Decoupling Systemic Risk into Endopathic and Exopathic Competing Risks Through Autoregressive Conditional Accelerated Fréchet Model

Jingyu Ji[a], Deyuan Li[b], Zhengjun Zhang[c,d,e]

[a] *Capital University of Economics and Business*, [b] *Fudan University*, [c] *University of Chinese Academy of Sciences*, [d] *Chinese Academy of Sciences*, [e] *University of Wisconsin*

Supplementary Material

In the supplementary material, we present the detailed proofs of Theorem 1, Proposition 1, Theorem 2 and 3, and Proposition 2, along with the auxiliary lemmas used in the proofs, and the expressions of the first order and the second order partial derivatives of the likelihood function in the paper, as well as three more simulation studies. We note that our new AcAF model is advanced from the AcF model, i.e., like the advance of GARCH model from ARCH model. Our proofs here have connections to the proofs in the AcF model, but due to adding an additional latent Fréchet process, the complexity level is much higher than the original proofs in the AcF, and as a result, the proofs in this supplementay file are nontrivial.

Corresponding authors: Deyuan Li, Department of Statistics and Data Science, School of Management, Fudan University, Shanghai, China. E-mail: deyuanli@fudan.edu.cn. Zhengjun Zhang, School of Economics and Management, and MOE Social Science Laboratory of Digital Economic Forecasts and Policy Simulation, University of Chinese Academy of Sciences, Beijing, China; Center for Forecasting Sciences, Chinese Academy of Sciences, Beijing, China; Department of Statistics, University of Wisconsin, Madison, WI, USA. E-mail: zhangzhengjun@ucas.ac.cn.

S1 Proof of Theorem 1

The proof of Theorem 1 is built upon some conclusions in Chan and Tong (1994) in the non-linear dynamic system. In the following, we assume that $\{\sigma_t, \alpha_{1t}, \alpha_{2t}\}$ comes from the model with $\boldsymbol{\theta} \in \Theta$ as specified in Theorem 1. Without loss of generality, we set the location parameter μ as 0. In our model, $\{\log \sigma_t, \log \alpha_{1t}, \log \alpha_{2t}\}$ forms a non-linear dynamic system according to the equations (2.6)-(2.8).

To fit $\{\sigma_t, \alpha_{1t}, \alpha_{2t}\}$ into the framework of Chan and Tong (1994), we reparameterize the autoregressive equations as follows:

$$\begin{aligned}
\log \sigma_t &= [\beta_0 - z_1 + \beta_1 \log \sigma_{t-1}] + [z_1 - \beta_2 \exp(-\beta_3(\sigma_{t-1} \max(Y_{1,t-1}^{1/\alpha_{1,t-1}}, Y_{2,t-1}^{1/\alpha_{2,t-1}})))], \\
\log \alpha_{1t} &= [\gamma_0 + z_2 + \gamma_1 \log \alpha_{1,t-1}] + [\gamma_2 \exp(-\gamma_3(\sigma_{t-1} \max(Y_{1,t-1}^{1/\alpha_{1,t-1}}, Y_{2,t-1}^{1/\alpha_{2,t-1}}))) - z_2], \\
\log \alpha_{2t} &= [\delta_0 + z_3 + \delta_1 \log \alpha_{2,t-1}] + [\delta_2 \exp(-\delta_3(\sigma_{t-1} \max(Y_{1,t-1}^{1/\alpha_{1,t-1}}, Y_{2,t-1}^{1/\alpha_{2,t-1}}))) - z_3],
\end{aligned}$$

where z_1 is a positive constant such that $0 < z_1 < \beta_2$ (e.g., we can set $z_1 = \beta_2/2$), z_2 is a positive constant such that $0 < z_2 = \gamma_2 \exp(\frac{\gamma_3}{\beta_3} \log(\frac{z_1}{\beta_2})) < \gamma_2$, and z_3 is a positive constant such that $0 < z_3 = \delta_2 \exp(\frac{\delta_3}{\beta_3} \log(\frac{z_1}{\beta_2})) < \delta_2$. The reason for defining z_1, z_2, z_3 as above will be made more clear in the proof of Lemma 2. Let $\mathbf{X}_t = (\log \sigma_t, \log \alpha_{1t}, \log \alpha_{2t})^T$ and

$$\begin{aligned}
T(\mathbf{X}_{t-1}) &= (\beta_0 - z_1 + \beta_1 \log \sigma_{t-1}, \gamma_0 + z_2 + \gamma_1 \log \alpha_{1,t-1}, \delta_0 + z_3 + \delta_1 \log \alpha_{2,t-1})^T, \\
S(\mathbf{X}_{t-1}, Y_{1,t-1}, Y_{2,t-1}) &= \Big(z_1 - \beta_2 \exp(-\beta_3(\sigma_{t-1} \max(Y_{1,t-1}^{1/\alpha_{1,t-1}}, Y_{2,t-1}^{1/\alpha_{2,t-1}}))), \\
&\qquad \gamma_2 \exp(-\gamma_3(\sigma_{t-1} \max(Y_{1,t-1}^{1/\alpha_{1,t-1}}, Y_{2,t-1}^{1/\alpha_{2,t-1}}))) - z_2, \\
&\qquad \delta_2 \exp(-\delta_3(\sigma_{t-1} \max(Y_{1,t-1}^{1/\alpha_{1,t-1}}, Y_{2,t-1}^{1/\alpha_{2,t-1}}))) - z_3\Big)^T.
\end{aligned}$$

We can rewrite the non-linear dynamic system of $\mathbf{X}_t$ as

$$\mathbf{X}_t = T(\mathbf{X}_{t-1}) + S(\mathbf{X}_{t-1}, Y_{1,t-1}, Y_{2,t-1}),$$

where $\{Y_{1t}\}$ and $\{Y_{2t}\}$ are two independent sequences of i.i.d. unit Fréchet random variables.

Following the terminologies in Chan and Tong (1994), we obtain that $T(\cdot)$ admits a compact attractor $\Lambda = (\frac{\beta_0 - z_1}{1-\beta_1}, \frac{\gamma_0 + z_2}{1-\gamma_1}, \frac{\delta_0 + z_3}{1-\delta_1})^T$, which is a singleton in $\mathbb{R}^3$, and the domain of attraction for Λ is $\mathbb{R}^3$. In other words, for any $\mathbf{x} \in \mathbb{R}^3$, we have the iterates $T^n(\mathbf{x}) \to \Lambda$ as $n \to \infty$. We further set $G = (\frac{\beta_0 - \beta_2}{1-\beta_1}, \frac{\beta_0}{1-\beta_1}) \times (\frac{\gamma_0}{1-\gamma_1}, \frac{\gamma_0+\gamma_2}{1-\gamma_1}) \times (\frac{\delta_0}{1-\delta_1}, \frac{\delta_0+\delta_2}{1-\delta_1})$, which is an open area in $\mathbb{R}^3$.

Then we are able to prove that the process $\mathbf{X}_t$ satisfies five conditions (named (a)-(e)) given in Theorem 1 of Chan and Tong (1994). Condition (a) (Λ has a dense orbit) is proved, because for any $\mathbf{x}$ in $\mathbb{R}^3$, $T^n(\mathbf{x}) \to \Lambda$ as $n \to \infty$ by the above argument. Condition (c) (Lipschitz continuous over G) is satisfied because $T(\cdot)$ is a Lipschitz continuous function. Next, we will verify the condition (b) (exponentially attracting) by leveraging our conclusion from Lemma 1.

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

Proof of Lemma 1. We only prove the result for $\log \alpha_{1t}$, the proofs for $\log \sigma_t$ and $\log \alpha_{2t}$ are similar. If $\log \alpha_{1t} > \frac{\gamma_0}{1-\gamma_1}$, then we have $\log \alpha_{1,t+1} = \gamma_0 + \gamma_1 \log \alpha_{1t} + \gamma_2 \exp(-\gamma_3 Q_t) > \gamma_0 + \gamma_1 \frac{\gamma_0}{1-\gamma_1} = \frac{\gamma_0}{1-\gamma_1}$. Similarly, we can prove that $\log \alpha_{1,t+1} < \gamma_0 + \gamma_1 \frac{\gamma_0+\gamma_2}{1-\gamma_1} + \gamma_2 = \frac{\gamma_0+\gamma_2}{1-\gamma_1}$ if $\log \alpha_{1t} < \frac{\gamma_0+\gamma_2}{1-\gamma_1}$. □

For the remaining part, we need to check the conditions (d) and (e), which are demonstrated by Lemma 2 and Lemma 3 below, respectively.

Lemma 2. *For any $\mathbf{x} \in G$, 0 is in the support of $|S(\mathbf{x}, Y_{1,t-1}, Y_{2,t-1})|$ where $|\cdot|$ is the norm of the vector, there exists a continuous and positive function $r(\mathbf{x})$ for $\mathbf{x} \in G$, such that the third step transition probability for $\mathbf{X}_t$, $P^3(\mathbf{x}, d\mathbf{y})$, has an absolutely continuous component whose p.d.f. is positive over $B(T^3(\mathbf{x}), r(\mathbf{x}))$ where $B(\mathbf{x}, r)$ denotes the open ball in G with center at $\mathbf{x}$ and radius equal to r.*

Proof of Lemma 2. Since $\sigma_{t-1}, \alpha_{1,t-1}, \alpha_{2,t-1} > 0$ and $0 < Y_{1,t-1}, Y_{2,t-1} < \infty$, it is easy to prove that for any $\mathbf{X}_{t-1}$, there always exists $(Y^*_{1,t-1}, Y^*_{2,t-1})$ depending on $\mathbf{X}_{t-1}$ such that

$$Q^*_{t-1} = \sigma_{t-1} \max\left((Y^*_{1,t-1})^{1/\alpha_{1,t-1}}, (Y^*_{2,t-1})^{1/\alpha_{2,t-1}}\right) = -\frac{1}{\beta_3}\log\left(\frac{z_1}{\beta_2}\right).$$

By the values of z_1, z_2 and z_3 defined above, we obtain that given $\mathbf{X}_{t-1}$, there exist $Y^*_{1,t-1}$ and $Y^*_{2,t-1}$ that make $|S(\mathbf{X}_{t-1}, Y^*_{1,t-1}, Y^*_{2,t-1})| = 0$. Hence for any $\mathbf{x} \in G$, 0 is in the support of $|S(\mathbf{x}, Y_{1,t-1}, Y_{2,t-1})|$. In addition, we denote $Q^* = -\frac{1}{\beta_3}\log(\frac{z_1}{\beta_2})$.

Next we verify that there exists a positive function $r(\mathbf{x})$ such that $P^3(\mathbf{x}, d\mathbf{y})$ has an absolutely continuous component whose p.d.f. is positive over $B(T^3(\mathbf{x}), r(\mathbf{x}))$. Given $\mathbf{X}_{t-1}$, for $\mathbf{X}_{t+2} = (\log\sigma_{t+2}, \log\alpha_{1,t+2}, \log\alpha_{2,t+2})^T$, we obtain the following equations:

$$\begin{aligned}
\log\sigma_{t+2} &= [\beta_0 - z_1 + \beta_1\log\sigma_{t+1}] + [z_1 - \beta_2\exp(-\beta_3 Q_{t+1})] \\
&= \beta_0 - z_1 + \beta_1\big\{\beta_0 - z_1 + \beta_1[\beta_0 - z_1 + \beta_1\log\sigma_{t-1}] + \beta_1[z_1 - \beta_2\exp(-\beta_3 Q_{t-1})] \\
&\quad + [z_1 - \beta_2\exp(-\beta_3 Q_t)]\big\} + [z_1 - \beta_2\exp(-\beta_3 Q_{t+1})] \\
&= \beta_0 - z_1 + \beta_1\big\{\beta_0 - z_1 + \beta_1[\beta_0 - z_1 + \beta_1\log\sigma_{t-1}]\big\} + \beta_1^2[z_1 - \beta_2\exp(-\beta_3 Q_{t-1})] \\
&\quad + \beta_1[z_1 - \beta_2\exp(-\beta_3 Q_t)] + [z_1 - \beta_2\exp(-\beta_3 Q_{t+1})] \\
&= T^3[\mathbf{X}_{t-1}][1] + \beta_1^2[z_1 - \beta_2\exp(-\beta_3 Q_{t-1})] + \beta_1[z_1 - \beta_2\exp(-\beta_3 Q_t)] \\
&\quad + [z_1 - \beta_2\exp(-\beta_3 Q_{t+1})].
\end{aligned}$$

Similarly,

$$\begin{aligned}
\log\alpha_{1,t+2} &= T^3[\mathbf{X}_{t-1}][2] + \gamma_1^2[\gamma_2\exp(-\gamma_3 Q_{t-1}) - z_2] + \gamma_1[\gamma_2\exp(-\gamma_3 Q_t) - z_2] \\
&\quad + [\gamma_2\exp(-\gamma_3 Q_{t+1}) - z_2], \\
\log\alpha_{2,t+2} &= T^3[\mathbf{X}_{t-1}][3] + \delta_1^2[\delta_2\exp(-\delta_3 Q_{t-1}) - z_3] + \delta_1[\delta_2\exp(-\delta_3 Q_t) - z_3] \\
&\quad + [\delta_2\exp(-\delta_3 Q_{t+1}) - z_3],
\end{aligned}$$

where $T^3[\mathbf{X}_{t-1}][1], T^3[\mathbf{X}_{t-1}][2]$ and $T^3[\mathbf{X}_{t-1}][3]$ denote the first, second and third components of $T^3(\mathbf{X}_{t-1})$, respectively. $\mathbf{X}_{t+2}$ is a function of Q_{t-1}, Q_t and Q_{t+1} given $\mathbf{X}_{t-1}$. We denote $\mathbf{X}_{t+2} = \mathbf{F}_{\mathbf{X}_{t-1}}(Q_{t-1}, Q_t, Q_{t+1})$. When $(Q^*_{t-1}, Q^*_t, Q^*_{t+1})^T = (Q^*, Q^*, Q^*)^T$, it follows that $\mathbf{X}_{t+2} = \mathbf{F}_{\mathbf{X}_{t-1}}(Q^*, Q^*, Q^*) = T^3(\mathbf{X}_{t-1})$ and the determinant of the Jacobian matrix of $\mathbf{X}_{t+2}$ at $(Q^*, Q^*, Q^*)^T$ is $\beta_2\beta_3\gamma_2\gamma_3\delta_2\delta_3 \exp(-(\beta_3 + \gamma_3 + \delta_3)Q^*)(\gamma_1 - \delta_1)(\beta_1 - \delta_1)(\beta_1 - \gamma_1)$, which is not zero since $\boldsymbol{\theta} \in \Theta$ and $\beta_1 \neq \gamma_1 \neq \delta_1$.

By the inverse function theorem, we know that there exists an open neighborhood at $\mathbf{X}_{t+2} = \mathbf{F}_{\mathbf{X}_{t-1}}(Q^*, Q^*, Q^*) = T^3[\mathbf{X}_{t-1}]$, denoted by $B(T^3(\mathbf{X}_{t-1}), r(\mathbf{X}_{t-1}))$ and another open neighborhood at $(Q^*, Q^*, Q^*)^T$ such that there is a one-to-one map between them.

Consider the sample space $\Omega = \{\boldsymbol{\omega}\}$, where $\boldsymbol{\omega} = (\omega_1, \omega_2, \omega_3)^T$ and $\triangle\boldsymbol{\omega} = (\triangle\omega_1, \triangle\omega_2, \triangle\omega_3)^T$. We need to prove that, for $\boldsymbol{\omega} \in B(T^3(\mathbf{X}_{t-1}), r(\mathbf{X}_{t-1}))$, $\lim_{\triangle\boldsymbol{\omega}\to\mathbf{0}} \frac{P(\boldsymbol{\omega}\leq\mathbf{X}_{t+2}\leq\boldsymbol{\omega}+\triangle\boldsymbol{\omega}|\mathbf{X}_{t-1}=\mathbf{x})}{\|\triangle\boldsymbol{\omega}\|}$ exists and is positive. We have

$$
\begin{aligned}
&P\left(\boldsymbol{\omega} \leq \mathbf{X}_{t+2} \leq \boldsymbol{\omega} + \triangle\boldsymbol{\omega} \mid \mathbf{X}_{t-1} = \mathbf{x}\right) \\
=&P\left(\boldsymbol{\omega} \leq F_{\mathbf{X}_{t-1}}(Q_{t-1}, Q_t, Q_{t+1}) \leq \boldsymbol{\omega} + \triangle\boldsymbol{\omega} \mid \mathbf{X}_{t-1} = \mathbf{x}\right) \\
=&P\Bigg(\boldsymbol{\omega} \leq T^3(\mathbf{x}) + \frac{\partial F_{\mathbf{X}_{t-1}}}{\partial Q_{t-1}}(Q_{t-1} - Q^*) + \frac{\partial F_{\mathbf{X}_{t-1}}}{\partial Q_t}(Q_t - Q^*) + \frac{\partial F_{\mathbf{X}_{t-1}}}{\partial Q_{t+1}}(Q_{t+1} - Q^*) \\
&\qquad \leq \boldsymbol{\omega} + \triangle\boldsymbol{\omega} \mid \mathbf{X}_{t-1} = \mathbf{x}\Bigg) \qquad\qquad \text{(S1.1)} \\
=&P\left(\mathbf{A}^{-1}(\boldsymbol{\omega} - T^3(\mathbf{x})) \leq \begin{pmatrix} Q_{t-1} - Q^* \\ Q_t - Q^* \\ Q_{t+1} - Q^* \end{pmatrix} \leq \mathbf{A}^{-1}(\boldsymbol{\omega} + \triangle\boldsymbol{\omega} - T^3(\mathbf{x})) \mid \mathbf{X}_{t-1} = \mathbf{x}\right),
\end{aligned}
$$

where

$$
\mathbf{A} = \begin{pmatrix}
\frac{\partial F_{\mathbf{X}_{t-1}}(Q^*,Q^*,Q^*)[1]}{\partial Q_{t-1}} & \frac{\partial F_{\mathbf{X}_{t-1}}(Q^*,Q^*,Q^*)[1]}{\partial Q_t} & \frac{\partial F_{\mathbf{X}_{t-1}}(Q^*,Q^*,Q^*)[1]}{\partial Q_{t+1}} \\
\frac{\partial F_{\mathbf{X}_{t-1}}(Q^*,Q^*,Q^*)[2]}{\partial Q_{t-1}} & \frac{\partial F_{\mathbf{X}_{t-1}}(Q^*,Q^*,Q^*)[2]}{\partial Q_t} & \frac{\partial F_{\mathbf{X}_{t-1}}(Q^*,Q^*,Q^*)[2]}{\partial Q_{t+1}} \\
\frac{\partial F_{\mathbf{X}_{t-1}}(Q^*,Q^*,Q^*)[3]}{\partial Q_{t-1}} & \frac{\partial F_{\mathbf{X}_{t-1}}(Q^*,Q^*,Q^*)[3]}{\partial Q_t} & \frac{\partial F_{\mathbf{X}_{t-1}}(Q^*,Q^*,Q^*)[3]}{\partial Q_{t+1}}
\end{pmatrix},
$$

and the determinant of matrix $\mathbf{A}$ is not equal to 0. Recall that $Q_t = \sigma_t \max(Y_{1,t}^{1/\alpha_{1,t}}, Y_{2,t}^{1/\alpha_{2,t}})$.

