

Dynamic Modeling via Autoregressive Conditional GB2 for Cross-Sectional Maxima of Financial Time Series Data

Ning Fan, Chunming Zhang & Zhengjun Zhang

To cite this article: Ning Fan, Chunming Zhang & Zhengjun Zhang (2025) Dynamic Modeling via Autoregressive Conditional GB2 for Cross-Sectional Maxima of Financial Time Series Data, Journal of Business & Economic Statistics, 43:4, 968-978, DOI: [10.1080/07350015.2025.2450488](https://doi.org/10.1080/07350015.2025.2450488)

To link to this article: <https://doi.org/10.1080/07350015.2025.2450488>



View supplementary material [↗](#)



Published online: 26 Feb 2025.



Submit your article to this journal [↗](#)



Article views: 331



View related articles [↗](#)



View Crossmark data [↗](#)



Dynamic Modeling via Autoregressive Conditional GB2 for Cross-Sectional Maxima of Financial Time Series Data

Ning Fan^a, Chunming Zhang^a , and Zhengjun Zhang^b 

^aDepartment of Statistics, University of Wisconsin-Madison, Madison, WI; ^bSchool of Economics and Management, and MOE Social Science Laboratory of Digital Economic Forecasts and Policy Simulation, University of Chinese Academy of Sciences; AMSS Center for Forecasting Sciences, Chinese Academy of Sciences, Beijing, China

ABSTRACT

This article introduces the “autoregressive conditional generalized beta distribution of the second kind” (AcGB2) to model the dynamic cross-sectional maxima of multivariate financial times series. The temporal dependence of the resulting univariate maxima time series is characterized by the parameter dynamics of the standard GB2 distribution, which offer versatility in approximating various distributions, including heavy-tailed distributions, through parameter adjustments. Consequently, the newly proposed AcGB2 modeling enhances flexibility in fitting real data, especially in scenarios where extreme value theory conditions are not met, as demonstrated through simulations. Real data analysis is conducted on three datasets: two with medium-sized cross-sectional dimensions (30 or fewer) and one with high dimension (100 or higher). These datasets are based on the daily negative simple-returns of 30 stocks in the Dow Jones Industrial Average, stocks in S&P 100, and stocks from 22 primary dealers, respectively. The article establishes stationary and ergodic solutions for this new time series model under mild parameter conditions and derives the consistency, asymptotic normality, and uniqueness of the statistical conditional maximum likelihood estimators.

ARTICLE HISTORY

Received January 2024
Accepted December 2024

KEYWORDS

Computational finance;
Extreme values; Nonlinear
time series; Parameter
dynamic; Systemic risk

1. Introduction

In economic and financial studies, indicators like GDP, CPI, PPI, and market indexes such as S&P 100, S&P 500, and Dow Jones 30 hold a central position. These metrics represent weighted averages of their respective constituents. Similarly, in climate studies, metrics like averages (means), medians, minimums, and maximums of precipitations, temperatures, air pollutants, etc., in a region are of significant concern to administrators and residents. Analogous scenarios are prevalent across various research domains. Univariate time series models have extensively explored each of these indicators and indexes. This article focuses on the dynamic index representing the worst-case scenario in a financial system, specifically by capturing the maxima of negative daily simple returns within a vector constituting the system (e.g., Dow Jones 30). Our proposed univariate time series model holds potential applications beyond financial systems, including climate studies and various other domains.

In the literature, an extensive examination of extreme events in financial markets, insurance, environmental studies, and various domains has heavily relied on classical extreme value theory and its statistical inferences. This theory finds particular relevance in datasets with sufficient length and when conditions align with the extreme value theorem. Numerous applications and proposed models can be found in the literature; for

instance, Diebold, Schuermann, and Stroughair (1998) deal with independent identically distributed series, Chavez-Demoulin, Embrechts, and Sardy (2014) use a nonparametric framework, Kelly and Jiang (2014) propose a panel estimation approach, Zhang and Schwaab (2017) study the dynamics of tail shape for the generalized Pareto distribution (GPD), Bali and Weinbaum (2007) work on high frequency returns. Additionally, Zhao, Zhang, and Chen (2018), Mao and Zhang (2018), Ji and Li (2021), and Lin and Zhang (2022) develop the autoregressive conditional Fréchet (AcF; see (9)) model to depict the evolving behavior of maxima in high-dimensional financial time series, amongst many others. Further, the discussion paper (Zhang 2021a, 2021b) explores current practices as the cross-sectional dimension, that is, the number of components, approaches infinity. Our article concentrates on examining time series characterized by worst-case scenarios within a fixed, medium-sized cross-sectional component of observations. This delineates that these worst-case scenarios are not inherently observed within the context of the time series under study. These ongoing efforts in statistical modeling and methodological development are driven by the pursuit of understanding extreme financial risks, such as stock market crashes, financial and economic crises, systematic and systemic risks, and effective worst-case scenario risk management. Consequently, studying extreme

return patterns and their impacts has become a crucial first step in achieving the broader goal of preventing market crashes and related financial disruptions. This article aims to advance both the existing theoretical framework and practical approaches in this field.

