436.
Drees, H., Segers, J., Warchol, M., 2015. Statistics for tail processes of Markov chains. Extremes 18 (3).
Embrechts, P., Resnick, S.I., Samorodnitsky, G., 1999. Extreme value theory as a risk management tool. N. Am. Actuar. J. 3, 30–41.
Engle, R.F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of UK inflation. Econometrica 50, 987–1007.
Engle, R., Russell, J., 1998. Autoregressive conditional duration: a new model for irregularly spaced transaction data. Econometrica 66, 1127–1162.
Fama, E., French, K., 1993. Common risk factors in the returns on stocks and bonds. J. Financ. Econ. 3–56.
Ferreira, A., de Haan, L., 2015. On the block maxima method in extreme value theory: PWM estimators. Ann. Statist. 43 (1), 276–298.
Francq, C., Zakoian, J., 2004. Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. Bernoulli 10 (4), 605–637.
Geweke, J., 1977. The dynamic factor analysis of economic time series. In: Aigner, D., Goldberger, A. (Eds.), Latent Variables in Socio-Economic Models. North-Holland, Amsterdam.
Hall, S., Swamy, P., Tavlas, G., 2016. Time-varying coefficient models: A proposal for selecting the coefficient driver sets. Macroecon. Dynam. 1–17.
Hansen, B., 1994. Autoregressive conditional density estimation. Internat. Econom. Rev. 35 (3), 705–730.
Harvey, A., 2013. Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series. Cambridge University Press, Cambridge, UK.
Harvey, A., Chakravarty, T., 2008. Beta-t-(E)GARCH, Discussion Paper. University of Cambridge.
Kelly, B., 2014. The dynamic power law model. Extremes 17 (4), 557–583.
Kelly, B., Jiang, H., 2014. Tail risk and asset prices. Rev. Financ. Stud. 27 (10), 2841–2871.
Kratz, M., Lok, Y.H., McNeil, A.J., 2016. Multinomial VaR Backtests: A simple implicit approach to backtesting expected shortfall, arXiv:1611.04851.
Kritzman, M., Li, Y., Page, S., Rigobon, R., 2011. Principal components as a measure of systemic risk. J. Portfolio Manage. 37 (4), 112–126.
Lam, C., Yao, Q., 2012. Factor modeling for high-dimensional time series: Inference for the number of factors. Ann. Statist. 40 (2), 694–726.
Laurini, F., Tawn, J.A., 2009. Regular variation and extremal dependence of GARCH residuals with application to market risk measures. Econometric Rev. 28, 146–169.
Leadbetter, M., Lindgren, G., Rootzen, H., 1983. Extremes and Related Properties of Random Sequences and Processes. Springer.
Lundbergh, S., Terasvirta, T., Van Dijk, D., 2003. Time-varying smooth transition autoregressive models. J. Bus. Econom. Statist. 21 (1), 104–128.
Makelainen, T., Schmidt, K., Styan, G., 1981. On the existence and uniqueness of the maximum likelihood estimate of a vector-valued parameter in fixed-size samples. Ann. Statist. 9, 758–767.
Malinowski, A., Schlather, M., Zhang, Z., 2015. Marked point process adjusted tail dependence analysis for high frequency financial data. Stat. Interface 8, 109–122.
Massacci, D., 2016. Tail risk dynamics in stock returns: Links to the macroeconomy and global markets connectedness. Manage. Sci. 1–18.
McNeil, A., Frey, R., 2000. Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. J. Empir. Finance 7 (34), 271–300.

Mossin, J., 1966. Equilibrium in a capital asset market. Econometrica 34 (4), 768–783.

Picklands, J., 1975. Statistical inference using extreme order statistics. Ann. Statist. 3 (1), 119–131.

Resnick, S.I., 1987. Extreme Values, Regular Variation and Point Processes. In: Springer Series in Operations Research and Financial Engineering, Springer.

Sharpe, W., 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. J. Finance 19 (3), 425–442.

Smith, R., 1985. Maximum likelihood estimation in a class of nonregular cases. Biometrika 72 (1), 67–90.

Smith, R., Goodman, D., 2000. Bayesian risk analysis, extremes and integrated risk management. Risk Books.

Stock, J., Watson, M., 2011. Dynamic factor models. In: Clements, M.P., Hendry, D.F. (Eds.), The Oxford Handbook of Economic Forecasting. Oxford University Press.

van Oordt, M., Zhou, C., 2016. Systematic tail risk. J. Financ. Quant. Anal. 51 (2), 685–705.

Zhang, X., Schwaab, B., 2017. Tail Risk in Government Bond Markets and ECB Asset Purchases, Working Paper.

Zhang, Z., Zhu, B., 2016. Copula structured M4 processes with application to high-Frequency financial data. J. Econometrics 194 (2), 231–241.