Then the right hand side of (S1.1) can be written as

$$P\Bigg(\mathbf{A}^{-1}(\boldsymbol{\omega}-T^3(\mathbf{x}))+Q^* \le \begin{pmatrix} \sigma_{t-1}\max(Y_{1,t-1}^{1/\alpha_{1,t-1}}, Y_{2,t-1}^{1/\alpha_{2,t-1}}) \\ \sigma_t \max(Y_{1,t}^{1/\alpha_{1,t}}, Y_{2,t}^{1/\alpha_{2,t}}) \\ \sigma_{t+1}\max(Y_{1,t+1}^{1/\alpha_{1,t+1}}, Y_{2,t+1}^{1/\alpha_{2,t+1}}) \end{pmatrix} \le \mathbf{A}^{-1}(\boldsymbol{\omega}+\triangle\boldsymbol{\omega}-T^3(\mathbf{x}))+Q^* \mid \mathbf{X}_{t-1}=\mathbf{x}\Bigg).$$

Since $Y_{1,t-1}, Y_{2,t-1}, Y_{1,t}, Y_{2,t}, Y_{1,t+1}, Y_{2,t+1}$ are i.i.d. continuous unit Fréchet random variables, we obtain that $\lim_{\triangle\boldsymbol{\omega}\to\mathbf{0}} \frac{P(\boldsymbol{\omega}\le\mathbf{X}_{t+2}\le\boldsymbol{\omega}+\triangle\boldsymbol{\omega}|\mathbf{X}_{t-1}=\mathbf{x})}{\|\triangle\boldsymbol{\omega}\|}$ exists and is positive for $\boldsymbol{\omega}\in B(T^3(\mathbf{X}_{t-1}), r(\mathbf{X}_{t-1}))$. Therefore, $P^3(\mathbf{x}, d\mathbf{y})$ has an absolutely continuous component whose p.d.f. is positive over $B(T^3(\mathbf{x}), r(\mathbf{x}))$. □

Lemma 3. $E\left(|S(\mathbf{X}_{t-1}, Y_{1,t-1}, Y_{2,t-1})| \mid \mathbf{X}_{t-1}\right)$ *is uniformly bounded above for* $\mathbf{X}_{t-1}\in G$.

Proof of Lemma 3. Notice that

$$\begin{aligned} |S(\mathbf{X}_{t-1}, Y_{1,t-1}, Y_{2,t-1})| =& \Big\{[z_1-\beta_2\exp(-\beta_3(\sigma_{t-1}\max(Y_{1,t-1}^{1/\alpha_{1,t-1}}, Y_{2,t-1}^{1/\alpha_{2,t-1}})))]^2 \\ &+[\gamma_2\exp(-\gamma_3(\sigma_{t-1}\max(Y_{1,t-1}^{1/\alpha_{1,t-1}}, Y_{2,t-1}^{1/\alpha_{2,t-1}})))-z_2]^2 \\ &+[\delta_2\exp(-\delta_3(\sigma_{t-1}\max(Y_{1,t-1}^{1/\alpha_{1,t-1}}, Y_{2,t-1}^{1/\alpha_{2,t-1}})))-z_3]^2\Big\}^{1/2} \\ \le& |z_1|+|\beta_2|+|z_2|+|\gamma_2|+|z_3|+|\gamma_3|, \end{aligned}$$

which is uniformly bounded above for $\mathbf{X}_{t-1}\in G$. □

Now we have verified that all five conditions of Theorem 1 in Chan and Tong (1994) are satisfied under the AcAF model. Hence $\{\sigma_t, \alpha_{1t}, \alpha_{2t}\}$, as a Markov chain on $G\subset\mathbb{R}^3$, is stationary and geometrically ergodic.

S2 Proof of Proposition 1

Given $\mathcal{F}_{t-1}$, by Proposition 1 in Zhao et al. (2018), we have, for $k = 1, 2$,

$$\frac{X_{k,t} - b_{k,p_k,t}}{a_{k,p_k,t}} \xrightarrow{d} \Psi_{\alpha_{kt}},$$

as $p_k \to \infty$, where $\Psi_{\alpha_{kt}}(x) = \exp(-x^{-\alpha_{kt}})$ denotes the distribution of a Fréchet type random variable with tail index $\alpha_{kt} > 0$.

We only illustrate the proof for Case 1 (i.e., $\alpha_{1t} < \alpha_{2t}$). The proofs for Cases 2 and 3 are similar. Recall that $a_{1,p_1,t} = (\sum_{i=1}^{p_1} \sigma_{it}^{\alpha_{1t}})^{1/\alpha_{1t}}$, $a_{2,p_2,t} = (\sum_{j=1}^{p_2} \tilde{\sigma}_{jt}^{\alpha_{2t}})^{1/\alpha_{2t}}$, $b_{1,p_1,t} = b_{2,p_2,t} = 0$, and that given $\mathcal{F}_{t-1}$, X_{1t} and X_{2t} are independent.

1. Since $p_1/p_2 \to C > 0$ or ∞, and $\sigma_{i,t}$, $\tilde{\sigma}_{j,t}$ are bounded, it is easy to show that $a_{1,p_1,t}/a_{2,p_2,t} \to \infty$. Then as $p_1, p_2 \to \infty$,

$$\begin{aligned} P\left(\frac{Q_t - b_{1,p_1,t}}{a_{1,p_1,t}} \le x\right) &= P\left(\max_{1\le i\le p_1}\{X_{1,i,t}\} \le a_{1,p_1,t}x\right) P\left(\max_{1\le j\le p_2}\{X_{2,j,t}\} \le a_{1,p_1,t}x\right) \\ &= P\left(\max_{1\le i\le p_1}\{X_{1,i,t}\} \le a_{1,p_1,t}x\right) P\left(\frac{\max_{1\le j\le p_2}\{X_{2,j,t}\}}{a_{2,p_2,t}} \le \frac{a_{1,p_1,t}}{a_{2,p_2,t}}x\right) \\ &\to \Psi_{\alpha_{1t}}(x). \end{aligned}$$

   where the limit follows by Theorem 2.2 in Cao and Zhang (2021).

2. By $p_1/p_2 \to 0$, $a_{1,p_1,t}/a_{2,p_2,t} \to a_t > 0$ and Theorem 2.2 in Cao and Zhang (2021), it follows that

$$\begin{aligned} P\left(\frac{Q_t - b_{1,p_1,t}}{a_{1,p_1,t}} \le x\right) &= P\left(\max_{1\le i\le p_1}\{X_{1,i,t}\} \le a_{1,p_1,t}x\right) P\left(\max_{1\le j\le p_2}\{X_{2,j,t}\} \le a_{1,p_1,t}x\right) \\ &\sim P\left(\max_{1\le i\le p_1}\{X_{1,i,t}\} \le a_{1,p_1,t}x\right) P\left(\max_{1\le j\le p_2}\{X_{2,j,t}\} \le a_{2,p_2,t}a_t x\right) \\ &\to \Psi_{\alpha_{1t}}(x)\Psi_{\alpha_{2t}}(a_t x). \end{aligned}$$

The proof of Proposition 1 is complete.

S3 Proof of consistency and asympototic normality

To prove Theorems 2, 3 and Proposition 2 in the paper, we first give Lemmas 4-16 and their proofs. Part of the proofs follows that in Francq et al. (2004) and Zhao et al. (2018).

In the following, we assume the conditions in Theorem 2 hold, i.e., Θ is a compact set of Θ_s and the observations $\{Q_t\}_{t=1}^n$ are generated from a stationary and ergodic AcAF model with true parameter $\boldsymbol{\theta}_0$ where $\boldsymbol{\theta}_0$ is in the interior of Θ. We use $Y_{1,n,k}$, $Y_{2,n,k}$ and $Q_{n,k}$ to denote the kth order statistics of $\{Y_{1t}\}_{t=1}^n$, $\{Y_{2t}\}_{t=1}^n$ and $\{Q_t\}_{t=1}^n$, respectively. 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 $\gamma_1, \delta_1, \beta_1$ in Θ by $C_b < 1$ and use C to denote a generic positive constant.

We first prove the identifiability of the AcAF model in Lemma 4.

Lemma 4 (Identifiability)**.** *If $Q_t(\boldsymbol{\theta}) = Q_t(\boldsymbol{\theta}_0)$ a.s. for all t, then $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. Here a.s. is for the infinite product space generated by $\{\cdots, Y_{1,-1}, Y_{2,-1}, Y_{1,0}, Y_{2,0}, Y_{1,1}, Y_{2,1}, Y_{1,2}, Y_{2,2}, \cdots\}$, where $Y_{i,t}$'s are i.i.d. unit Fréchet random variables.*

Proof of Lemma 4. We denote $\sigma_t = \sigma_t(\boldsymbol{\theta})$, $\alpha_{1t} = \alpha_{1t}(\boldsymbol{\theta})$, $\alpha_{2t} = \alpha_{2t}(\boldsymbol{\theta})$, $\sigma_t^0 = \sigma_t(\boldsymbol{\theta}_0)$, $\alpha_{1t}^0 = \alpha_{1t}(\boldsymbol{\theta}_0)$, $\alpha_{2t}^0 = \alpha_{2t}(\boldsymbol{\theta}_0)$.

Suppose there exist $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_0$ such that $Q_t(\boldsymbol{\theta}) = Q_t(\boldsymbol{\theta}_0)$ a.s., then

$$\mu + \sigma_t \max(Y_{1t}^{1/\alpha_{1t}}, Y_{2t}^{1/\alpha_{2t}}) = \mu_0 + \sigma_t^0 \max(Y_{1t}^{1/\alpha_{1t}^0}, Y_{2t}^{1/\alpha_{2t}^0}), \quad a.s.$$

Since $Y_{1,n,1} \searrow 0$ and $Y_{2,n,1} \searrow 0$ a.s., by the boundness of $(\sigma_t, \alpha_{1t}, \alpha_{2t})^T$ and $(\sigma_t^0, \alpha_{1t}^0, \alpha_{2t}^0)^T$, we have $\mu = \mu_0$. By rearrangement, we obtain

$$\frac{\sigma_t}{\sigma_t^0} = \frac{\max(Y_{1t}^{1/\alpha_{1t}^0}, Y_{2t}^{1/\alpha_{2t}^0})}{\max(Y_{1t}^{1/\alpha_{1t}}, Y_{2t}^{1/\alpha_{2t}})}.$$

Denote $\mathcal{F}_t = \sigma(\cdots, Y_{1,t-1}, Y_{2,t-1}, Y_{1t}, Y_{2t})$, we know that $Y_{1t}, Y_{2t} \perp \mathcal{F}_{t-1}$ and $\alpha_{1t}, \alpha_{2t}, \sigma_t$, $\alpha_{1t}^0, \alpha_{2t}^0, \sigma_t^0 \in \mathcal{F}_{t-1}$. Since $\sigma_t/\sigma_t^0 \in \mathcal{F}_{t-1}$, it is easy to verify that σ_t/σ_t^0 equals to a constant.

In general, we assume $\sigma_t/\sigma_t^0 = 1$, then we have $\sigma_t = \sigma_t^0$ a.s., i.e. $\max(Y_{1t}^{1/\alpha_{1t}^0}, Y_{2t}^{1/\alpha_{2t}^0}) = \max(Y_{1t}^{1/\alpha_{1t}}, Y_{2t}^{1/\alpha_{2t}})$ for all t.

Since $Y_{1t}, Y_{2t} \perp \mathcal{F}_{t-1}$, $Y_{1t} \perp Y_{2t}$ and Y_{1t}, Y_{2t} are two continuous random variables, the above equation holds if and only if $\sigma_t(\boldsymbol{\theta}) = \sigma_t(\boldsymbol{\theta}_0)$, $\alpha_{1t}(\boldsymbol{\theta}) = \alpha_{1t}(\boldsymbol{\theta}_0)$, $\alpha_{2t}(\boldsymbol{\theta}) = \alpha_{2t}(\boldsymbol{\theta}_0)$ a.s. From the autoregressive equation of $\log \alpha_{1t}$, if $\alpha_{1t}(\boldsymbol{\theta}) = \alpha_{1t}(\boldsymbol{\theta}_0)$ a.s., we have

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

By the same argument as above, since $\alpha_{1,t-1} \in \mathcal{F}_{t-2}$ and $Q_{t-1} \notin \mathcal{F}_{t-2}$, we 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 $\delta_0 = \delta_0^0$, $\delta_1 = \delta_1^0$, $\delta_2 = \delta_2^0$ and $\delta_3 = \delta_3^0$; $\beta_0 = \beta_0^0$, $\beta_1 = \beta_1^0$, $\beta_2 = \beta_2^0$ and $\beta_3 = \beta_3^0$. Hence $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ for all t. □

Given the parameter $\boldsymbol{\theta}$ and an initial value $(\sigma_1, \alpha_{11}, \alpha_{21})^T$, $\{\sigma_t, \alpha_{1t}, \alpha_{2t}\}_{t=1}^n$ can be recovered recursively by their autoregressive equations. In the following, we use $\sigma_t(\boldsymbol{\theta}), \alpha_{1t}(\boldsymbol{\theta}), \alpha_{2t}(\boldsymbol{\theta})$ ($\sigma_t, \alpha_{1t}, \alpha_{2t}$ for simplicity) to denote the scale parameter series and the tail indices series based on $\boldsymbol{\theta}$ and true initial value $(\sigma_1^0, \alpha_{11}^0, \alpha_{21}^0)^T$, and use $\tilde{\sigma}_t(\boldsymbol{\theta}), \tilde{\alpha}_{1t}(\boldsymbol{\theta}), \tilde{\alpha}_{2t}(\boldsymbol{\theta})$ (or $\tilde{\sigma}_t, \tilde{\alpha}_{1t}, \tilde{\alpha}_{2t}$ for simplicity) to denote the ones based on $\boldsymbol{\theta}$ and an arbitrary initial value $(\tilde{\sigma}_1, \tilde{\alpha}_{11}, \tilde{\alpha}_{21})^T$. We denote the unobserved true hidden process by $\sigma_t(\boldsymbol{\theta}_0)$, $\alpha_{1t}(\boldsymbol{\theta}_0)$, $\alpha_{2t}(\boldsymbol{\theta}_0)$ (or $\sigma_t^0, \alpha_{1t}^0, \alpha_{2t}^0$ for simplicity).

By the compactness of Θ and the boundness of $\gamma_2 \exp(-\gamma_3 Q_{t-1})$, $\delta_2 \exp(-\delta_3 Q_{t-1})$ and $-\beta_2 \exp(-\beta_3 Q_{t-1})$, there exist uniform lower bound and upper bound of $\{\sigma_t, \alpha_{1t}, \alpha_{2t}\}$ and $\{\tilde{\sigma}_t, \tilde{\alpha}_{1t}, \tilde{\alpha}_{2t}\}$ for all $\boldsymbol{\theta} \in \Theta$. We denote the lower bound by $(\sigma_L, \alpha_{1L}, \alpha_{2L})^T$ and the upper bound by $(\sigma_U, \alpha_{1U}, \alpha_{2U})^T$. The uniform boundedness plays a key role in the following proof.

Given $(\sigma_t, \alpha_{1t}, \alpha_{2t})^T$, the conditional log-likelihood function $l_t(\boldsymbol{\theta})$ of Q_t is,

$$\begin{aligned} l_t(\boldsymbol{\theta}) = &\log \left[\alpha_{1t}\sigma_t^{\alpha_{1t}}(Q_t - \mu)^{-\alpha_{1t}-1} + \alpha_{2t}\sigma_t^{\alpha_{2t}}(Q_t - \mu)^{-\alpha_{2t}-1}\right] \\ &- \sigma_t^{\alpha_{1t}}(Q_t - \mu)^{-\alpha_{1t}} - \sigma_t^{\alpha_{2t}}(Q_t - \mu)^{-\alpha_{2t}}. \end{aligned}$$

By conditional independence, the log-likelihood function

$$L_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{t=1}^{n} l_t(\boldsymbol{\theta}) = \frac{1}{n}\sum_{t=1}^{n}\Bigg[\log\left[\alpha_{1t}\sigma_t^{\alpha_{1t}}(Q_t-\mu)^{-\alpha_{1t}-1} + \alpha_{2t}\sigma_t^{\alpha_{2t}}(Q_t-\mu)^{-\alpha_{2t}-1}\right] - \sigma_t^{\alpha_{1t}}(Q_t-\mu)^{-\alpha_{1t}} - \sigma_t^{\alpha_{2t}}(Q_t-\mu)^{-\alpha_{2t}}\Bigg].$$

We use $\tilde{l}_t(\boldsymbol{\theta})$ and $\tilde{L}_n(\boldsymbol{\theta})$ to denote the corresponding log-likelihood functions when $(\tilde{\sigma}_t, \tilde{\alpha}_{1t}, \tilde{\alpha}_{2t})^T$ is used.

Lemma 5 gives the result about the behavior of score function and Fisher information matrix at the true parameter $\boldsymbol{\theta}_0$ given true initial value $(\sigma_1^0, \alpha_{11}^0, \alpha_{21}^0)^T$.

Lemma 5. *Under the conditions in Theorem 2,* $\mathrm{E}_{\boldsymbol{\theta}_0}\left[\frac{\partial}{\partial\boldsymbol{\theta}} l_t(\boldsymbol{\theta}_0)\right] = \mathbf{0}$. *For* $\mathbf{M_0}$*, the Fisher information matrix at* $\boldsymbol{\theta}_0$*, we have* $\mathbf{M_0} = \mathrm{Var}_{\boldsymbol{\theta}_0}\left[\frac{\partial}{\partial\boldsymbol{\theta}} l_t(\boldsymbol{\theta}_0)\right] = -\mathrm{E}_{\boldsymbol{\theta}_0}[\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T} l_t(\boldsymbol{\theta}_0)]$*, and* $\mathbf{M_0}$ *is also well defined and positive definite.*

Proof of Lemma 5. For the first part: $\mathrm{E}_{\boldsymbol{\theta}_0}[\frac{\partial}{\partial\boldsymbol{\theta}} l_t(\boldsymbol{\theta}_0)] = \mathbf{0}$. We define $f_t(q_t, \boldsymbol{\theta}_0) = f_t(q_t, \boldsymbol{\theta}_0 | \sigma_t, \alpha_{1t}, \alpha_{2t})$ as the conditional p.d.f. of Q_t given $(\sigma_t, \alpha_{1t}, \alpha_{2t})^T$. After interchanging the integration operator with the differential operator, we obtain

$$\begin{aligned}\mathrm{E}_{\boldsymbol{\theta}_0}\left[\frac{\partial \log f_t(q_t,\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\right] &= \int \frac{\partial \log f_t(q_t,\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} f_t(q_t,\boldsymbol{\theta}_0)dq_t \\ &= \int \frac{1}{f_t(q_t,\boldsymbol{\theta}_0)}\frac{\partial f_t(q_t,\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} f_t(q_t,\boldsymbol{\theta}_0)dq_t = \int \frac{\partial f_t(q_t,\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} dq_t.\end{aligned}$$

Note that $(\sigma_t^0, \alpha_{1t}^0, \alpha_{2t}^0)^T$ is bounded between $[\sigma_L, \sigma_U] \times [\alpha_{1L}, \alpha_{1U}] \times [\alpha_{2L}, \alpha_{2U}]$, then it is easy to find a function g such that $|\frac{\partial f_t(q_t,\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}| \le g(q_t)$ and $\int g(q_t)dq_t < \infty$ for all $\boldsymbol{\theta} \in (\boldsymbol{\theta}_0 - \boldsymbol{\epsilon}, \boldsymbol{\theta}_0 + \boldsymbol{\epsilon})$ and some $\boldsymbol{\epsilon} > \mathbf{0}$. Then we get $\int \frac{\partial f_t(q_t,\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} dq_t = \frac{\partial}{\partial\boldsymbol{\theta}}\int f_t(q_t,\boldsymbol{\theta}_0)dq_t = \mathbf{0}$ by dominated convergence theorem, which gives that $\mathrm{E}_{\boldsymbol{\theta}_0}[\frac{\partial}{\partial\boldsymbol{\theta}} l_t(\boldsymbol{\theta}_0)] = \mathbf{0}$.