While it is intuitive to model maxima using extreme value theory and related statistical models, numerous application scenarios exist where the extreme value theory falls short. This limitation emerges due to various constraints, including insufficient sample sizes, restricted cross-sectional dimensions in multivariate time series, unknown dependencies between random variables. Consequently, there is a pressing need for alternative models capable of effectively capturing maxima characterized by high skewness and long tails. This article aims to develop such alternative models.

The standard “generalized beta distribution of the second kind” (GB2) has drawn much attention (McDonald and Xu 1995) due to its adeptness in describing highly skewed and long-tailed data. Specifically, the GB2(a, b, p, q) distribution family, composed of one scale parameter (b) and three shape parameters (a controlling both tails, p controlling the left tail, and q controlling the right tail), offers versatility and richness. This flexibility allows the distribution to adopt various shapes, including J-shaped, bell-shaped, long-tailed, light-tailed, right-skewed, and left-skewed distributions. Notably, the GB2 distribution family encompasses extreme value distributions as special or limiting cases, making it suitable for handling long-tailed and highly skewed data. Furthermore, compared to extreme value distributions, the GB2 distribution family exhibits greater versatility. Consequently, when handling datasets lacking significant skewness, the GB2 model, with its flexibility accommodating a broad spectrum of data shapes, is anticipated to outperform extreme value distributions. However, while there is potential to further generalize the GB2 family for increased modeling adaptability, such extensions may introduce greater inference complexity, interpretational challenges, and the risk of overfitting in non-linear time series models.

Considering the preceding discussions and research, we propose a novel framework called “dynamic GB2” to model the time-varying behavior of cross-sectional maxima within financial time series. By allowing the parameters of the GB2 distribution to vary over time, this model can effectively capture the time-varying patterns inherent in the data. Essentially, the GB2-based time series model accurately represents the highly-skewed and long-tailed characteristics of financial data (maxima) while ensuring enhanced flexibility. This article introduces the autoregressive conditional GB2 model, referred to as the AcGB2 model, which shares the same dynamic parameter structure as the AcF model. Similar to the AcF model, the AcGB2 model exhibits broad applicability, extending its utility beyond financial contexts. Its application spans the analysis of systemic risks across diverse domains, including financial institutions, extreme climate conditions, seismic activities, forest fires, and more. Next, we outline some additional features offered by the AcGB2 model in financial econometric time series modeling and inference.

Compared to the AcF model, the AcGB2 model offers the following advantages: (i) Enhanced Flexibility: The AcGB2 model demonstrates greater flexibility, particularly when

the number of assets is medium-sized, benefiting from the fact that Fréchet-type distributions are limit cases of GB2 distributions. (ii) Absence of Constraints: The AcGB2 model avoids constraints related to block size choices and threshold selections, challenges commonly encountered in extreme value applications. (iii) Improved Market Description: Demonstrated superiority in describing market features, as evidenced in the real data section, including Table 9, compared to the AcF model. (iv) Enhanced Prediction: Our analysis reveals that AcF models tend to over-predict the worst market scenarios of Dow Jones 30 while under-predicting those of S&P 100 and primary dealers during critical periods. Additionally, we propose leveraging exponential transformations of the maxima of observations to broaden the applicability of both AcF and AcGB2 models across diverse data scenarios, including those with normally distributed observations. With this transformation, our newly proposed model can be viewed as capturing the dynamics of the worst simple-returns using dynamic parameters. Notably, it stands as the first model, besides the AcF models commonly applied to negative log-returns.