For $\mathbf{M_0} = \mathrm{Var}_{\boldsymbol{\theta}_0}(\frac{\partial}{\partial\boldsymbol{\theta}} l_t(\boldsymbol{\theta}_0))$, we have

$$\mathrm{E}_{\boldsymbol{\theta}_0}\left[\frac{\partial^2 \log f_t(q_t,\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right] = \int \frac{\partial}{\partial\boldsymbol{\theta}}\left(\frac{1}{f_t(q_t,\boldsymbol{\theta}_0)}\frac{\partial f_t(q_t,\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}^T}\right) f_t(q_t,\boldsymbol{\theta}_0)dq_t$$
$$= \int\left(-\frac{\frac{\partial f_t(q_t,\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}}{f_t^2(q_t,\boldsymbol{\theta}_0)}\frac{\partial f_t(q_t,\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}^T} + \frac{1}{f_t(q_t,\boldsymbol{\theta}_0)}\frac{\partial^2 f_t(q_t,\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right) f_t(q_t,\boldsymbol{\theta}_0)dq_t$$
$$= -\int \frac{1}{f_t^2(q_t,\boldsymbol{\theta}_0)}\frac{\partial f_t(q_t,\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\frac{\partial f_t(q_t,\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}^T} f_t(q_t,\boldsymbol{\theta}_0)dq_t + \int\frac{\partial^2 f_t(q_t,\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}dq_t$$
$$= -\int\left[\frac{1}{f_t(q_t,\boldsymbol{\theta}_0)}\frac{\partial f_t(q_t,\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\right]\left[\frac{1}{f_t(q_t,\boldsymbol{\theta}_0)}\frac{\partial f_t(q_t,\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}^T}\right] f_t(q_t,\boldsymbol{\theta}_0)dq_t + \int\frac{\partial^2 f_t(q_t,\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}dq_t,$$

in which we have

$$\int\frac{\partial^2 f_t(q_t,\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}dq_t = \frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\int f_t(q_t,\boldsymbol{\theta}_0)dq_t = 0.$$

Then we obtain $\mathbf{M_0} = \mathrm{Var}_{\boldsymbol{\theta}_0}(\frac{\partial}{\partial\boldsymbol{\theta}} l_t(\boldsymbol{\theta}_0)) = -\mathrm{E}_{\boldsymbol{\theta}_0}[\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T} l_t(\boldsymbol{\theta}_0)]$. When t goes to infinity, the sequence $\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T} l_t(\boldsymbol{\theta}_0)$ is strictly stationary, and their expectations are the same. Hence $\mathbf{M}_0$ is independent with t and is well defined, i.e., $\mathbf{M}_0 < \infty$. Moreover, since $Y_{1t}, Y_{2t} \perp \mathcal{F}_{t-1}$, we can observe that there does not exist a $\mathbf{c} \in \mathbb{R}^{13}$ such that $\mathbf{c}^T \frac{\partial}{\partial\boldsymbol{\theta}} l_t(\boldsymbol{\theta}_0) = \mathbf{0}$ a.s. Thus $\mathbf{M}_0$ is positive definite. □

In Lemma 6, we will show that the expectations of the items in $\frac{\partial}{\partial\boldsymbol{\theta}} L_n(\boldsymbol{\theta}_0)$ and $\frac{\partial^2}{\partial\theta_i\partial\theta_j} L_n(\boldsymbol{\theta}_0)$ exist, which serve as building blocks for the proof of latter lemmas.

Lemma 6. *Under the conditions in Theorem 2, we have*

(a) for any $\alpha > 0$, $\frac{1}{n}\sum_{t=1}^n (Q_t - \mu_0)^{-\alpha} \to_p E_{\boldsymbol{\theta}_0}(Q_1 - \mu_0)^{-\alpha} < \infty$,

(b) for any positive integer k, $\frac{1}{n}\sum_{t=1}^n \left[\log(Q_t - \mu_0)\right]^k \to_p E_{\boldsymbol{\theta}_0}\left[\log(Q_1 - \mu_0)\right]^k < \infty$.

Proof of Lemma 6. For (a), by the boundness of the scale parameter $\{\sigma_t^0\}$ and tail indices $\{\alpha_{1t}^0\}$ and $\{\alpha_{2t}^0\}$, we have

$$Q_t - \mu_0 > \sigma_L \min\big\{\max(Y_{1t}^{1/\alpha_{1L}}, Y_{2t}^{1/\alpha_{2L}}), \max(Y_{1t}^{1/\alpha_{1L}}, Y_{2t}^{1/\alpha_{2U}}),$$
$$\max(Y_{1t}^{1/\alpha_{1U}}, Y_{2t}^{1/\alpha_{2L}}), \max(Y_{1t}^{1/\alpha_{1U}}, Y_{2t}^{1/\alpha_{2U}})\big\}.$$

Since Y_{1t} and Y_{2t} are two i.i.d. unit Fréchet random variables, it is easy to show that $E_{\boldsymbol{\theta}_0}(Q_t - \mu_0)^{-\alpha} < \infty$ for any $\alpha > 0$. The result of (a) follows the ergodicity of our model and pointwise ergodicity theorem in Birkhoff (1931).

For (b), we have, for any positive integer k,

$$
\begin{aligned}
|\log(Q_t - \mu_0)|^k &= |\log \sigma_t + \log \max(Y_{1t}^{1/\alpha_{1t}}, Y_{2t}^{1/\alpha_{2t}})|^k \\
&= \begin{cases} |\log \sigma_t + (\log Y_{1t})/\alpha_{1t}|^k & \text{if } Y_{1t}^{1/\alpha_{1t}} \geq Y_{2t}^{1/\alpha_{2t}} \\ |\log \sigma_t + (\log Y_{2t})/\alpha_{1t}|^k & \text{if } Y_{1t}^{1/\alpha_{1t}} < Y_{2t}^{1/\alpha_{2t}} \end{cases} \\
&\leq 2^{k-1}\left[C + \max\left((|\log Y_{1t}|^k)/\alpha_{1L}^k, (|\log Y_{2t}|^k)/\alpha_{2L}^k\right)\right] \\
&\leq 2^{k-1}\left[C + \max\left(1/\alpha_{1L}^k, 1/\alpha_{2L}^k\right)\max\left(|\log Y_{1t}|^k, |\log Y_{2t}|^k\right)\right],
\end{aligned}
$$

where $C = |\log \sigma_t|^k$ and the first equation follows the fact that $|x+y|^k \leq \frac{1}{2}(|x|^k + |y|^k)$. It is known that both $\log(Y_{1t})$ and $\log(Y_{2t})$ follow Gumbel distribution thus $E_{\boldsymbol{\theta}_0}(|\log Y_{it}|^k) < \infty$, $i = 1, 2$, then we obtain $E_{\boldsymbol{\theta}_0}[\max\left(|\log Y_{1t}|^k, |\log Y_{2t}|^k\right)] < \infty$. The result of (b) follows from the ergodicity of the AcAF model and pointwise ergodicity theorem in Birkhoff (1931). □

The main technical difficulty is that the support of Q_t depends on the unknown location parameter μ_0. Lemma 7 to Lemma 15 aim to solve this difficulty by establishing uniform convergence between $\frac{1}{n}\sum_{t=1}^n f(Q_t - \mu_n)$ and $\frac{1}{n}\sum_{t=1}^n f(Q_t - \mu_0)$ for μ_n within a neighborhood of μ_0, where $f(\cdot)$ denotes the generic function that appears in the first and second order derivatives of $\tilde{L}_n(\boldsymbol{\theta}_0)$. The main result is stated in Lemma 15.

Recall that $Q_{n,1} = \min_{1 \leq t \leq n} Q_t$. Lemma 7 provides an asymptotic bound on the distance between $Q_{n,1}$ and μ_0, indicating that $Q_{n,1}$ converges to μ_0 at a rate which is slower than polynomial.

Lemma 7. *Under the conditions in Theorem 2,*

$$
Q_{n,1} - \mu_0 \geq A_n = O_p\left((\log n)^{-1/\max(\alpha_{1L}, \alpha_{2L})}\right).
$$

Proof of Lemma 7. When $Y_{1t}, Y_{2t} < 1$, we have

$$Q_t - \mu_0 = \sigma_t \max(Y_{1t}^{1/\alpha_{1t}}, Y_{2t}^{1/\alpha_{2t}}) \geq \sigma_L \max(Y_{1t}^{1/\alpha_{1L}}, Y_{2t}^{1/\alpha_{2L}}).$$

Since $Y_{1,n,1}, Y_{2,n,1} < 1$ as $n \to \infty$, we can obtain, as $n \to \infty$,

$$Q_{n,1} - \mu_0 \geq \sigma_L \max(Y_{1,n,1}^{1/\alpha_{1L}}, Y_{2,n,1}^{1/\alpha_{2L}}) \geq \sigma_L \max(Y_{1,n,1}, Y_{2,n,1})^{1/\max(\alpha_{1L},\alpha_{2L})}, \ \ a.s.$$

Since $(\log n) \max(Y_{1,n,1}, Y_{2,n,1}) \to_p 1$, we have

$$\sigma_L \max(Y_{1,n,1}^{1/\alpha_{1L}}, Y_{2,n,1}^{1/\alpha_{2L}}) = O_p\left((\log n)^{-1/\max(\alpha_{1L},\alpha_{2L})}\right).$$

Then $Q_{n,1} - \mu_0 \geq O_p\left((\log n)^{-1/\max(\alpha_{1L},\alpha_{2L})}\right)$. □

In Lemma 8, we state the foundation for the uniform convergence results of the first and second derivatives of $L_n(\boldsymbol{\theta})$ given in Lemmas 12 and 15.

Lemma 8. *Denote (a) $S_n^\alpha(\mu) = n^{-1}\sum_{k=1}^n (Q_{n,k}-\mu)^{-\alpha}$, $\alpha > 0$ or (b) $S_n^\alpha(\mu) = n^{-1}\sum_{k=1}^n \log(Q_{n,k}-\mu)$ or (c) $S_n^\alpha(\mu) = n^{-1}\sum_{k=1}^n (Q_{n,k}-\mu)^{-\alpha}[\log(Q_{n,k}-\mu)]^m$ for $m = 1, 2, 3$. Under the conditions in Theorem 2, given positive sequence τ_n such that $\tau_n \sim n^{-r}$, $r > 0$, the following result holds uniformly over $|\mu_n - \mu_0| < \tau_n$,*

$$|S_n^\alpha(\mu_n) - S_n^\alpha(\mu_0)| \leq O_p(\tau_n).$$

Proof of Lemma 8. By Lemma 7, it follows that $Q_{n,1} - \mu_0 \geq O_p\left((\log n)^{-1/\max(\alpha_{1L},\alpha_{2L})}\right)$. Since τ_n's convergence rate to 0 is faster than $O_p\left((\log n)^{-1/\max(\alpha_{1L},\alpha_{2L})}\right)$, we obtain

$$\begin{aligned} Q_{n,1} - \mu_n = Q_{n,1} - \mu_0 - (\mu_n - \mu_0) &\geq O_p((\log n)^{\frac{1}{\max(\alpha_{1L},\alpha_{2L})}}) - (\mu_n - \mu_0) \\ &\geq O_p\left((\log n)^{-1/\max(\alpha_{1L},\alpha_{2L})}\right) - \tau_n > 0. \end{aligned}$$

Therefore $(Q_t - \mu_n)^{-\alpha}$ 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 obtain

$$\begin{aligned}
|S_n^\alpha(\mu_n) - S_n^\alpha(\mu_0)| &\le \frac{1}{n}\sum_{k=1}^n |(Q_{n,k}-\mu_n)^{-\alpha} - (Q_{n,k}-\mu_0)^{-\alpha}| \\
&\le \frac{1}{n}\sum_{k=1}^n \frac{\alpha|\mu_n-\mu_0|}{\min\{Q_{n,k}-\mu_n, Q_{n,k}-\mu_0\}^{\alpha+1}} \\
&\le \frac{\tau_n}{n}\sum_{k=1}^n \frac{\alpha}{(Q_{n,k}-\mu_n)^{\alpha+1}} \\
&= \frac{\tau_n}{n}\sum_{k=1}^n \frac{\alpha}{(Q_{n,k}-\mu_0+\mu_0-\mu_n)^{\alpha+1}} \\
&\le \frac{\tau_n}{n}\sum_{k=1}^n \frac{\alpha}{(Q_{n,k}-\mu_0-\tau_n)^{\alpha+1}},
\end{aligned}$$

where the second inequality follows the fact that $f(x) = x^{-\alpha}$ is a local Lipschitz function, i.e., $|f(x)-f(y)| \le |\max\{f'(x), f'(y)\}||x-y|$, so $|(Q_{n,k}-\mu_n)^{-\alpha} - (Q_{n,k}-\mu_0)^{-\alpha}| \le \max\{\alpha(Q_{n,k}-\mu_n)^{-(\alpha+1)}, \alpha(Q_{n,k}-\mu_0)^{-(\alpha+1)}\}|\mu_n-\mu_0|$.

Since $Q_{n,1}-\mu_0 \ge O_p((\log n)^{-\frac{1}{\max(\alpha_{1L},\alpha_{2L})}})$, for any fixed $0<\rho<1$, we have $P(\rho(Q_{n,1}-\mu_0) > \tau_n) \to 1$, so $P(\rho(Q_{n,k}-\mu_0) > \tau_n, \text{for all } 1 \le k \le n) \to 1$.

With probability going to 1, we have

$$\begin{aligned}
\frac{\tau_n}{n}\sum_{k=1}^n \frac{\alpha}{(Q_{n,k}-\mu_0-\tau_n)^{\alpha+1}} &\le \frac{\tau_n}{n}\sum_{k=1}^n \frac{\alpha}{[(Q_{n,k}-\mu_0)(1-\rho)]^{\alpha+1}} \\
&= \tau_n\left[\frac{1}{n}\sum_{k=1}^n \frac{1}{(Q_{n,k}-\mu_0)^{\alpha+1}}\right]\frac{\alpha}{(1-\rho)^{\alpha+1}}. \\
&= O_p(\tau_n),
\end{aligned}$$

which follows from $\frac{1}{n}\sum_{t=1}^n (Q_t-\mu_0)^{-\alpha} < \infty$ in Lemma 6(a), then $\frac{1}{n}\sum_{k=1}^n \frac{1}{(Q_{n,k}-\mu_0)^{\alpha+1}} < \infty$. Hence we obtain $|S_n^\alpha(\mu_n) - S_n^\alpha(\mu_0)| \le O_p(\tau_n)$. For $\mu_n < \mu_0$, the proof is similar.

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

$$\begin{aligned}
|S_n^\alpha(\mu_n) - S_n^\alpha(\mu_0)| \leq& \frac{1}{n}\sum_{k=1}^n |\log(Q_{n,k}-\mu_n) - \log(Q_{n,k}-\mu_0)| \\
=& \frac{1}{n}\sum_{k=1}^n \log\left(1 + \frac{\mu_n-\mu_0}{Q_{n,k}-\mu_n}\right) \\
\leq& \frac{\tau_n}{n}\sum_{k=1}^n \frac{1}{Q_{n,k}-\mu_n} \\
=& O_p(\tau_n),
\end{aligned}$$

where the last inequality follows from the fact that $\log(1+x) < x$ when $x > 0$ and the last equality follows the results 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.

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

$$\begin{aligned}
|S_n^\alpha(\mu_n) - S_n^\alpha(\mu_0)| \leq& \frac{1}{n}\sum_{k=1}^n (Q_{n,k}-\mu_n)^{-\alpha}|\log(Q_{n,k}-\mu_n) - \log(Q_{n,k}-\mu_0)| \\
&+ \frac{1}{n}\sum_{k=1}^n |(Q_{n,k}-\mu_n)^{-\alpha} - (Q_{n,k}-\mu_0)^{-\alpha}||\log(Q_{n,k}-\mu_0)|.
\end{aligned}$$

For the first term in the sum,

$$\begin{aligned}
&\frac{1}{n}\sum_{k=1}^n (Q_{n,k}-\mu_n)^{-\alpha}|\log(Q_{n,k}-\mu_n) - \log(Q_{n,k}-\mu_0)| \\
=& \frac{1}{n}\sum_{k=1}^n (Q_{n,k}-\mu_n)^{-\alpha}\log\left(1 + \frac{\mu_n-\mu_0}{Q_{n,k}-\mu_n}\right) \\
\leq& \frac{\tau_n}{n}\sum_{k=1}^n (Q_{n,k}-\mu_n)^{-(\alpha+1)} \\
=& O_p(\tau_n).
\end{aligned}$$

For the second term in the sum,

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

where the last inequality follows from the Cauchy-Schwarz inequality. The last equality follows from the Lemma 6 and the results for $S_n^{\alpha}(\mu)=n^{-1}\sum_{k=1}^{n}(Q_{n,k}-\mu)^{-\alpha}$.

When $m=2$, we have

$$
\begin{aligned}
|S_n^{\alpha}(\mu_n)-S_n^{\alpha}(\mu_0)|\leq&\frac{1}{n}\sum_{k=1}^{n}(Q_{n,k}-\mu_n)^{-\alpha}\log(Q_{n,k}-\mu_n)|\log(Q_{n,k}-\mu_n)-\log(Q_{n,k}-\mu_0)| \\
+&\frac{1}{n}\sum_{k=1}^{n}|(Q_{n,k}-\mu_n)^{-\alpha}\log(Q_{n,k}-\mu_n)-(Q_{n,k}-\mu_0)^{-\alpha}\log(Q_{n,k}-\mu_0)|\log(Q_{n,k}-\mu_0).
\end{aligned}
$$

For the first term in the sum,

$$
\begin{aligned}
&\frac{1}{n}\sum_{k=1}^{n}(Q_{n,k}-\mu_n)^{-\alpha}\log(Q_{n,k}-\mu_n)|\log(Q_{n,k}-\mu_n)-\log(Q_{n,k}-\mu_0)| \\
=&\frac{1}{n}\sum_{k=1}^{n}(Q_{n,k}-\mu_n)^{-\alpha}\log(Q_{n,k}-\mu_n)\log\left(1+\frac{\mu_n-\mu_0}{Q_{n,k}-\mu_n}\right) \\
\leq&\frac{\tau_n}{n}\sum_{k=1}^{n}(Q_{n,k}-\mu_n)^{-(\alpha+1)}|\log(Q_{n,k}-\mu_n)| \\
=&O_p(\tau_n).
\end{aligned}
$$

For the second term in the sum,

$$
\begin{aligned}
&\frac{1}{n}\sum_{k=1}^{n}|(Q_{n,k}-\mu_n)^{-\alpha}\log(Q_{n,k}-\mu_n)-(Q_{n,k}-\mu_0)^{-\alpha}\log(Q_{n,k}-\mu_0)|\cdot\log(Q_{n,k}-\mu_0)\\
\leq&\frac{\tau_n}{n}\sum_{k=1}^{n}\left[\frac{1+(\alpha)|\log(Q_{n,k}-\mu_0)|}{(Q_{n,k}-\mu_n)^{\alpha+1}}\right]|\log(Q_{n,k}-\mu_0)|\\
=&\frac{\tau_n}{n}\sum_{k=1}^{n}(Q_{n,k}-\mu_n)^{-\alpha-1}\left[|\log(Q_{n,k}-\mu_0)|+\alpha|\log(Q_{n,k}-\mu_0)|^2\right]\\
\leq&\tau_n\left(\frac{1}{n}\sum_{k=1}^{n}\frac{1}{(Q_{n,k}-\mu_n)^{2\alpha+2}}\right)^{1/2}\left(\frac{1}{n}\sum_{k=1}^{n}\left[|\log(Q_{n,k}-\mu_0)|+\alpha|\log(Q_{n,k}-\mu_0)|^2\right]^2\right)^{1/2}\\
=&O_p(\tau_n),
\end{aligned}
$$

When $m=3$, we have

$$
\begin{aligned}
|S_n^\alpha(\mu_n)-S_n^\alpha(\mu_0)|\leq&\frac{1}{n}\sum_{k=1}^{n}(Q_{n,k}-\mu_n)^{-\alpha}\log^2(Q_{n,k}-\mu_n)|\log(Q_{n,k}-\mu_n)-\log(Q_{n,k}-\mu_0)|\\
+&\frac{1}{n}\sum_{k=1}^{n}|(Q_{n,k}-\mu_n)^{-\alpha}\log^2(Q_{n,k}-\mu_n)-(Q_{n,k}-\mu_0)^{-\alpha}\log^2(Q_{n,k}-\mu_0)|\log(Q_{n,k}-\mu_0).
\end{aligned}
$$

According to some conclusions that we got in the process of proving the case of $m=1$ and $m=2$, it is easy to verify that $|S_n^\alpha(\mu_n)-S_n^\alpha(\mu_0)|=O_p(\tau_n)$.

For $\mu_n<\mu_0$, the proof is similar. Then we complete the proof of (c) in Lemma 8. □

Lemmas 9 and 10 state that the supremum of the difference between the first n values of σ_t and σ_t^0 (so as α_{1t} and α_{1t}^0; α_{2t} and α_{2t}^0), which is impacted by the parameter difference $|\boldsymbol{\theta}-\boldsymbol{\theta}_0|$, converges at the rate of τ_n uniformly over t. This convergence rate also holds for their partial derivatives.

Lemma 9. *Denote $\boldsymbol{\Gamma}=(\gamma_0,\gamma_1,\gamma_2,\gamma_3)^T$ and $\boldsymbol{\Gamma_0}=(\gamma_0^0,\gamma_1^0,\gamma_2^0,\gamma_3^0)^T$, if $\|\boldsymbol{\Gamma}-\boldsymbol{\Gamma_0}\|<\tau_n$ and $\tau_n\searrow 0$, under the conditions in Theorem 2, we have*

$$(a)\ \sup_{1\leq t\leq n}\left|\alpha_{1t}-\alpha_{1t}^0\right|=O(\tau_n),$$

$$(b)\ \sup_{1\leq t\leq n}\left|\frac{\partial\alpha_{1t}}{\partial\Gamma_i}-\frac{\partial\alpha_{1t}^0}{\partial\Gamma_i}\right|=O(\tau_n),\ \textit{for } i=1,2,3,4,$$

$$(c)\ \sup_{1\leq t\leq n}\left|\frac{\partial^2\alpha_{1t}}{\partial\Gamma_i\partial\Gamma_j}-\frac{\partial^2\alpha_{1t}^0}{\partial\Gamma_i\partial\Gamma_j}\right|=O(\tau_n),\ \textit{for all } i,j=1,2,3,4,$$

uniformly over $\|\mathbf{\Gamma} - \mathbf{\Gamma_0}\| < \tau_n$.

Proof of Lemma 9. Here we illustrate our proof of (a), the proofs of (b) and (c) are similar. The domain of α_{1t} is bounded so the function $\exp(\cdot)$ defined on a compact set is Lipschitz continuous. Then it is equivalent to prove that $\sup_{1\le t\le n} |\log \alpha_{1t} - \log \alpha_{1t}^0| = O(\tau_n)$.

By repeatedly applying the autoregressive equation, we can obtain

$$\log \alpha_{1t} = \gamma_0 \sum_{k=1}^{t-1} \gamma_1^{k-1} + \gamma_2 \sum_{k=1}^{t-1} \gamma_1^{k-1} \exp(-\gamma_3 Q_{t-k}) + \gamma_1^{t-1} \log \alpha_{11}^0.$$

We have

$$\begin{aligned} |\log \alpha_{1t} - \log \alpha_{1t}^0| \le & \left| \gamma_0 \sum_{k=1}^{t-1} \gamma_1^{k-1} - \gamma_0^0 \sum_{k=1}^{t-1} (\gamma_1^0)^{k-1} \right| + \left| \gamma_1^{t-1} \log \alpha_{11}^0 - (\gamma_1^0)^{t-1} \log \alpha_{11}^0 \right| \\ & + \left| \gamma_2 \sum_{k=1}^{t-1} \gamma_1^{k-1} \exp(-\gamma_3 Q_{t-k}) - \gamma_2^0 \sum_{k=1}^{t-1} (\gamma_1^0)^{k-1} \exp(-\gamma_3^0 Q_{t-k}) \right|. \end{aligned}$$

We know that $\sum_{k=1}^{t} \gamma_1^{k-1} < \frac{1}{1-\gamma_1} \le \frac{1}{1-C_b}$ and

$$\begin{aligned} \left| \sum_{k=1}^{t-1} \gamma_1^{k-1} - \sum_{k=1}^{t-1} (\gamma_1^0)^{k-1} \right| \le & \left| \frac{1}{1-\gamma_1} - \frac{1}{1-\gamma_1^0} \right| + \left| \frac{\gamma_1^{t-1}}{1-\gamma_1} - \frac{(\gamma_1^0)^{t-1}}{1-\gamma_1^0} \right| \\ = & \left| \frac{1}{1-\gamma_1} - \frac{1}{1-\gamma_1^0} \right| + \left| \frac{(1-\gamma_1^0)\gamma_1^{t-1} - (1-\gamma_1)(\gamma_1^0)^{t-1}}{(1-\gamma_1)(1-\gamma_1^0)} \right| \\ = & \left| \frac{1}{1-\gamma_1} - \frac{1}{1-\gamma_1^0} \right| + \left| \frac{(\gamma_1^{t-1} - (\gamma_1^0)^{t-1}) - \gamma_1^0 \gamma_1 (\gamma_1^{t-2} - (\gamma_1^0)^{t-2})}{(1-\gamma_1)(1-\gamma_1^0)} \right|. \end{aligned}$$

Since $(\gamma_1^{t-1} - (\gamma_1^0)^{t-1}) = (\gamma_1 - \gamma_1^0)(\gamma_1^{t-2} + \gamma_1^{t-3}\gamma_1^0 + \cdots + (\gamma_1^0)^{t-2}) \le (\gamma_1 - \gamma_1^0) C_b^{t-2}(t-1) = o(\tau_n)$, then we obtain

$$\left| \sum_{k=1}^{t-1} \gamma_1^{k-1} - \sum_{k=1}^{t-1} (\gamma_1^0)^{k-1} \right| \le \frac{\tau_n}{(1-C_b)^2} + o(\tau_n) = O(\tau_n).$$

It is easy to see that the first two terms of the sum are $O(\tau_n)$ for any $1 \le t \le n$. For the third term, we have

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

The first two terms of the sum are $O(\tau_n)$ for any $1 \leq t \leq n$ by the boundness of $\exp(-\gamma_3 Q_{t-k})$.

For the third term, we have,

$$\begin{aligned}&\gamma_2^0\sum_{k=1}^{t-1}(\gamma_1^0)^{k-1}|\exp(-\gamma_3 Q_{t-k})-\exp(-\gamma_3^0 Q_{t-k})|\\ =&\gamma_2^0\sum_{k=1}^{t-1}(\gamma_1^0)^{k-1}Q_{t-k}\exp(-\gamma_{3k}' Q_{t-k})|\gamma_3-\gamma_3^0|\\ =&O(\tau_n),\end{aligned}$$

where γ_{3k}' is a positive number between γ_3 and γ_3^0 depending on Q_{t-k}, and $\gamma_{3k}' \to \gamma_3^0$ uniformly over all $k > 1$. By the compactness of Θ, $0 < C \leq \gamma_{3k}'$ for all $k \geq 1$. The first equality follows from mean value theorem and the second equality follows from the uniform boundness of $Q_{t-k}\exp(-\gamma_{3k}' Q_{t-k})$. □

Lemma 10. *Denote* $\mathbf{\Phi} = (\delta_0, \delta_1, \delta_2, \delta_3)^T$ *and* $\mathbf{\Phi_0} = (\delta_0^0, \delta_1^0, \delta_2^0, \delta_3^0)^T$, *if* $\|\mathbf{\Phi} - \mathbf{\Phi_0}\| < \tau_n$ *and* $\tau_n \searrow 0$, *under the conditions in Theorem 2, we have*

$$(a)\ \sup_{1\leq t\leq n}\left|\alpha_{2t}-\alpha_{2t}^0\right| = O(\tau_n),$$

$$(b)\ \sup_{1\leq t\leq n}\left|\frac{\partial\alpha_{2t}}{\partial\Phi_i}-\frac{\partial\alpha_{2t}^0}{\partial\Phi_i}\right| = O(\tau_n),\ \textit{for all } i=1,2,3,4,$$

$$(c)\ \sup_{1\leq t\leq n}\left|\frac{\partial^2\alpha_{2t}}{\partial\Phi_i\partial\Phi_j}-\frac{\partial^2\alpha_{2t}^0}{\partial\Phi_i\partial\Phi_j}\right| = O(\tau_n),\ \textit{for all } i,j=1,2,3,4,$$

uniformly over $\|\mathbf{\Phi} - \mathbf{\Phi_0}\| < \tau_n$.

Denote $\mathbf{\Psi} = (\beta_0, \beta_1, \beta_2, \beta_3)^T$ *and* $\mathbf{\Psi_0} = (\beta_0^0, \beta_1^0, \beta_2^0, \beta_3^0)^T$, *if* $\|\mathbf{\Psi} - \mathbf{\Psi_0}\| < \tau_n$ *and* $\tau_n \searrow 0$, *under the conditions in Theorem 2, we have*

$$(d) \sup_{1\le t\le n} \left|\sigma_t - \sigma_t^0\right| = O(\tau_n),$$

$$(e) \sup_{1\le t\le n} \left| \frac{\partial \sigma_t}{\partial \Psi_i} - \frac{\partial \sigma_t^0}{\partial \Psi_i} \right| = O(\tau_n), \text{ for all } i = 1,2,3,4,$$

$$(f) \sup_{1\le t\le n} \left| \frac{\partial^2 \sigma_t}{\partial \Psi_i \partial \Psi_j} - \frac{\partial^2 \sigma_t^0}{\partial \Psi_i \partial \Psi_j} \right| = O(\tau_n), \text{ for all } i,j = 1,2,3,4,$$

uniformly over $\|\mathbf{\Psi} - \mathbf{\Psi_0}\| < \tau_n$.

Proof of Lemma 10. The proof is similar to that of Lemma 9. □

Lemma 11 is used to build blocks for the proof of Lemma 12.

Lemma 11. *Suppose* $\tau_n \sim n^{-r}$, $r > 0$ *and* $\sup_{1\le t\le n} |\alpha_{1t} - \alpha'_{1t}| = O(\tau_n)$ *where* $\{\alpha_{1t}\}$ *and* $\{\alpha'_{1t}\}$ *represent two different series of tail index. Under the conditions in Theorem 2, we have*

$$\frac{1}{n}\sum_{t=1}^{n} |(Q_t - \mu_n)^{-\alpha_{1t}} - (Q_t - \mu_n)^{-\alpha'_{1t}}| = O_p(\tau_n)$$

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

Similarly, suppose $\sup_{1\le t\le n} |\alpha_{2t} - \alpha'_{2t}| = O(\tau_n)$ *where* $\{\alpha_{2t}\}$ *and* $\{\alpha'_{2t}\}$ *represent two different series of tail index. Under the conditions in Theorem 2, we have*

$$\frac{1}{n}\sum_{t=1}^{n} |(Q_t - \mu_n)^{-\alpha_{2t}} - (Q_t - \mu_n)^{-\alpha'_{2t}}| = O_p(\tau_n)$$

uniformly over $|\mu_n - \mu_0| < \tau_n$. *The same results holds for* $\frac{1}{n}\sum_{t=1}^{k} |Q_t - \mu_n)^{-\alpha_{2t}} - (Q_t - \mu_n)^{-\alpha'_{2t}}|[\log(Q_t - \mu_n)]^k$, $k = 1, 2$.

Proof of Lemma 11. We just give the proof for the case of $\frac{1}{n}\sum_{t=1}^{n} |(Q_t - \mu_n)^{-\alpha_{1t}} - (Q_t - \mu_n)^{-\alpha'_{1t}}|$, the proofs of other cases are similar. Without loss of generality, we assume $\alpha'_{1t} > \alpha_{1t}$, the proof for the other direction is the same. By mean value theorem, we have

$$\begin{aligned}&\frac{1}{n}\sum_{t=1}^{n}|(Q_t-\mu_n)^{-\alpha_{1t}}-(Q_t-\mu_n)^{-\alpha'_{1t}}|\le\frac{C}{n}\sum_{t=1}^{n}(Q_t-\mu_n)^{-\alpha^*_{1t}}|\log(Q_t-\mu_n)|\tau_n\\ \le&\frac{\tau_n C}{n}\sum_{t=1}^{n}\left((Q_t-\mu_n)^{-\alpha_{1L}}+(Q_t-\mu_n)^{-\alpha_{1U}}\right)|\log(Q_t-\mu_n)|=O_p(\tau_n),\end{aligned}$$

where $\alpha^*_{1t}\in(\alpha_{1t},\alpha'_{1t})$. The last equality follows from Lemma 8. □

Lemma 12 gives the uniform convergence results of the second order derivatives of $L_n(\boldsymbol{\theta})$ over a neighborhood of $\boldsymbol{\theta}_0$, which is used in the proof of Lemma 15(a). We denote $m_{\theta_i\theta_j}(\theta_0)=-E_{\boldsymbol{\theta}_0}\left[\frac{\partial^2}{\partial\theta_i\partial\theta_j}l_1(\boldsymbol{\theta}_0)\right]$.

Lemma 12. *Under the conditions in Theorem 2, for all second order derivatives of $L_n(\boldsymbol{\theta}_n)$, we have $\frac{\partial^2}{\partial\theta_i\partial\theta_j}L_n(\boldsymbol{\theta}_n)\to_p -m_{\theta_i\theta_j}(\boldsymbol{\theta}_0)$, uniformly over $\|\boldsymbol{\theta}_n-\boldsymbol{\theta}_0\|<\tau_n$, where $\tau_n\sim n^{-r}$, $r>0$.*

Proof of Lemma 12. We just give the proof for the case of $\frac{\partial^2}{\partial\mu^2}L_n(\boldsymbol{\theta}_n)$, the proofs of other cases are similar. Note that the first and second order of the partial derivatives of $L_n(\cdot)$ are measurable functions of the stationary and ergodic series $\{Q_t\}$, so they are also ergodic and strictly stationary. By the pointwise ergodicity theorem [Birkhoff (1931)], we have $\frac{\partial^2}{\partial\mu^2}L_n(\boldsymbol{\theta}_0)\to_p -m_{\mu\mu}(\boldsymbol{\theta}_0)$, so we only need to prove that $\frac{\partial^2}{\partial\mu^2}L_n(\boldsymbol{\theta}_n)-\frac{\partial^2}{\partial\mu^2}L_n(\boldsymbol{\theta}_0)\to_p 0$ uniformly over the claimed region.