The rest of the article is structured as follows: Section 2 introduces the dynamic “AcGB2 model” in (4), (5), and (8) and explores stationarity and ergodicity. Section 3 develops the parameter estimation method. Section 4 conducts a simulation study to analyze the effect of sample size on estimation accuracy, demonstrating the flexibility of the AcGB2 model through comparisons with other models. This section also discusses two non-time-varying parameters of the GB2 distribution and evaluates the impact of the number of time series on fitting the GB2 distribution. Section 5 focuses on the real data application, employing stock returns from companies in the Dow Jones Industrial Average (DJIA), companies in S&P 100, and 22 primary dealers, while assessing goodness of fit. Section 6 provides the conclusion. In the online supplementary file, technical derivations are presented in Appendix A, additional details in Appendix B, scale parameter and tail index dynamics for the real data in Appendix C, with tables and figures included in Appendix D.

2. Autoregressive Conditional GB2 (AcGB2) Model

2.1. Connecting GB2 Distributions to Extreme Value Distributions

In the Introduction, we discussed some basic properties of GB2 distributions and their link to extreme value distributions. Due to the limited application of GB2 in finance, we outline the most basic mathematical properties of standard GB2 distributions based on McDonald and Xu (1995) and Yang, Frees, and Zhang (2011) here.

Consider a random variable $X \sim \text{GB2}(a, b, p, q)$, whose probability density function (p.d.f.) is given by:

$$f_X(x; a, b, p, q) = \frac{|a|}{B(p, q)x} (x/b)^{ap} \{1 + (x/b)^a\}^{-(p+q)}, \quad x > 0. \quad (1)$$

Here, b is the scale parameter, $|a| > 0$, $p > 0$, and $q > 0$ are three shape parameters, and $B(p, q)$ represents the beta function with parameters p and q . Parameter a controls both tails, p controls

the left tail, and q controls the right tail. When $Y = \log(X)$, it follows an EGB2($\sigma = 1/a, b, p, q$) distribution.

Observing the limit properties of the GB2 distribution in Appendix A.7, we can see that the density of X is heavy-tailed because it varies regularly at ∞ with an index of $-aq - 1$ and varies regularly at 0 with an index of $ap - 1$. We observe the following limit property:

$$\lim_{q \rightarrow \infty} f_X(x; a, q^{1/a} \beta, p, q) = |a| x^{ap-1} e^{-(x/\beta)^a} / \{\beta^{ap} \Gamma(p)\}. \quad (2)$$

The right-hand side of (2) contains the Fréchet-type density function as a special case with specific parameter choices, which provides mathematical motivation for our proposed models in Section 2.2.

The standard GB2 distribution has been widely applied in various fields. McDonald and Xu (1995) explored the properties and applications of the generalized beta distribution. Corrado (2001) used the GB2 distribution in option pricing, while Kleiber and Kotz (2003) employed it in actuarial science. Frees and Valdez (2008) used the four-parameter GB2 distribution for insurance claims. Papalexiou and Koutsoyiannis (2012) demonstrated the suitability of the GB2 distribution for global rainfall data. Caivano and Harvey (2014) focused on times series models using the exponential GB2 (EGB2) conditional distribution. Chiang and Li (2015) modeled asset returns incorporating skewness, kurtosis, and outliers using the EGB2 distribution. Additionally, Chen and Singh (2017) showcased its fitting capacity for flood frequency analysis. Higbee and McDonald (2021) compared the GB2 and skewed generalized log- t distributions with finance, amongst other applications. In the multivariate context, Yang, Frees, and Zhang (2011) extended the GB2 distribution to the MGB2 copula for multivariate data, applied to pricing synthetic CDO (Cui and Ma 2014) and modeling guaranteed unitized participating life insurance (Zheng et al. 2019). One primary reason for the attention garnered by GB2 is its association with various distributions, including extreme value distributions, especially in modeling long-tailed and highly skewed data. These distributions are often treated as limits or special cases of GB2. For instance, fixing the second shape parameter p to 1 yields the Burr type 12 distribution, while allowing the third shape parameter q to approach infinity results in the generalized gamma distribution. McDonald and Xu (1995) summarized the relationships among GB2 family distributions using distribution trees, encompassing common distributions such as gamma, generalized gamma, Weibull, Chi-square, log-normal, log-logistic, F, exponential, Dagum, Lomax, Fisk, Burr type 3, and Burr type 12. Most recently, Lin, Liu, and Zhang (2023) introduced a novel EGB2 option pricing formula that outperforms the Black-Scholes formula and other jump diffusion models.