By definition, we obtain

$$
\begin{aligned}
&\frac{\partial^2}{\partial\mu^2}L_n(\boldsymbol{\theta}_n)-\frac{\partial^2}{\partial\mu^2}L_n(\boldsymbol{\theta}_0)\\
=&\frac{1}{n}\sum_{t=1}^{n}\Big[\alpha_{1t}^0(\alpha_{1t}^0+1)(\sigma_t^0)^{\alpha_{1t}^0}(Q_t-\mu_0)^{-(\alpha_{1t}^0+2)}-\alpha_{1t}(\alpha_{1t}+1)\sigma_t^{\alpha_{1t}}(Q_t-\mu_n)^{-(\alpha_{1t}+2)}\\
&+\alpha_{2t}^0(\alpha_{2t}^0+1)(\sigma_t^0)^{\alpha_{2t}^0}(Q_t-\mu_0)^{-(\alpha_{2t}^0+2)}-\alpha_{2t}(\alpha_{2t}+1)\sigma_t^{\alpha_{2t}}(Q_t-\mu_n)^{-(\alpha_{2t}+2)}\Big]\\
&+\frac{1}{n}\sum_{t=1}^{n}\Big[\frac{\alpha_{1t}(\alpha_{1t}+1)(\alpha_{1t}+2)\sigma_t^{\alpha_{1t}}(Q_t-\mu_n)^{-\alpha_{1t}-3}+\alpha_{2t}(\alpha_{2t}+1)(\alpha_{2t}+2)\sigma_t^{\alpha_{2t}}(Q_t-\mu_n)^{-\alpha_{2t}-3}}{\alpha_{1t}\sigma_t^{\alpha_{1t}}(Q_t-\mu_n)^{-\alpha_{1t}-1}+\alpha_{2t}\sigma_t^{\alpha_{2t}}(Q_t-\mu_n)^{-\alpha_{2t}-1}}\\
&-\frac{\alpha_{1t}^0(\alpha_{1t}^0+1)(\alpha_{1t}^0+2)(\sigma_t^0)^{\alpha_{1t}^0}(Q_t-\mu_0)^{-\alpha_{1t}^0-3}+\alpha_{2t}^0(\alpha_{2t}^0+1)(\alpha_{2t}^0+2)(\sigma_t^0)^{\alpha_{2t}^0}(Q_t-\mu_0)^{-\alpha_{2t}^0-3}}{\alpha_{1t}^0(\sigma_t^0)^{\alpha_{1t}^0}(Q_t-\mu_0)^{-\alpha_{1t}^0-1}+\alpha_{2t}^0(\sigma_t^0)^{\alpha_{2t}^0}(Q_t-\mu_0)^{-\alpha_{2t}^0-1}}\Big]\\
&-\frac{1}{n}\sum_{t=1}^{n}\Big[\frac{\big(\alpha_{1t}(\alpha_{1t}+1)\sigma_t^{\alpha_{1t}}(Q_t-\mu_n)^{-\alpha_{1t}-2}+\alpha_{2t}(\alpha_{2t}+1)\sigma_t^{\alpha_{2t}}(Q_t-\mu_n)^{-\alpha_{2t}-2}\big)^2}{\big(\alpha_{1t}\sigma_t^{\alpha_{1t}}(Q_t-\mu_n)^{-\alpha_{1t}-1}+\alpha_{2t}\sigma_t^{\alpha_{2t}}(Q_t-\mu_n)^{-\alpha_{2t}-1}\big)^2}\\
&-\frac{\Big(\alpha_{1t}^0(\alpha_{1t}^0+1)(\sigma_t^0)^{\alpha_{1t}^0}(Q_t-\mu_0)^{-\alpha_{1t}^0-2}+\alpha_{2t}^0(\alpha_{2t}^0+1)(\sigma_t^0)^{\alpha_{2t}^0}(Q_t-\mu_0)^{-\alpha_{2t}^0-2}\Big)^2}{\Big(\alpha_{1t}^0(\sigma_t^0)^{\alpha_{1t}^0}(Q_t-\mu_0)^{-\alpha_{1t}^0-1}+\alpha_{2t}^0(\sigma_t^0)^{\alpha_{2t}^0}(Q_t-\mu_0)^{-\alpha_{2t}^0-1}\Big)^2}\Big]\\
=:&M_1+M_2+M_3.
\end{aligned}
$$

For M_1, we have

$$
\begin{aligned}
&\Big|\frac{1}{n}\sum_{t=1}^{n}\big[\alpha_{1t}^0(\alpha_{1t}^0+1)(\sigma_t^0)^{\alpha_{1t}^0}(Q_t-\mu_0)^{-(\alpha_{1t}^0+2)}-\alpha_{1t}(\alpha_{1t}+1)\sigma_t^{\alpha_{1t}}(Q_t-\mu_n)^{-(\alpha_{1t}+2)}\\
&+\alpha_{2t}^0(\alpha_{2t}^0+1)(\sigma_t^0)^{\alpha_{2t}^0}(Q_t-\mu_0)^{-(\alpha_{2t}^0+2)}-\alpha_{2t}(\alpha_{2t}+1)\sigma_t^{\alpha_{2t}}(Q_t-\mu_n)^{-(\alpha_{2t}+2)}\big]\Big|\\
\le&\frac{1}{n}\sum_{t=1}^{n}\alpha_{1t}^0(\alpha_{1t}^0+1)(\sigma_t^0)^{\alpha_{1t}^0}\Big|(Q_t-\mu_0)^{-(\alpha_{1t}^0+2)}-(Q_t-\mu_n)^{-(\alpha_{1t}^0+2)}\Big|\\
&+\frac{1}{n}\sum_{t=1}^{n}\alpha_{1t}^0(\alpha_{1t}^0+1)(\sigma_t^0)^{\alpha_{1t}^0}\Big|(Q_t-\mu_n)^{-(\alpha_{1t}^0+2)}-(Q_t-\mu_n)^{-(\alpha_{1t}+2)}\Big|\\
&+\frac{1}{n}\sum_{t=1}^{n}\Big|\alpha_{1t}^0(\alpha_{1t}^0+1)(\sigma_t^0)^{\alpha_{1t}^0}-\alpha_{1t}(\alpha_{1t}+1)\sigma_t^{\alpha_{1t}}\Big|(Q_t-\mu_n)^{-(\alpha_{1t}+2)}\\
&+\frac{1}{n}\sum_{t=1}^{n}\alpha_{2t}^0(\alpha_{2t}^0+1)(\sigma_t^0)^{\alpha_{2t}^0}\Big|(Q_t-\mu_0)^{-(\alpha_{2t}^0+2)}-(Q_t-\mu_n)^{-(\alpha_{2t}^0+2)}\Big|\\
&+\frac{1}{n}\sum_{t=1}^{n}\alpha_{2t}^0(\alpha_{2t}^0+1)(\sigma_t^0)^{\alpha_{2t}^0}\Big|(Q_t-\mu_n)^{-(\alpha_{2t}^0+2)}-(Q_t-\mu_n)^{-(\alpha_{2t}+2)}\Big|\\
&+\frac{1}{n}\sum_{t=1}^{n}\Big|\alpha_{2t}^0(\alpha_{2t}^0+1)(\sigma_t^0)^{\alpha_{2t}^0}-\alpha_{2t}(\alpha_{2t}+1)\sigma_t^{\alpha_{2t}}\Big|(Q_t-\mu_n)^{-(\alpha_{2t}+2)}\\
=:&N_1+N_2+N_3+N_4+N_5+N_6.
\end{aligned}
$$

By Lemma 9(a) and Lemma 10(a), we know that $\sup_{1\le t\le n}|\alpha_{1t}-\alpha_{1t}^0| = O(\tau_n)$, $\sup_{1\le t\le n}|\alpha_{2t}-\alpha_{2t}^0| = O(\tau_n)$. N_1 and N_4 go to zero by Lemma 8. N_2 and N_5 go to zero by Lemma 11. N_3 and N_6 go to zero by the boundness of $\{\sigma_t, \alpha_{1t}, \alpha_{2t}\}$, the differentiable continuity of $\alpha_{1t}(\alpha_{1t}+1)\sigma_t^{\alpha_{1t}}$, $\alpha_{2t}(\alpha_{2t}+1)\sigma_t^{\alpha_{2t}}$ with respect to $\sigma_t, \alpha_{1t}, \alpha_{2t}$ and Lemma 9(a), Lemma 10(a).

For M_2 and M_3, through the proof process similar to the above, it is easy to verify that both of these two terms go to 0 in probability. □

We have already proved $\|L_n(\boldsymbol{\theta}) - L_n(\boldsymbol{\theta}_0)\| \to_p 0$ when $\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\| = O_p(\tau_n)$. Note that our ultimate goal is to establish uniform convergence result about $\tilde{L}_n(\boldsymbol{\theta})$. Lemmas 13 and 14 state that the impact of arbitrary initial value $(\tilde{\sigma}_1, \tilde{\alpha}_{11}, \tilde{\alpha}_{21})^T$ on the behavior of $\tilde{L}_n(\boldsymbol{\theta})$ is asymptotically negligible over a neighborhood of $\boldsymbol{\theta}_0$.

Lemma 13. *Under the conditions in Theorem 2, there exist positive constants C and $0 < C_b < 1$ such that for all $\boldsymbol{\theta} \in \Theta$, $t \ge 1$ and $i, j = 1, 2, 3, 4$,*

$$
\begin{aligned}
&(a)|\alpha_{1t}-\tilde{\alpha}_{1t}| \le CC_b^{t-1},\ |\alpha_{2t}-\tilde{\alpha}_{2t}| \le CC_b^{t-1},\ |\sigma_t-\tilde{\sigma}_t| \le CC_b^{t-1};\\
&(b)\left|\frac{\partial \alpha_{1t}}{\partial \Gamma_i}-\frac{\partial \tilde{\alpha}_{1t}}{\partial \Gamma_i}\right| \le tCC_b^{t-1},\ \left|\frac{\partial \alpha_{2t}}{\partial \Phi_i}-\frac{\partial \tilde{\alpha}_{2t}}{\partial \Phi_i}\right| \le tCC_b^{t-1},\ \left|\frac{\partial \sigma_t}{\partial \Psi_i}-\frac{\partial \tilde{\sigma}_t}{\partial \Psi_i}\right| \le tCC_b^{t-1};\\
&(c)\left|\frac{\partial^2 \alpha_{1t}}{\partial \Gamma_i\partial\Gamma_j}-\frac{\partial^2 \tilde{\alpha}_{1t}}{\partial \Gamma_i\partial\Gamma_j}\right| \le t^2CC_b^{t-1},\ \left|\frac{\partial^2 \alpha_{2t}}{\partial \Phi_i\partial\Phi_j}-\frac{\partial^2 \tilde{\alpha}_{2t}}{\partial \Phi_i\partial\Phi_j}\right| \le t^2CC_b^{t-1},\\
&\quad\left|\frac{\partial^2 \sigma_t}{\partial \Psi_i\partial\Psi_j}-\frac{\partial^2 \tilde{\sigma}_t}{\partial \Psi_i\partial\Psi_j}\right| \le t^2CC_b^{t-1}.
\end{aligned}
$$

Proof of Lemma 13. The proof follows by direct calculation, so we omit the details. □

Lemma 14. *Under the conditions in Theorem 2, we have*

$$
\frac{1}{n}\sum_{t=1}^{n}\left|(Q_t-\mu_n)^{-\alpha_{1t}}-(Q_t-\mu_n)^{-\tilde{\alpha}_{1t}}\right| \to_p 0,
$$

and

$$
\frac{1}{n}\sum_{t=1}^{n}\left|(Q_t-\mu_n)^{-\alpha_{2t}}-(Q_t-\mu_n)^{-\tilde{\alpha}_{2t}}\right| \to_p 0,
$$

uniformly over $|\mu_n - \mu_0| < \tau_n$*, where* $\tau_n \sim n^{-r}$*,* $r > 0$*. The same result holds for*

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

and

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

Proof of Lemma 14. We just give the proof for the case of $\frac{1}{n}\sum_{t=1}^{n}|(Q_t-\mu_n)^{-\alpha_{1t}} - (Q_t-\mu_n)^{-\tilde{\alpha}_{1t}}| \to_p 0$, the proofs for other cases are similar. By Lemma 13(a), we have $|\alpha_{1t} - \tilde{\alpha}_{1t}| \le CC_b^{t-1}$. Assume that $\tilde{\alpha}_{1t} > \alpha_{1t}$, the proof for the other direction is the same. By mean value theorem,

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n}\left|(Q_t-\mu_n)^{-\alpha_{1t}}-(Q_t-\mu_n)^{-\tilde{\alpha}_{1t}}\right| &\le \frac{C}{n}\sum_{t=1}^{n}(Q_t-\mu_n)^{-\alpha_{1t}^*}|\log(Q_t-\mu_n)|C_b^{t-1}\\ \le&\frac{C}{n}\sum_{t=1}^{n}\{(Q_t-\mu_n)^{-\alpha_{1L}}+(Q_t-\mu_n)^{-\alpha_{1U}}\}|\log(Q_t-\mu_n)|C_b^{t-1}\to_p 0,\end{aligned}$$

where $\alpha_{1t}^* \in (\alpha_{1t}, \tilde{\alpha}_{1t})$. The result follows Lemma 8 and that

$$E_{\boldsymbol{\theta}_0}\left[\sum_{t=1}^{n}((Q_t-\mu_n)^{-\alpha_{1L}}+(Q_t-\mu_n)^{-\alpha_{1U}})|\log(Q_t-\mu_n)|C_b^{t-1}\right]<\infty.$$

Then we obtain our conclusion for Lemma 14. □

Lemma 15 can be utilized to prove Theorems 2 and 3 in the article.

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

(a) for all second order derivatives of $\tilde{L}_n(\boldsymbol{\theta})$*, we have* $\frac{\partial^2}{\partial\theta_i\partial\theta_j}\tilde{L}_n(\boldsymbol{\theta}) \to_p -m_{\theta_i\theta_j}(\boldsymbol{\theta}_0)$*, uniformly over* $\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\| < \tau_n$*, where* $\tau_n \sim n^{-r}$*,* $r > 0$*.*

(b) for the score function of $\tilde{L}_n(\boldsymbol{\theta})$*, we have* $(\tau_n^*)^{-1}\left(\frac{\partial}{\partial\boldsymbol{\theta}}\tilde{L}_n(\boldsymbol{\theta}_0) - \frac{\partial}{\partial\boldsymbol{\theta}}L_n(\boldsymbol{\theta}_0)\right) \to_p 0$ *if* $n\tau_n^* \to \infty$.

Proof of Lemma 15. For the proof of part (a), first, we see that $\frac{\partial^2}{\partial\theta_i\partial\theta_j}\tilde{L}_n(\boldsymbol{\theta}) - \frac{\partial^2}{\partial\theta_i\partial\theta_j}L_n(\boldsymbol{\theta}) \to_p 0$ holds for any $\boldsymbol{\theta}$ by Lemma 13 and Lemma 14. In addition, we also know that $\frac{\partial^2}{\partial\theta_i\partial\theta_j}L_n(\boldsymbol{\theta}) -$

$\frac{\partial^2}{\partial\theta_i\partial\theta_j}L_n(\boldsymbol{\theta}_0) \to_p 0$ holds uniformly over $\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\| < \tau_n$ by Lemma 12. Thus, we get $\frac{\partial^2}{\partial\theta_i\partial\theta_j}\tilde{L}_n(\boldsymbol{\theta})-\frac{\partial^2}{\partial\theta_i\partial\theta_j}L_n(\boldsymbol{\theta}_0) \to_p 0$ over $\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\| < \tau_n$.

For the proof of part (b), here we just prove the case of $\frac{\partial}{\partial\mu}\tilde{L}_n(\boldsymbol{\theta}_0)$, the proof for other first order partial derivatives are similar. For convenience, we define $g(x,y) = yx^y$ and $h(x,y) = y(y+1)x^y$ for $x,y>0$. By the fact that $|\sigma_t - \tilde{\sigma}_t| \le CC_b^{t-1}$, $|\alpha_{1t}-\tilde{\alpha}_{1t}| \le CC_b^{t-1}$, $|\alpha_{2t}-\tilde{\alpha}_{2t}| \le CC_b^{t-1}$, we can verify that $|g(\sigma_t,\alpha_{1t})-g(\tilde{\sigma}_t,\tilde{\alpha}_{1t})| \le CC_b^{t-1}$, $|g(\sigma_t,\alpha_{2t})-g(\tilde{\sigma}_t,\tilde{\alpha}_{2t})| \le CC_b^{t-1}$, $|h(\sigma_t,\alpha_{1t})-h(\tilde{\sigma}_t,\tilde{\alpha}_{1t})| \le CC_b^{t-1}$ and $|h(\sigma_t,\alpha_{2t})-h(\tilde{\sigma}_t,\tilde{\alpha}_{2t})| \le CC_b^{t-1}$ hold by utilizing Lemma 13 after assuming σ_L is greater than zero. Next, we obtain

$$
\begin{aligned}
&\frac{1}{\tau_n^*}\left(\frac{\partial}{\partial\mu}\tilde{L}_n(\boldsymbol{\theta}_0)-\frac{\partial}{\partial\mu}L_n(\boldsymbol{\theta}_0)\right)\\
=&\frac{1}{n\tau_n^*}\sum_{t=1}^{n}\Bigg[\frac{\tilde{\alpha}_{1t}(\tilde{\alpha}_{1t}+1)\tilde{\sigma}_t^{\tilde{\alpha}_{1t}}(Q_t-\mu_0)^{-(\tilde{\alpha}_{1t}+2)}+\tilde{\alpha}_{2t}(\tilde{\alpha}_{2t}+1)\tilde{\sigma}_t^{\tilde{\alpha}_{2t}}(Q_t-\mu_0)^{-(\tilde{\alpha}_{2t}+2)}}{\tilde{\alpha}_{1t}\tilde{\sigma}_t^{\tilde{\alpha}_{1t}}(Q_t-\mu_0)^{-(\tilde{\alpha}_{1t}+1)}+\tilde{\alpha}_{2t}\tilde{\sigma}_t^{\tilde{\alpha}_{2t}}(Q_t-\mu_0)^{-(\tilde{\alpha}_{2t}+1)}}\\
&\qquad -\tilde{\alpha}_{1t}(\tilde{\sigma}_t)^{\tilde{\alpha}_{1t}}(Q_t-\mu_0)^{-(\tilde{\alpha}_{1t}+1)}-\tilde{\alpha}_{2t}(\tilde{\sigma}_t)^{\tilde{\alpha}_{2t}}(Q_t-\mu_0)^{-(\tilde{\alpha}_{2t}+1)}\\
&\qquad -\frac{\alpha_{1t}(\alpha_{1t}+1)\sigma_t^{\alpha_{1t}}(Q_t-\mu_0)^{-(\alpha_{1t}+2)}+\alpha_{2t}(\alpha_{2t}+1)\sigma_t^{\alpha_{2t}}(Q_t-\mu_0)^{-(\alpha_{1t}+2)}}{\alpha_{1t}\sigma_t^{\alpha_{1t}}(Q_t-\mu_0)^{-(\alpha_{1t}+1)}+\alpha_{2t}\sigma_t^{\alpha_{2t}}(Q_t-\mu_0)^{-(\alpha_{2t}+1)}}\\
&\qquad +\alpha_{1t}\sigma_t^{\alpha_{1t}}(Q_t-\mu_0)^{(-\alpha_{1t}+1)}+\alpha_{2t}\sigma_t^{\alpha_{2t}}(Q_t-\mu_0)^{(-\alpha_{2t}+1)}\Bigg]\\
=&\frac{1}{n\tau_n^*}\sum_{t=1}^{n}\left[\frac{g(\sigma_t,\alpha_{1t})}{(Q_t-\mu_0)^{\alpha_{1t}+1}}-\frac{g(\tilde{\sigma}_t,\tilde{\alpha}_{1t})}{(Q_t-\mu_0)^{\tilde{\alpha}_{1t}+1}}\right]+\frac{1}{n\tau_n^*}\sum_{t=1}^{n}\left[\frac{g(\sigma_t,\alpha_{2t})}{(Q_t-\mu_0)^{\alpha_{2t}+1}}-\frac{g(\tilde{\sigma}_t,\tilde{\alpha}_{2t})}{(Q_t-\mu_0)^{\tilde{\alpha}_{2t}+1}}\right]\\
&+\frac{1}{n\tau_n^*}\sum_{t=1}^{n}\left[\frac{\frac{h(\tilde{\sigma}_t,\tilde{\alpha}_{1t})}{(Q_t-\mu_0)^{\tilde{\alpha}_{1t}+2}}+\frac{h(\tilde{\sigma}_t,\tilde{\alpha}_{2t})}{(Q_t-\mu_0)^{\tilde{\alpha}_{2t}+2}}}{\frac{g(\tilde{\sigma}_t,\tilde{\alpha}_{1t})}{(Q_t-\mu_0)^{\tilde{\alpha}_{1t}+1}}+\frac{g(\tilde{\sigma}_t,\tilde{\alpha}_{2t})}{(Q_t-\mu_0)^{\tilde{\alpha}_{2t}+1}}}-\frac{\frac{h(\sigma_t,\alpha_{1t})}{(Q_t-\mu_0)^{\alpha_{1t}+2}}+\frac{h(\sigma_t,\alpha_{2t})}{(Q_t-\mu_0)^{\alpha_{2t}+2}}}{\frac{g(\sigma_t,\alpha_{1t})}{(Q_t-\mu_0)^{\alpha_{1t}+1}}+\frac{g(\sigma_t,\alpha_{2t})}{(Q_t-\mu_0)^{\alpha_{2t}+1}}}\right]\\
=:&I_1+I_2+I_3.
\end{aligned}
$$

We have

$$
\begin{aligned}
I_1 =&\frac{1}{n\tau_n^*}\sum_{t=1}^{n}\left[\frac{g(\sigma_t,\alpha_{1t})-g(\tilde{\sigma}_t,\tilde{\alpha}_{1t})}{(Q_t-\mu_0)^{\alpha_{1t}+1}}\right]\\
&+\frac{1}{n\tau_n^*}\sum_{t=1}^{n}\left[g(\tilde{\sigma}_t,\tilde{\alpha}_{1t})\left[(Q_t-\mu_0)^{-(\alpha_{1t}+1)}-(Q_t-\mu_0)^{-(\tilde{\alpha}_{1t}+1)}\right]\right].
\end{aligned}
$$

In the above equation, the first term is bounded by $\frac{C}{n\tau_n^*}\sum_{t=1}^{n}\frac{C_b^{t-1}}{(Q_t-\mu_0)^{\alpha_{1t}+1}}$, and it goes to zero in probability since $n\tau_n^*\to\infty$ and $\mathrm{E}_{\boldsymbol{\theta}_0}\left[\sum_{t=1}^{\infty}C_b^{t-1}(Q_t-\mu_0)^{-\alpha}\right]<\infty$ for all $\alpha>0$. The same argument applies to the second term after applying mean value theorem. Then we obtain $I_1 \to_p 0$.

Through the same proof, it is easy to verify that $I_2 \to_p 0$ and $I_3 \to_p 0$ as $n\tau_n^* \to \infty$. $\square$

Lemma 16 gives the standard martingale CLT, which displays the asymptotic distribution of the score function.

Lemma 16. *Under the conditions in Theorem 2,*

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \xrightarrow{d} N(\mathbf{0}, \mathbf{M}_0{}^{-1}),$$

where $\mathbf{M}_0$ *is the Fisher information matrix at* $\boldsymbol{\theta}_0$.

Proof of Lemma 16. We prove this result by using CLT for martingale difference in Billingsley (1961), we have

$$\mathrm{E}_{\boldsymbol{\theta}_0}\left[\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\bigg|\mathcal{F}_{t-1}\right] = \mathbf{0},\ \mathrm{VaR}_{\boldsymbol{\theta_0}}\left(\frac{\partial \mathbf{l_t}(\boldsymbol{\theta_0})}{\partial \boldsymbol{\theta}}\right) = \mathbf{M_0} < \infty.$$

So for $\boldsymbol{\lambda}^T \in \mathbb{R}^{13}$, $\{\boldsymbol{\lambda}^T \frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}, \mathcal{F}_t\}$ is a square-integrable stationary martingale difference. Note that the sequences $\{\sigma_t, \alpha_{1t}, \alpha_{2t}\}$ and $\{Q_t\}$ are both stationary and ergodic, so the sequences $\frac{\partial \alpha_{1t}}{\partial \Gamma}$, $\frac{\partial \alpha_{2t}}{\partial \Phi}$ and $\frac{\partial \sigma_t}{\partial \Psi}$ are also strictly stationary and ergodic. We also know that $\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \theta_i}$, for $i = 1, \cdots, 13$, are generated from $\sigma_t, \alpha_{1t}, \alpha_{2t}, Q_t, \frac{\partial \alpha_{1t}}{\partial \Gamma}, \frac{\partial \alpha_{2t}}{\partial \Phi}, \frac{\partial \sigma_t}{\partial \Psi}$, so they also follow the properties of strict stationarity and ergodicity. Then by CLT of Billingsley (1961) and Wold-Cramér device, we obtain the conclusion of Lemma 16. $\square$

Proof of Theorem 2. Let $\{\tau_n\}$ be any sequence such that $\tau_n \sim n^{-r}$ and $n^{1/2}\tau_n \to \infty$, i.e. $0 < r < 1/2$. Let $t \in \mathbb{R}, \boldsymbol{y} \in \mathbb{R}^{12}$ and define $f_n(t, \boldsymbol{y}) = \tau_n^{-2}\tilde{L}_n(\mu_0 + \tau_n t, \boldsymbol{\phi}^0 + \tau_n \boldsymbol{y})$. We denote $\boldsymbol{\phi}^0 = (\beta_0^0, \beta_1^0, \beta_2^0, \beta_3^0, \gamma_0^0, \gamma_1^0, \gamma_2^0, \gamma_3^0, \delta_0^0, \delta_1^0, \delta_2^0, \delta_3^0)^T$.

By Taylor expansion we have,

$$\begin{aligned}\frac{\partial}{\partial t} f_n(t, \boldsymbol{y}) =& \tau_n^{-1} \frac{\partial \tilde{L}_n(\mu_0 + \tau_n t, \boldsymbol{\phi}^0 + \tau_n \boldsymbol{y})}{\partial \mu} \\ =& \tau_n^{-1} \frac{\partial \tilde{L}_n(\mu_0, \boldsymbol{\phi}^0)}{\partial \mu} + \frac{\partial^2 \tilde{L}_n(\mu^*, \boldsymbol{\phi}^*)}{\partial \mu^2} t + \sum_{i=1}^{12} \frac{\partial^2 \tilde{L}_n(\mu^*, \boldsymbol{\phi}^*)}{\partial \mu \partial \boldsymbol{\phi}_i} y_i \\ =& \tau_n^{-1} \left(\frac{\partial \tilde{L}_n(\mu_0, \boldsymbol{\phi}^0)}{\partial \mu} - \frac{\partial L_n(\mu_0, \boldsymbol{\phi}^0)}{\partial \mu} \right) + \tau_n^{-1} \left(\frac{\partial L_n(\mu_0, \boldsymbol{\phi}^0)}{\partial \mu} \right) \\ &+ \frac{\partial^2 \tilde{L}_n(\mu^*, \boldsymbol{\phi}^*)}{\partial \mu^2} t + \sum_{i=1}^{12} \frac{\partial^2 \tilde{L}_n(\mu^*, \boldsymbol{\phi}^*)}{\partial \mu \partial \boldsymbol{\phi}_i} y_i,\end{aligned}$$

where the second equality comes from the Taylor expansion of $\frac{\partial \tilde{L}_n(\mu_0 + \tau_n t, \boldsymbol{\phi}^0 + \tau_n \boldsymbol{y})}{\partial \mu}$ at $(\mu_0, \boldsymbol{\phi}^0)$, and we have $|\mu^* - \mu_0| < \tau_n t$, $\left\|\boldsymbol{\phi}^* - \boldsymbol{\phi}^0\right\| < \tau_n \|\boldsymbol{y}\|$ due to mean value theorem. Hence, the first term goes to zero by Lemma 15(b). The second term goes to zero by Lemma 16 and the fact that $\sqrt{n}\tau_n \to \infty$. By Lemma 15(a), the last two terms converge uniformly over $t^2 + \|\boldsymbol{y}\|^2 \leq 1$, i.e.

$$\frac{\partial^2 \tilde{L}_n(\mu^*, \boldsymbol{\phi}^*)}{\partial \mu^2} t + \sum_{i=1}^{12} \frac{\partial^2 \tilde{L}_n(\mu^*, \boldsymbol{\phi}^*)}{\partial \mu \partial \boldsymbol{\phi}_i} y_i \to_p -m_{\mu\mu}(\boldsymbol{\theta}_0) t - \sum_{i=1}^{12} m_{\mu \boldsymbol{\phi}_i}(\boldsymbol{\theta}_0) y_i,$$

in which $m_{\mu\mu}(\boldsymbol{\theta}_0) = -E_{\boldsymbol{\theta}_0}[\frac{\partial^2}{\partial \mu^2} l_1(\boldsymbol{\theta}_0)]$ and $m_{\mu \boldsymbol{\phi}_i}(\boldsymbol{\theta}_0) = -E_{\boldsymbol{\theta}_0}[\frac{\partial^2}{\partial \mu \partial \boldsymbol{\phi}_i} l_1(\boldsymbol{\theta}_0)]$.

Then we obtain

$$\frac{\partial}{\partial t} f_n(t, \boldsymbol{y}) = -m_{\mu\mu}(\boldsymbol{\theta}_0) t - \sum_{i=1}^{12} m_{\mu \boldsymbol{\phi}_i}(\boldsymbol{\theta}_0) y_i + o_p(1).$$

Similarly, we obtain

$$\frac{\partial}{\partial y_i} f_n(t, \boldsymbol{y}) = -m_{\mu \boldsymbol{\phi}_i}(\boldsymbol{\theta}_0) t - \sum_{i=1}^{12} m_{\boldsymbol{\phi}_i \boldsymbol{\phi}_j}(\boldsymbol{\theta}_0) + o_p(1), \quad \text{for } i = 1, 2, \cdots, 12,$$

where $o_p(1)$'s decay uniformly over $t^2 + \|\boldsymbol{y}\|^2 \leq 1$. Let $t^2 + \|\boldsymbol{y}\|^2 = 1$, and we have

$$\begin{aligned}& t \frac{\partial f_n(t, \boldsymbol{y})}{\partial t} + \sum_{i=1}^{12} y_i \frac{\partial f_n(t, \boldsymbol{y})}{\partial y_i} \\ =& - t^2 m_{\mu\mu}(\boldsymbol{\theta}_0) - 2t \sum_{i=1}^{12} y_i m_{\mu \boldsymbol{\phi}_i}(\boldsymbol{\theta}_0) - \sum_{i=1}^{12} \sum_{j=1}^{12} y_i y_j m_{\boldsymbol{\phi}_i \boldsymbol{\phi}_j}(\boldsymbol{\theta}_0) + o_p(1) < 0,\end{aligned}$$

where the negative sign follows from the fact that the Fisher information matrix $\mathbf{M}_0$ is positive definite. According to the above arguments and Lemma 5 in Smith (1985), we obtain that f_n has a local maximum over the open set $t^2 + \|\boldsymbol{y}\|^2 < 1$ with probability going to one. So there exists a sequence of local maximizer $\hat{\boldsymbol{\theta}}_n$ of $\tilde{L}_n$ such that $\hat{\boldsymbol{\theta}}_n \to_p \boldsymbol{\theta}_0$ and $||\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0|| \leq \tau_n$, where $\tau_n \sim n^{-r}$, $0 < r < 1/2$. □

Proof of Theorem 3. By Taylor expansion, we obtain

$$\mathbf{0} = \frac{\partial \tilde{L}_n(\hat{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta}} = \frac{\partial \tilde{L}_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} + \frac{\partial^2 \tilde{L}_n(\boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0),$$

where $\boldsymbol{\theta}^* = \lambda \hat{\boldsymbol{\theta}}_n + (1-\lambda)\boldsymbol{\theta}_0$ with $0 \leq \lambda \leq 1$. Therefore, we obtain

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = -\left(\frac{\partial^2 \tilde{L}_n(\boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}\right)^{-1} \sqrt{n}\frac{\partial \tilde{L}_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}},$$

in which we have $-\frac{\partial^2 \tilde{L}_n(\boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \to_p I(\boldsymbol{\theta}_0) = -\mathrm{E}_{\boldsymbol{\theta}_0}[\frac{\partial^2 l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}]$ by Lemma 15(a). In addition, $\sqrt{n}\frac{\partial \tilde{L}_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}$ converges to $N(\mathbf{0}, I(\boldsymbol{\theta}_0))$ in distribution by Lemmas 15(b) and 16. In the end, after utilizing Slutsky theorem, we conclude that $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}, \mathbf{M}_0^{-1})$. □

Proof of Proposition 2. We use δ to denote a generic small positive value and $\boldsymbol{\phi} = (\beta_0, \beta_1, \beta_2, \beta_3, \gamma_0, \gamma_1, \gamma_2, \gamma_3, \delta_0, \delta_1, \delta_2, \delta_3)^T$. Recall that in Proposition 2, $V_n = \{\boldsymbol{\theta} \in \Theta \mid \mu \leq cQ_{n,1} + (1-c)\mu_0\}$. For any $0 < c < 1$, we have $\mu_0 < cQ_{n,1} + (1-c)\mu_0 < Q_{n,1}$ and $(Q_{n,1} + (1-c)\mu_0) \searrow \mu_0$ a.s. Denote $\Theta_n^\delta = \{\boldsymbol{\theta} \in V_n \mid ||\boldsymbol{\theta} - \boldsymbol{\theta}_0|| \geq \delta\}$, $\Theta_n^\mu = \{\boldsymbol{\theta} \in V_n \mid ||\boldsymbol{\theta} - \boldsymbol{\theta}_0|| \geq \delta, \mu > \mu_0\}$ and $\Theta^\delta = \{\boldsymbol{\theta} \in V_n \mid ||\boldsymbol{\theta} - \boldsymbol{\theta}_0|| \geq \delta, \mu \leq \mu_0\}$. Obviously, $\Theta_n^\delta = \Theta_n^\mu \cup \Theta^\delta$.

We first prove that

$$\text{(I)} \quad \text{for any } \delta > 0,\ P(\sup_{\Theta_n^\delta} \tilde{L}_n(\boldsymbol{\theta}) \geq \tilde{L}_n(\boldsymbol{\theta}_0)) \to 0 \text{ as } n \to \infty.$$

By the same argument in Lemmas 8 and 14, it can be proved that $\sup_{\Theta_n^\delta} |\tilde{L}_n(\boldsymbol{\theta}) - L_n(\boldsymbol{\theta})| \to_p$

0 and $\sup_{\Theta_n^{\mu}} |L_n(\mu, \boldsymbol{\phi}) - L_n(\mu_0, \boldsymbol{\phi})| \to_p 0$ as $n \to \infty$. Then we obtain

$$\begin{aligned}\sup_{\Theta_n^{\delta}} \tilde{L}_n(\boldsymbol{\theta}) =& \sup_{\Theta_n^{\delta}} L_n(\boldsymbol{\theta}) + o_p(1) = \max\big(\sup_{\Theta^{\delta}} L_n(\boldsymbol{\theta}), \sup_{\Theta_n^{\mu}} L_n(\boldsymbol{\theta})) + o_p(1) \\ =& \max\big(\sup_{\Theta^{\delta}} L_n(\boldsymbol{\theta}), \sup_{\Theta_n^{\mu}} L_n(\mu_0, \boldsymbol{\phi})\big) + o_p(1) \le \sup_{\Theta^{\delta/2}} L_n(\boldsymbol{\theta}) + o_p(1).\end{aligned}$$

The last inequality follows from the fact that $Q_{n,1} \searrow \mu_0$ a.s.. With probability going to one, we have $\{\boldsymbol{\phi}|\boldsymbol{\phi} \in \Theta_n^{\mu}\} \subseteq \{\boldsymbol{\phi}|\boldsymbol{\phi} \in \Theta^{\delta/2}\}$. Following similar proof procedures given in Lemmas 14 and 15, we have $\tilde{L}_n(\boldsymbol{\theta}_0) = L_n(\boldsymbol{\theta}_0) + o_p(1) \to_p \mathrm{E}_{\boldsymbol{\theta}_0}[l_1(\boldsymbol{\theta}_0)]$. The rest proof for (I) follows from the proof of Proposition 2 in Dombry et al. (2015).

Denote $\Theta_n^{\delta c} = \{\boldsymbol{\theta} \in V_n| \, \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| < \delta\}$, $\Theta_n^{\mu c} = \{\boldsymbol{\theta} \in V_n| \, \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| < \delta, \mu > \mu_0\}$ and $\Theta^{\delta c} = \{\boldsymbol{\theta} \in V_n| \, \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| < \delta, \mu \le \mu_0\}$. Note that $\Theta_n^{\delta c} = \Theta_n^{\mu c} \cup \Theta^{\delta c}$.

Next we prove that there exists a $\delta^* > 0$ such that, as $n \to \infty$,

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

According to our results given in Lemmas 8 and 14, we obtain

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

and

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

Since $l_t(\boldsymbol{\theta})$ is a continuous function, it can be proved that $\sup_{\Theta^{\delta c}} \left| \frac{\partial^2}{\partial \theta_i \partial \theta_j} l_t(\boldsymbol{\theta}) \right|$ is integrable. Therefore, by the properties of stationarity and ergodicity and the uniform law of large numbers, we obtain

$$\sup_{\Theta^{\delta c}} \left| \frac{\partial^2}{\partial \theta_i \partial \theta_j} L_n(\mu, \boldsymbol{\phi}) - \mathrm{E}_{\boldsymbol{\theta}_0}\left(\frac{\partial^2}{\partial \theta_i \partial \theta_j} l_1(\mu, \boldsymbol{\phi})\right) \right| \to_p 0, \text{ as } n \to \infty.$$

So $\mathrm{E}_{\boldsymbol{\theta}_0}\left(\frac{\partial^2}{\partial \theta_i \partial \theta_j} l_1(\boldsymbol{\theta}_0)\right) = -(\mathbf{M_0})_{ij}$, where $\mathbf{M_0}$ is positive definite by Lemma 5.

Furthermore, the function $\mathrm{E}_{\boldsymbol{\theta}_0}\left(\frac{\partial^2}{\partial \theta_i \partial \theta_j} l_1(\boldsymbol{\theta})\right)$ is continuous, so there exists a $\delta^* > 0$ such that $\mathrm{E}_{\boldsymbol{\theta}_0}\left(\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} l_1(\boldsymbol{\theta})\right)$ is negative definite for all $\boldsymbol{\theta} \in \Theta^{\delta^* c}$. Together the above arguments, we can prove (II).

Utilizing our conclusion from (I), we obtain that the global maximizer of $\tilde{L}_n(\boldsymbol{\theta})$ over V_n is located within $\Theta_n^{\delta^* c}$ with probability going to one. By Theorem 2, we know that there exists a sequence $\hat{\boldsymbol{\theta}}_n$ of local maximizer of $\tilde{L}_n(\boldsymbol{\theta})$ such that $||\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0|| \le \tau_n$, where $\tau_n = O_p(n^{-r})$, $0 < r < 1/2$. So we have $P(\hat{\boldsymbol{\theta}}_n \in \Theta_n^{\delta^* c}) \to 1$. Also, we have $\frac{\partial}{\partial \boldsymbol{\theta}} \tilde{L}_n(\hat{\boldsymbol{\theta}}_n) = \mathbf{0}$. Combining with our proof of (II) and using the conclusion of Theorem 2.6 in Mäkeläinen et al. (1981), we finally claim the conclusion of Proposition 2. □

In the next section we will illustrate expressions of the first order and the second order partial derivatives of the likelihood function.