2.2. The Proposed AcGB2 Model

The data we analyze comprises a time series of maxima, denoted as M_t :

$$M_t = \max_{1 \leq i \leq m} X_{i,t}, \quad 1 \leq t \leq n, \quad (3)$$

where $\{X_{i,t}\}_{i=1}^m$ represent m (fixed) time series. Here, M_t reflects the worst-case scenario in a financial system (e.g., market, institution, or portfolio) at time t . It is linked to systemic risk. These

$X_{i,t}$ values are related to the negative simple-return of the i th stock at time t . Conventional time series models that focus solely on mean or volatility, such as ARMA, ARIMA, ARCH, and GARCH (Fan and Yao 2003), cannot either fit $\{M_t\}_{t=1}^n$ well or offer sufficient economic explanations (Teyssière and Kirman, 2007 and Rootzen and Finkenstadt, 2003). To model the maxima time series $\{M_t\}_{t=1}^n$ when m is fixed or its growth violates the extreme value theorem, we propose a novel dynamic “autoregressive conditional generalized beta distribution of second kind” (AcGB2) framework:

$$\begin{aligned} M_t &= \mu + B_t Y_t^{1/A_t}, \quad 1 \leq t \leq n, \\ \{Y_t\}_{t=1}^n &\text{ are iid variables following the prime beta}(p, q) \\ &\text{distribution, equivalent to GB2}(a = 1, b = 1, p, q) \text{ as} \\ &\text{illustrated in (1)}, \end{aligned} \quad (4)$$

where μ is a location parameter, A_t controls both tails as a shape parameter, B_t stands as a scale parameter.

In (2), a can be either negative or positive. In extreme value theory and its applications, the tail (shape) index (parameter) ξ of a generalized extreme value distribution (see the formula after Remark 8) has three cases: negative, zero, and positive. Most financial applications set ξ and its time-varying tail indexes to be positive. In this work, we also set A_t to be positive, with a dynamic structure similar to that of AcF. When A_t s are a constant, it is related to $1/\xi$ in the literature.

Remark 1. If Y_t follows GB2($a = 1, b = 1, p, q$) as shown in Appendix A.6 part (ii), it is demonstrated that $M_t - \mu = B_t Y_t^{1/A_t}$ follows GB2(A_t, B_t, p, q), where A_t and B_t depend on the history up to time $t - 1$. In model (4), the parameters μ , p , and q are constants, remaining unchanged over time. Here, μ signifies the market's overall best performance relative to negative simple-returns. If $\mu > 1$, it implies that at least one stock's price among m stocks drops daily compared to the prior day's price. The scale parameter B_t can be interpreted as the market volatility, while the shape parameter A_t may relate to the volatility of market volatilities. Parameters p and q , hidden in (4), can be set as arbitrary constants (e.g., $p = 1, q = 1$) similar to many existing models like standard normal, t , and unit Fréchet distributions. In our study, p and q are unknown parameters, enhancing the model's flexibility.

After defining the model in (4), we then specify the stochastic dynamic structures for A_t and B_t . Existing literature presents several well-studied dynamic structures, including quadratic forms in ARCH models by Engle (1982), GARCH models by Bollerslev (1986), logarithmic and exponential forms in Lundbergh, Terasvirta, and Van Dijk (2003), Hall, Swamy, and Tavlás (2016), Zhao, Zhang, and Chen (2018), and their extensions discussed in the Introduction. Following Zhao, Zhang, and Chen (2018) and assuming $A_t > 0$, we adopt the following forms:

$$\begin{aligned} \log(A_t) &= \alpha_0 + \alpha_1 \log(A_{t-1}) + \alpha_2 \exp(-\alpha_3 M_{t-1}), \\ \log(B_t) &= \beta_0 + \beta_1 \log(B_{t-1}) - \beta_2 \exp(-\beta_3 M_{t-1}), \end{aligned} \quad (5)$$

which can be further expressed as

$$\begin{aligned} \log(A_t) &= \alpha_0 + \alpha_1 \log(A_{t-1}) \\ &\quad + \alpha_2 \exp\{-\alpha_3 \cdot (\mu + B_{t-1} Y_{t-1}^{1/A_{t-1}})\}, \end{aligned}$$

$$\log(B_t) = \beta_0 + \beta_1 \log(B_{t-1}) - \beta_2 \exp\{-\beta_3 \cdot (\mu + B_{t-1} Y_{t-1}^{1/A_{t-1}})\}. \quad (6)$$