S4 First and the second order partial derivatives of $l_t(\boldsymbol{\theta})$

In this section, we give the formulas for $\frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ and $\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}$. Denote $\Gamma = (\gamma_0, \gamma_1, \gamma_2, \gamma_3)^T$, i.e., we use $\boldsymbol{\Gamma}$ as a generic symbol for $(\gamma_0, \gamma_1, \gamma_2, \gamma_3)^T$. Similarly, we set $\boldsymbol{\Phi} = (\delta_0, \delta_1, \delta_2, \delta_3)^T$ and $\boldsymbol{\Psi} = (\beta_0, \beta_1, \beta_2, \beta_3)^T$.

The log-likelihood function is

$$\begin{aligned} l_t(\boldsymbol{\theta}) &= \log\left[\alpha_{1t}\sigma_t^{\alpha_{1t}}(Q_t-\mu)^{-\alpha_{1t}-1} + \alpha_{2t}\sigma_t^{\alpha_{2t}}(Q_t-\mu)^{-\alpha_{2t}-1}\right] - \sigma_t^{\alpha_{1t}}(Q_t-\mu)^{-\alpha_{1t}} - \sigma_t^{\alpha_{2t}}(Q_t-\mu)^{-\alpha_{2t}} \\ &= \log\left[\frac{\alpha_{1t}}{\sigma_t}\left(\frac{Q_t-\mu}{\sigma_t}\right)^{-\alpha_{1t}-1} + \frac{\alpha_{2t}}{\sigma_t}\left(\frac{Q_t-\mu}{\sigma_t}\right)^{-\alpha_{2t}-1}\right] - \left(\frac{Q_t-\mu}{\sigma_t}\right)^{-\alpha_{1t}} - \left(\frac{Q_t-\mu}{\sigma_t}\right)^{-\alpha_{2t}}. \end{aligned}$$

S4. FIRST AND THE SECOND ORDER PARTIAL DERIVATIVES OF $L_T(\boldsymbol{\theta})$

For the first order partial derivatives, we have, for $i = 1, 2, 3, 4$,

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

$$\frac{\partial l_t(\boldsymbol{\theta})}{\partial \Gamma_i} = \left[\frac{\left(\frac{Q_t-\mu}{\sigma_t}\right)^{-(\alpha_{1t}+1)} - \alpha_{1t}\left(\frac{Q_t-\mu}{\sigma_t}\right)^{-(\alpha_{1t}+1)}\log\left(\frac{Q_t-\mu}{\sigma_t}\right)}{\alpha_{1t}\left(\frac{Q_t-\mu}{\sigma_t}\right)^{-(\alpha_{1t}+1)} + \alpha_{2t}\left(\frac{Q_t-\mu}{\sigma_t}\right)^{-(\alpha_{2t}+1)}} + \left(\frac{Q_t-\mu}{\sigma_t}\right)^{-\alpha_{1t}}\log\left(\frac{Q_t-\mu}{\sigma_t}\right)\right]\frac{\partial \alpha_{1t}}{\partial \Gamma_i},$$

$$\frac{\partial l_t(\boldsymbol{\theta})}{\partial \Phi_i} = \left[\frac{\left(\frac{Q_t-\mu}{\sigma_t}\right)^{-(\alpha_{2t}+1)} - \alpha_{2t}\left(\frac{Q_t-\mu}{\sigma_t}\right)^{-(\alpha_{2t}+1)}\log\left(\frac{Q_t-\mu}{\sigma_t}\right)}{\alpha_{1t}\left(\frac{Q_t-\mu}{\sigma_t}\right)^{-(\alpha_{1t}+1)} + \alpha_{2t}\left(\frac{Q_t-\mu}{\sigma_t}\right)^{-(\alpha_{2t}+1)}} + \left(\frac{Q_t-\mu}{\sigma_t}\right)^{-\alpha_{2t}}\log\left(\frac{Q_t-\mu}{\sigma_t}\right)\right]\frac{\partial \alpha_{2t}}{\partial \Phi_i},$$

and

$$\frac{\partial l_t(\boldsymbol{\theta})}{\partial \Psi_i} = \Bigg[\frac{\alpha_{1t}^2\left(\frac{Q_t-\mu}{\sigma_t}\right)^{-(\alpha_{1t}+1)} + \alpha_{2t}^2\left(\frac{Q_t-\mu}{\sigma_t}\right)^{-(\alpha_{2t}+1)}}{\alpha_{1t}\sigma_t\left(\frac{Q_t-\mu}{\sigma_t}\right)^{-(\alpha_{1t}+1)} + \alpha_{2t}\sigma_t\left(\frac{Q_t-\mu}{\sigma_t}\right)^{-(\alpha_{2t}+1)}} - \frac{\alpha_{1t}}{\sigma_t}\left(\frac{Q_t-\mu}{\sigma_t}\right)^{-\alpha_{1t}}$$
$$- \frac{\alpha_{2t}}{\sigma_t}\left(\frac{Q_t-\mu}{\sigma_t}\right)^{-\alpha_{2t}}\Bigg]\frac{\partial \sigma_t}{\partial \Psi_i}.$$

For the second order partial derivatives, we have, for $i, j = 1, 2, 3, 4$,

$$\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial \mu^2} = \frac{\alpha_{1t}(\alpha_{1t}+1)(\alpha_{1t}+2)\sigma_t^{\alpha_{1t}}(Q_t-\mu)^{-\alpha_{1t}-3} + \alpha_{2t}(\alpha_{2t}+1)(\alpha_{2t}+2)\sigma_t^{\alpha_{2t}}(Q_t-\mu)^{-\alpha_{2t}-3}}{\alpha_{1t}\sigma_t^{\alpha_{1t}}(Q_t-\mu)^{-\alpha_{1t}-1} + \alpha_{2t}\sigma_t^{\alpha_{2t}}(Q_t-\mu)^{-\alpha_{2t}-1}}$$
$$- \frac{\left(\alpha_{1t}(\alpha_{1t}+1)\sigma_t^{\alpha_{1t}}(Q_t-\mu)^{-\alpha_{1t}-2} + \alpha_{2t}(\alpha_{2t}+1)\sigma_t^{\alpha_{2t}}(Q_t-\mu)^{-\alpha_{2t}-2}\right)^2}{(\alpha_{1t}\sigma_t^{\alpha_{1t}}(Q_t-\mu)^{-\alpha_{1t}-1} + \alpha_{2t}\sigma_t^{\alpha_{2t}}(Q_t-\mu)^{-\alpha_{2t}-1})^2}$$
$$- \alpha_{1t}(\alpha_{1t}+1)\sigma_t^{\alpha_{1t}}(Q_t-\mu)^{-\alpha_{1t}-2} - \alpha_{2t}(\alpha_{2t}+1)\sigma_t^{\alpha_{2t}}(Q_t-\mu)^{-\alpha_{2t}-2},$$

$$\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial \mu \partial \Gamma_i} = \frac{\sigma^{\alpha_{1t}}(Q_t-\mu)^{-1-\alpha_{1t}}}{\left(\alpha_{2t}\left(\frac{Q_t-\mu}{\sigma_t}\right)^{\alpha_{1t}} + \alpha_{1t}\left(\frac{Q_t-\mu}{\sigma_t}\right)^{\alpha_{2t}}\right)^2} \Bigg\{ \alpha_{2t}^2 \left(\frac{Q_t-\mu}{\sigma_t}\right)^{2\alpha_{1t}} \left(-1 + \alpha_{1t}\log\left(\frac{Q_t-\mu}{\sigma_t}\right)\right)$$
$$+ \alpha_{1t}^2 \left(\frac{Q_t-\mu}{\sigma_t}\right)^{2\alpha_{2t}} \left(-1 + \left(\frac{Q_t-\mu}{\sigma_t}\right)^{\alpha_{1t}} + \alpha_{1t}\log\left(\frac{Q_t-\mu}{\sigma_t}\right)\right)$$
$$+ \alpha_{2t}\left(\frac{Q_t-\mu}{\sigma_t}\right)^{\alpha_{1t}+\alpha_{2t}} \Bigg[2\alpha_{1t}\bigg(-1 + \alpha_{1t}\log\left(\frac{Q_t-\mu}{\sigma_t}\right)\bigg)$$
$$- \left(\frac{Q_t-\mu}{\sigma_t}\right)^{\alpha_{1t}} \left(-2\alpha_{1t} + \alpha_{2t} + \alpha_{1t}(\alpha_{1t}-\alpha_{2t})\log\left(\frac{Q_t-\mu}{\sigma_t}\right)\right)\Bigg]\Bigg\} \frac{\partial \alpha_{1t}}{\partial \Gamma_i},$$

$$\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial \mu \partial \Phi_i} = \frac{\sigma^{\alpha_{2t}}(Q_t-\mu)^{-1-\alpha_{2t}}}{\left(\alpha_{2t}\left(\frac{Q_t-\mu}{\sigma_t}\right)^{\alpha_{1t}} + \alpha_{1t}\left(\frac{Q_t-\mu}{\sigma_t}\right)^{\alpha_{2t}}\right)^2} \Bigg\{ \alpha_{1t}^2 \left(\frac{Q_t-\mu}{\sigma_t}\right)^{2\alpha_{2t}} \left(-1 + \alpha_{2t}\log\left(\frac{Q_t-\mu}{\sigma_t}\right)\right)$$
$$+ \alpha_{2t}^2 \left(\frac{Q_t-\mu}{\sigma_t}\right)^{2\alpha_{1t}} \left(-1 + \left(\frac{Q_t-\mu}{\sigma_t}\right)^{\alpha_{2t}} + \alpha_{2t}\log\left(\frac{Q_t-\mu}{\sigma_t}\right)\right)$$
$$+ \alpha_{1t}\left(\frac{Q_t-\mu}{\sigma_t}\right)^{\alpha_{1t}+\alpha_{2t}} \Bigg[2\alpha_{2t}\bigg(-1 + \alpha_{2t}\log\left(\frac{Q_t-\mu}{\sigma_t}\right)\bigg)$$
$$- \left(\frac{Q_t-\mu}{\sigma_t}\right)^{\alpha_{2t}} \cdot \left(\alpha_{1t} - 2\alpha_{2t} + \alpha_{2t}(-\alpha_{1t}+\alpha_{2t})\log\left(\frac{Q_t-\mu}{\sigma_t}\right)\right)\Bigg]\Bigg\} \frac{\partial \alpha_{2t}}{\partial \Phi_i},$$

$$\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial \mu \partial \Psi_i} = \frac{{\sigma_t}^{1+\alpha_{1t}+\alpha_{2t}}(Q_t-\mu)^{-1-\alpha_{1t}-\alpha_{2t}}}{\left(\alpha_{2t}\sigma_t\left(\frac{Q_t-\mu}{\sigma_t}\right)^{\alpha_{1t}} + \alpha_{1t}\sigma_t\left(\frac{Q_t-\mu}{\sigma_t}\right)^{\alpha_{2t}}\right)^2} \Bigg[-\alpha_{2t}^4 \left(\frac{Q_t-\mu}{\sigma_t}\right)^{3\alpha_{1t}}$$
$$- \alpha_{1t}^4 \left(\frac{Q_t-\mu}{\sigma_t}\right)^{3\alpha_{2t}} + \alpha_{2t}\bigg(-\alpha_{1t}\alpha_{2t}(\alpha_{1t}+2\alpha_{2t}) + \alpha_{1t}(\alpha_{1t}-\alpha_{2t})^2 \left(\frac{Q_t-\mu}{\sigma_t}\right)^{\alpha_{2t}}\bigg)$$
$$\left(\frac{Q_t-\mu}{\sigma_t}\right)^{2\alpha_{1t}+\alpha_{2t}} - \alpha_{1t}^2\alpha_{2t}(2\alpha_{1t}+\alpha_{2t})\left(\frac{Q_t-\mu}{\sigma_t}\right)^{\alpha_{1t}+2\alpha_{2t}} \Bigg] \frac{\partial \sigma_t}{\partial \Psi_i},$$

$$\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial\Gamma_i\partial\Phi_j} = -\Bigg[\frac{\left(\frac{Q_t-\mu}{\sigma_t}\right)^{\alpha_{1t}+\alpha_{2t}}\left(-1+\alpha_{1t}\log\left(\frac{Q_t-\mu}{\sigma_t}\right)\right)\left(-1+\alpha_{2t}\log\left(\frac{Q_t-\mu}{\sigma_t}\right)\right)}{\left(\alpha_{2t}\left(\frac{Q_t-\mu}{\sigma_t}\right)^{\alpha_{1t}}+\alpha_{1t}\left(\frac{Q_t-\mu}{\sigma_t}\right)^{\alpha_{2t}}\right)^2}\Bigg]\frac{\partial\alpha_{1t}}{\partial\Gamma_i}\frac{\partial\alpha_{2t}}{\partial\Phi_j},$$

$$\begin{aligned}\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial\Gamma_i\partial\Psi_j} = \frac{1}{\sigma_t}\Bigg[&\left(\frac{Q_t-\mu}{\sigma_t}\right)^{-\alpha_{1t}}\left(-1+\alpha_{1t}\log\left(\frac{Q_t-\mu}{\sigma_t}\right)\right)\\ &+\frac{\alpha_{1t}(\alpha_{1t}-\alpha_{2t})\left(\frac{Q_t-\mu}{\sigma_t}\right)^{2\alpha_{2t}}\left(-1+\alpha_{1t}\log\left(\frac{Q_t-\mu}{\sigma_t}\right)\right)}{\left(\alpha_{2t}\left(\frac{Q_t-\mu}{\sigma_t}\right)^{\alpha_{1t}}+\alpha_{1t}\left(\frac{Q_t-\mu}{\sigma_t}\right)^{\alpha_{2t}}\right)^2}\\ &+\frac{\left(\frac{Q_t-\mu}{\sigma_t}\right)^{\alpha_{2t}}\left(2\alpha_{1t}-\alpha_{2t}+\alpha_{1t}(-\alpha_{1t}+\alpha_{2t})\log\left(\frac{Q_t-\mu}{\sigma_t}\right)\right)}{\alpha_{2t}\left(\frac{Q_t-\mu}{\sigma_t}\right)^{\alpha_{1t}}+\alpha_{1t}\left(\frac{Q_t-\mu}{\sigma_t}\right)^{\alpha_{2t}}}\Bigg]\frac{\partial\alpha_{1t}}{\partial\Gamma_i}\frac{\partial\sigma_t}{\partial\Psi_j},\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial\Phi_i\partial\Psi_j} = \frac{1}{\sigma_t}\Bigg[&\left(\frac{Q_t-\mu}{\sigma_t}\right)^{-\alpha_{2t}}\left(-1+\alpha_{2t}\log\left(\frac{Q_t-\mu}{\sigma_t}\right)\right)\\ &+\frac{\alpha_{2t}(\alpha_{2t}-\alpha_{1t})\left(\frac{Q_t-\mu}{\sigma_t}\right)^{2\alpha_{1t}}\left(-1+\alpha_{2t}\log\left(\frac{Q_t-\mu}{\sigma_t}\right)\right)}{\left(\alpha_{1t}\left(\frac{Q_t-\mu}{\sigma_t}\right)^{\alpha_{2t}}+\alpha_{2t}\left(\frac{Q_t-\mu}{\sigma_t}\right)^{\alpha_{1t}}\right)^2}\\ &+\frac{\left(\frac{Q_t-\mu}{\sigma_t}\right)^{\alpha_{1t}}\left(2\alpha_{2t}-\alpha_{1t}+\alpha_{2t}(-\alpha_{2t}+\alpha_{1t})\log\left(\frac{Q_t-\mu}{\sigma_t}\right)\right)}{\alpha_{1t}\left(\frac{Q_t-\mu}{\sigma_t}\right)^{\alpha_{2t}}+\alpha_{2t}\left(\frac{Q_t-\mu}{\sigma_t}\right)^{\alpha_{1t}}}\Bigg]\frac{\partial\alpha_{2t}}{\partial\Phi_i}\frac{\partial\sigma_t}{\partial\Psi_j},\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial\Gamma_i\partial\Gamma_j} = \Bigg[&-\left(\frac{Q_t-\mu}{\sigma_t}\right)^{-\alpha_{1t}}\log\left(\frac{Q_t-\mu}{\sigma_t}\right)^2+\frac{\left(\frac{Q_t-\mu}{\sigma_t}\right)^{\alpha_{2t}}\log\left(\frac{Q_t-\mu}{\sigma_t}\right)\left(-2+\alpha_{1t}\log\left(\frac{Q_t-\mu}{\sigma_t}\right)\right)}{\alpha_{2t}\left(\frac{Q_t-\mu}{\sigma_t}\right)^{\alpha_{1t}}+\alpha_{1t}\left(\frac{Q_t-\mu}{\sigma_t}\right)^{\alpha_{2t}}}\\ &-\frac{\left(\frac{Q_t-\mu}{\sigma_t}\right)^{2\alpha_{2t}}\left(-1+\alpha_{1t}\log\left(\frac{Q_t-\mu}{\sigma_t}\right)\right)^2}{\left(\alpha_{2t}\left(\frac{Q_t-\mu}{\sigma_t}\right)^{\alpha_{1t}}+\alpha_{1t}\left(\frac{Q_t-\mu}{\sigma_t}\right)^{\alpha_{2t}}\right)^2}\Bigg]\frac{\partial^2\alpha_{1t}}{\partial\Gamma_i\partial\Gamma_j},\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial\Phi_i\partial\Phi_j} = \Bigg[&-\left(\frac{Q_t-\mu}{\sigma_t}\right)^{-\alpha_{2t}}\log\left(\frac{Q_t-\mu}{\sigma_t}\right)^2+\frac{\left(\frac{Q_t-\mu}{\sigma_t}\right)^{\alpha_{1t}}\log\left(\frac{Q_t-\mu}{\sigma_t}\right)\left(-2+\alpha_{2t}\log\left(\frac{Q_t-\mu}{\sigma_t}\right)\right)}{\alpha_{1t}\left(\frac{Q_t-\mu}{\sigma_t}\right)^{\alpha_{2t}}+\alpha_{2t}\left(\frac{Q_t-\mu}{\sigma_t}\right)^{\alpha_{1t}}}\\ &-\frac{\left(\frac{Q_t-\mu}{\sigma_t}\right)^{2\alpha_{1t}}\left(-1+\alpha_{2t}\log\left(\frac{Q_t-\mu}{\sigma_t}\right)\right)^2}{\left(\alpha_{1t}\left(\frac{Q_t-\mu}{\sigma_t}\right)^{\alpha_{2t}}+\alpha_{2t}\left(\frac{Q_t-\mu}{\sigma_t}\right)^{\alpha_{1t}}\right)^2}\Bigg]\frac{\partial^2\alpha_{2t}}{\partial\Phi_i\partial\Phi_j},\end{aligned}$$

and

$$\begin{aligned}\frac{\partial^2 l_t(\boldsymbol{\theta})}{\partial\Psi_i\partial\Psi_j} = \frac{1}{{\sigma_t}^2}\Bigg[&\alpha_{2t}\left(-1-(-1+\alpha_{2t})\left(\frac{Q_t-\mu}{\sigma_t}\right)^{-\alpha_{2t}}\right)-\alpha_{1t}(\alpha_{1t}-1)\left(\frac{Q_t-\mu}{\sigma_t}\right)^{-\alpha_{1t}}\\ &+\frac{\alpha_{1t}(-\alpha_{1t}+\alpha_{2t})(1-\alpha_{1t}+\alpha_{2t})\left(\frac{Q_t-\mu}{\sigma_t}\right)^{\alpha_{2t}}}{\alpha_{2t}\left(\frac{Q_t-\mu}{\sigma_t}\right)^{\alpha_{1t}}+\alpha_{1t}\left(\frac{Q_t-\mu}{\sigma_t}\right)^{\alpha_{2t}}}-\frac{\alpha_{1t}^2(\alpha_{1t}-\alpha_{2t})^2\left(\frac{Q_t-\mu}{\sigma_t}\right)^{2\alpha_{2t}}}{\left(\alpha_{2t}\left(\frac{Q_t-\mu}{\sigma_t}\right)^{\alpha_{1t}}+\alpha_{1t}\left(\frac{Q_t-\mu}{\sigma_t}\right)^{\alpha_{2t}}\right)^2}\Bigg]\\ &\frac{\partial\sigma_t}{\partial\Psi_i}\frac{\partial\sigma_t}{\partial\Psi_j}.\end{aligned}$$

S5 Simulation Study

S5.1 Performance of the cMLE under X_{it} in the max domain of attraction

In this section, we conduct one more simulation to demonstrate the estimation to the setting where X_{it} is in the max domain of attraction but not exact Fréchet. We simulate random variables $X_{1,i,t}$ and $X_{2,i,t}$ $(i = 1, 2, \cdots, p)$ and their distributions are both in the Fréchet domain of attraction. Let $Q_{1t} = \max_{1\le i\le p} X_{1,i,t}$, $Q_{2t} = \max_{1\le i\le p} X_{2,i,t}$, and $Q_t = \max(Q_{1t}, Q_{2t}) = \max(\max_{1\le i\le p} X_{1,i,t}, \max_{1\le i\le p} X_{2,i,t})$. For a sufficiently large p, the approximated distribution of Q_{jt} is Fréchet distribution with location parameter μ, shape parameter σ_t and scale parameter α_{jt}, for $j = 1, 2$.

Recent studies by Vernic (2006), Eling (2012) and Bernardi et al. (2012) have shown that financial data sets can be reasonably fitted by skewed distributions such as the skew-t distribution (Azzalini and Capitanio (2003)). Specifically, we simulate $X_{j,i,t}$ $(j = 1, 2)$ from the following model,

$$X_{j,i,t} = \frac{\sigma_t}{a_{j,p,t}} Z_{j,i,t} + \mu, \quad i = 1, \cdots, p,$$

where $Z_{j,i,t}$ are i.i.d. random variables (across i, j, and t) generated from the skew-t distribution $\mathrm{ST}(\beta, \nu_{jt})$ with degree of freedom ν_{jt} and the skew parameter β. Note that for a skew-t distribution with degree of freedom ν, the extreme value index is $1/\nu$ and the tail index is ν. We draw $\beta \sim U[1, 2]$, and set $\nu_{jt} = \alpha_{jt}$. The dynamics of σ_t and α_{jt} $(j = 1, 2)$ are modeled in the same way as in equations (2.2)-(2.4) of the main text. According to the norming constant for the limiting distribution of the sample maximum for the skew-t distribution (Peng et al. (2016)), we set $a_{j,p,t} = \left(\frac{2p\Gamma\left(\frac{\nu_{jt}+1}{2}\right)}{\Gamma\left(\frac{\nu_{jt}}{2}\right)\sqrt{\nu_{jt}\pi}} \nu_{jt}^{\frac{\nu_{jt}-1}{2}} T_{\nu_{jt}+1}(\beta\sqrt{\nu_{jt}+1}) \right)^{\frac{1}{\nu_{jt}}}$, where $T_{\nu_{jt}+1}(\cdot)$ is the cdf of the standard Student t distribution with degree of freedom $\nu_{jt}+1$.

We set $p=1000$, and simulate $Q_t=\max(Q_{1t},Q_{2t})=\max\left(\max_{1\le i\le p}X_{1,i,t},\max_{1\le i\le p}X_{2,i,t}\right)$. We generate data with the following parameters $(\beta_0,\beta_1,\beta_2,\beta_3,\gamma_0,\gamma_1,\gamma_2,\gamma_3,\delta_0,\delta_1,\delta_2,\delta_3,\mu)^T=(-0.244,0.787,0.066,8.111,0.230,0.755,0.417,7.114,-0.035,0.907,0.425,4.861,-0.227)^T$, which is the same as in Section 4 of the main text. We investigate the performance of cMLE with sample sizes $N=1000,2000,5000$, and 10000. For each sample size, we conduct 100 experiments. The parameter estimation results are presented in Table A1, including the average of the estimates and the standard deviation from the 100 experiments.

In Table A1, except γ_2, β_2, and μ, the estimated values of all other parameters are very close to their counterparts in Table 1 obtained from the simulation using the exact Fréchet distribution in Section 4. Even for γ_2, β_2, and μ, the estimated values are reasonably good and acceptable. In the estimation of advanced nonlinear time series with parameter dynamics, it is common that the estimation of parameters can be challenging when the data generating processes are not exactly the same as the assumed one. These observations suggest that our estimated models in real data analysis are meaningful and reliable.

S5.2 Comparison with the autoregressive conditional Fréchet model

In this section, we compare our AcAF model with the autoregressive conditional Fréchet (AcF) model. The AcF model is one of the time-varying GEV models that can be used to model time series data of maxima. The AcF model converts $\max(Y_{1t}^{1/\alpha_{1t}},Y_{2t}^{1/\alpha_{2t}})$ in equation (2.1) of the main text to Y_t^{1/α_t} and contains the dynamic structures for σ_t and α_t same to equations (2.2) and (2.3) of the main text. More details can be found in Zhao et al. (2018). The way to fulfill the comparison is as follows. First, we simulate data $\{Q_t\}$ with a length of 1000 from the AcAF

Table A1: Numerical results for the performance of the cMLE under X_{it} in the max domain of attraction with sample sizes 1000, 2000, 5000, and 10000. Mean and S.D. are the sample mean and standard deviation of the cMLE's obtained from 100 simulations.

		$N = 1000$		$N = 2000$		$N = 5000$		$N = 10000$	
Parameter	True value	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
γ_0	0.230	0.291	0.195	0.321	0.178	0.419	0.124	0.442	0.105
γ_1	0.755	0.762	0.156	0.756	0.133	0.693	0.095	0.688	0.077
γ_2	0.417	0.207	0.207	0.166	0.140	0.162	0.097	0.166	0.067
γ_3	7.114	8.380	5.403	9.178	5.526	8.066	4.340	9.460	4.120
δ_0	−0.035	−0.002	0.029	−0.002	0.019	−0.004	0.012	−0.005	0.009
δ_1	0.907	0.892	0.041	0.892	0.024	0.893	0.015	0.894	0.011
δ_2	0.425	0.462	0.137	0.461	0.090	0.447	0.049	0.442	0.036
δ_3	4.861	5.722	2.124	5.472	1.281	5.189	0.747	5.018	0.633
β_0	−0.244	−0.233	0.205	−0.191	0.163	−0.115	0.058	−0.098	0.036
β_1	0.787	0.537	0.369	0.608	0.303	0.747	0.124	0.771	0.084
β_2	0.066	−0.039	0.048	−0.042	0.024	−0.040	0.013	−0.039	0.009
β_3	8.111	7.271	4.730	7.424	4.451	8.774	3.915	9.593	2.839
μ	−0.227	−0.464	0.050	−0.473	0.038	−0.488	0.019	−0.507	0.032

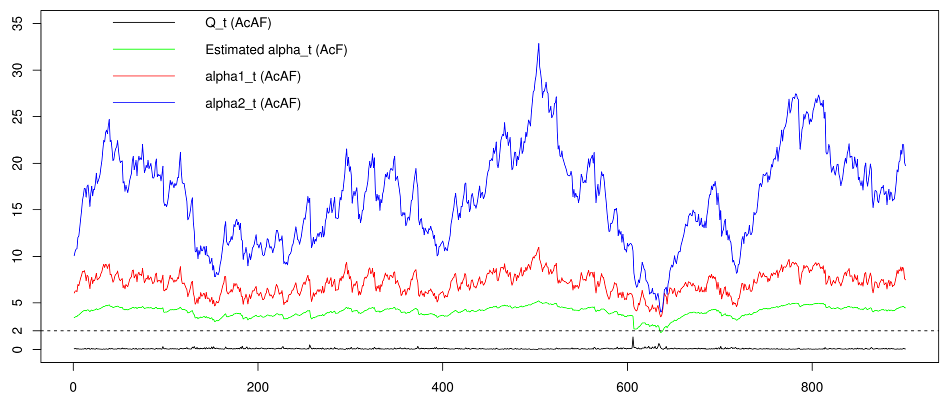


Figure 1: Estimated tail index $\{\hat{\alpha}_t\}$ (green) by the AcF model, recovered tail indices $\{\alpha_{1t}\}$ (red), $\{\alpha_{2t}\}$ (blue) by the AcAF model and simulated $\{Q_t\}$ (black) by the AcAF model. All plotted series omit the first 100 data points.

model by the parameters mentioned in Section 4, and use the simulated data and parameters to recover two tail indices $\{\alpha_{1t}\}$ and $\{\alpha_{2t}\}$ through the evolution structures (2.3) and (2.4) in the main text. Then we fit the AcF model on the simulated data to estimate its tail index $\hat{\alpha}_t$. The simulated $\{Q_t\}$, recovered tail indices $\{\alpha_{1t}\}$ and $\{\alpha_{2t}\}$ by the AcAF model and the fitted tail index $\{\hat{\alpha}_t\}$ by the AcF model are presented in Figure 1.

From Figure 1, we can see that the AcF model seems to give a significantly lower estimation of the tail index than the AcAF model. An under-estimated tail index implies over-estimated tail risk, which in turn may result in higher reserve requirements and other expenses for financial institutions, and in turn lead to reduced liquidity of financial institutions. If the potential loss (risk) that a financial system faces based on the known market information has multiple sources, that is, the loss data is not i.i.d., the tail index estimated by the AcF model is inaccurate and very different from the real tail risk. The recovered tail indices $\{\alpha_{1t}\}$ and $\{\alpha_{2t}\}$ by the AcAF model are all larger than 2, hence the conditional mean and variance of the simulated maxima

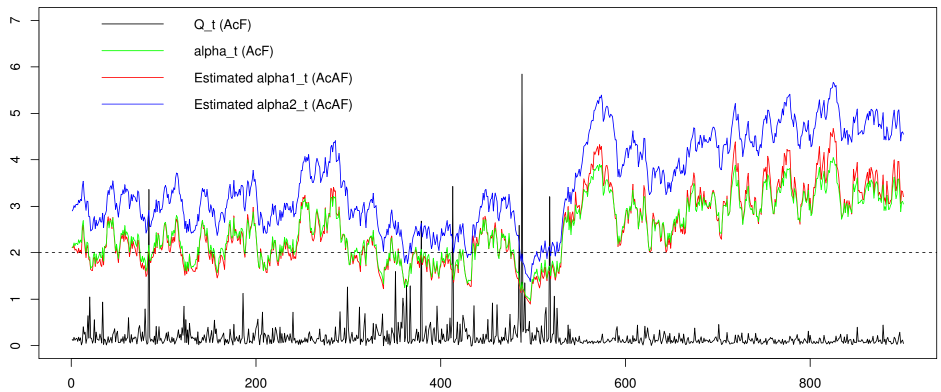


Figure 2: Recovered tail index $\{\alpha_t\}$ (green) by the AcF model, estimated tail indices $\{\hat{\alpha}_{1t}\}$ (red), $\{\hat{\alpha}_{2t}\}$ (blue) by the AcAF model and simulated $\{Q_t\}$ (black) by the AcF model. All plotted series omit the first 100 data.

of maxima $\{Q_t\}$ from the AcAF model always exist. We note that the estimated tail index $\{\hat{\alpha}_t\}$ by the AcF model is less than 2 at some points, which means that the variance does not exist. This phenomenon indicates that when $\{Q_t\}$ follows an accelerated Fréchet distribution with finite first and second moments, fitting observed Q_t series to the AcF model may lead to a wrong conclusion of the second (or even the first) moment being infinite, which raises a question in practice, how to test the moment conditions in time series. We will study this issue in a further project.

We also simulate $\{Q_t\}$ with a length of 1000 by the AcF model with the parameters provided in Zhao et al. (2018), then apply the AcAF model to estimate parameters. Similarly, we recover the tail index $\{\alpha_t\}$ of the simulated data and plot it with the estimated tail indices $\{\hat{\alpha}_{1t}\}$ and $\{\hat{\alpha}_{2t}\}$ of the AcAF model in Figure 2. As can be seen from Figure 2, the tail index recovered by the AcF model almost coincides with $\{\hat{\alpha}_{1t}\}$ estimated by the AcAF model, i.e., the green line in Figure 2 coincides with the red line. In the AcAF model, tail risk is mainly characterized by the dominant sequence and the $\{\hat{\alpha}_{1t}\}$ series captures the dominant information, which indicates that an AcAF model fitting is acceptable for the data generated from an AcF model. And the conditional mean and variance of the maxima of maxima $\{Q_t\}$ do not always exist.

In summary, using the patterns in Figures 1 and 2, we can visually conclude which model, AcF or AcAF, to be used in real data analysis and make inferences.

S5.3 Convergence of maxima of maxima in factor model

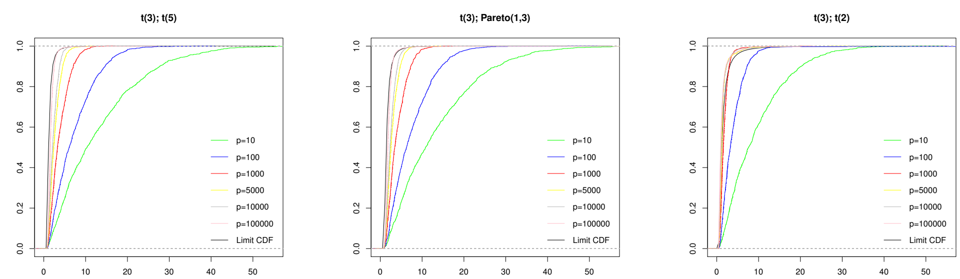
In this section, we conduct numerical experiments to investigate the finite sample behavior of Q_t described in Corollary 1 of the main text. Specifically, we study the convergence of the marginal distribution of Q_t to its accelerated Fréchet limit under a one-time period factor model. To simplify notation,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

$$X_i = \beta_i Z + \sigma_i \max(\epsilon_{1i}, \epsilon_{2i}), \quad i = 1, \cdots, p,$$

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 σ_i's are i.i.d. random variables generated from $\frac{1}{2}U(0,0.09) + \frac{1}{2}U(0.01,0.08)$ such that all σ_i's are moderate in $(0.005, 0.085)$. This setting roughly matches the pattern of GARCH(1,1) fitted average volatilities of 505 different stocks in S&P 500.

Random variables ϵ_{1i}'s, ϵ_{2i}'s are independent and their distributions are in the Fréchet domain of attraction. We will select different combinations of ϵ_{1i} and ϵ_{2i} according to the relationship between two tail indices. Specifically, we will choose the following three cases to perform our simulation study: (i) $t(3)$ and $t(5)$; (ii) $t(3)$ and Pareto$(1,3)$; (iii) $t(3)$ and $t(2)$. Here $t(\nu)$ represents t-distribution with degree of freedom ν, and Pareto(x_m, α) represents Pareto distribution with c.d.f. $F(x) = 1 - (x_m/x)^\alpha$ for $x \geq x_m$. This setting corresponds to the three cases in Corollary 1.

We set $Q = \max_{1\leq i\leq p} X_i$ and compare the finite sample empirical distribution of Q and its corresponding limit stated in Corollary 1 under different ϵ_{1i}, ϵ_{2i} and p. For each $(\epsilon_{1i}, \epsilon_{2i}, p)^T$ combination, 1000 sets of i.i.d. $\{X_i\}_{i=1}^p$ are generated, resulting in 1000 sampled $Q = \max_{1\leq i\leq p} X_i$. Figure 3 plots the empirical c.d.f. of the normalized Q in Corollary 1 along with the corresponding limiting accelerated Fréchet distribution. It can be clearly see in Figure 3 that as p increases, the empirical distribution of Q approaches its accelerated Fréchet limit. A larger tail index requires a larger p for accurate approximation.



(a) (b) (c)

Figure 3: Finite sample empirical distribution of the maxima of maxima Q and its corresponding accelerated Fréchet limit, with different combinations of p and distributions of $(\epsilon_{1i}, \epsilon_{2i})^T$ in the factor model.

Bibliography

Azzalini, A. and Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(2):367–389.

Bernardi, M., Maruotti, A., and Petrella, L. (2012). Skew mixture models for loss distributions: a bayesian approach. *Insurance: Mathematics and Economics*, 51(3):617–623.

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

Birkhoff, G. D. (1931). Proof of the ergodic theorem. *Proceedings of the National Academy of Sciences*, 17(12):656–660.

Cao, W. and Zhang, Z. (2021). New extreme value theory for maxima of maxima. *Statistical Theory and Related Fields*, 5(3):232–252.

Chan, K.-S. and Tong, H. (1994). A note on noisy chaos. *Journal of the Royal Statistical Society: Series B (Methodological)*, 56(2):301–311.

Dombry, C. et al. (2015). Existence and consistency of the maximum likelihood estimators for the extreme value index within the block maxima framework. *Bernoulli*, 21(1):420–436.

Eling, M. (2012). Fitting insurance claims to skewed distributions: Are the skew-normal and skew-student good models? *Insurance: Mathematics and Economics*, 51(2):239–248.

Francq, C., Zakoian, J.-M., et al. (2004). Maximum likelihood estimation of pure garch and arma-garch processes. *Bernoulli*, 10(4):605–637.

Mäkeläinen, T., Schmidt, K., and Styan, G. P. (1981). On the existence and uniqueness of the maximum likelihood estimate of a vector-valued parameter in fixed-size samples. *The Annals of Statistics*, 9(4):758–767.

Peng, Z., Li, C., and Nadarajah, S. (2016). Extremal properties of the skew-t distribution. *Statistics & Probability Letters*, 112:10–19.

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

Vernic, R. (2006). Multivariate skew-normal distributions with applications in insurance. *Insurance: Mathematics and Economics*, 38(2):413–426.

Zhao, Z., Zhang, Z., and Chen, R. (2018). Modeling maxima with autoregressive conditional Fréchet model. *Journal of Econometrics*, 207(2):325–351.