The parameters in these expressions include $\alpha_0 \in \mathbb{R}$, $\beta_0 \in \mathbb{R}$, $\mu \in \mathbb{R}$, $0 \leq \alpha_1 < 1$, $0 \leq \beta_1 < 1$, $\alpha_1 \neq \beta_1$, $\alpha_2 > 0$, $\beta_2 > 0$, $\alpha_3 > 0$, and $\beta_3 > 0$. For notational simplicity, we define the parameter vector $\theta = (\mu, \alpha_0, \alpha_1, \alpha_2, \alpha_3, \beta_0, \beta_1, \beta_2, \beta_3, p, q)^\top$ within the parameter set

$$\Theta = \{\theta : \alpha_0 \in \mathbb{R}, \beta_0 \in \mathbb{R}, \mu \in \mathbb{R}; 0 \leq \alpha_1 < 1, 0 \leq \beta_1 < 1, \alpha_1 \neq \beta_1; \alpha_2 > 0, \alpha_3 > 0, \beta_2 > 0, \beta_3 > 0; p > 0, q > 0\}. \quad (7)$$

Remark 2. The conditions $0 \leq \alpha_1 < 1$ and $0 \leq \beta_1 < 1$ are technical assumptions necessary to prove the stationarity and ergodicity of the model. Additionally, the condition $\alpha_1 \neq \beta_1$ is essential for parameter identifiability. Setting $\alpha_3 > 0$ and $\beta_3 > 0$ helps avoid issues related to computing overflow values of $\exp(-\alpha_3 M_{t-1})$ and $\exp(-\beta_3 M_{t-1})$ for extremely large M_{t-1} values. Considering the tendency for large maxima to cluster around specific periods, that is, clustering, a substantial M_{t-1} typically precedes a large M_t . This implies that $\log(A_t)$ decreases with M_{t-1} , while $\log(B_t)$ increases with M_{t-1} , as represented by $\alpha_2 > 0$ and $\beta_2 > 0$.

With the established model, the k th moment of $M_t - \mu$ exists only when $-A_t p < k < A_t q$:

$$E\{(M_t - \mu)^k | \mathcal{F}_{t-1}\} = \frac{B_t^k \Gamma(p + k/A_t) \Gamma(q - k/A_t)}{\Gamma(p) \Gamma(q)}.$$

It is important to note that the moment condition $-A_t p < k < A_t q$ introduces technical complexities compared to the proofs developed in the AcF models.

The conditional probability density function (p.d.f.) of M_t takes the form:

$$f_{M_t}(m_t | \mathcal{F}_{t-1}) = \frac{1}{B(p, q)} \frac{A_t (m_t - \mu)^{A_t p - 1}}{B_t^{A_t p} \{1 + (\frac{m_t - \mu}{B_t})^{A_t}\}^{p+q}}, \quad \text{for } m_t > \mu, \quad (8)$$

where $\mathcal{F}_{t-1}$ denotes the sigma-field generated by $\{M_s : s \leq t - 1\}$. Using this model and notations, we will explore the stationary and ergodic properties and conditions in Section 2.3.

Remark 3. Note that while we adopt similar parameter dynamics as Zhao, Zhang, and Chen (2018), there exist significant differences between AcGB2 models and AcF models: (i) AcGB2 focuses on fixed medium-sized multivariate time series, whereas AcF deals with a growing numbers of high-dimensional time series; (ii) the moment existence condition requires AcGB2 models to have additional dynamically satisfied parameter constraints to apply the strong law of large numbers in time series, introducing further technical challenges. It is worth mentioning that models with analogous dynamic structures are common in the literature, provided their underlying distributional assumptions and dependence structures differ, leading to different performance and financial and economic interpretations.

For the sake of notation convenience, the AcF model is represented as

$$\begin{cases} Q_t &= \mu + \sigma_t Y_t^{1/\alpha_t}, \\ \log(\alpha_t) &= \gamma_0 + \gamma_1 \log(\alpha_{t-1}) + \gamma_2 \exp(-\gamma_3 Q_{t-1}), \\ \log(\sigma_t) &= \beta_0 + \beta_1 \log(\sigma_{t-1}) - \beta_2 \exp(-\beta_3 Q_{t-1}). \end{cases} \quad (9)$$

For comparison:

- In (9), $\{Q_t\}$ represents a sequence of maxima, denoted as $Q_t = \max_{1 \leq i \leq m} X_{i,t}$, similar to M_t in (3). However, $\{Y_t\}$ is a sequence of iid unit Fréchet random variables, characterized by the distribution function $F(x) = \exp(-1/x)$ for $x > 0$. The cross-sectional dimension approaches infinity, that is, $m \rightarrow \infty$.
- In our proposed AcGB2 model (4), $\{Y_t\}_{t=1}^n$ are iid variables conforming to the prime beta(p, q) distribution, equivalent to GB2($a = 1, b = 1, p, q$). The cross-sectional dimension m is fixed.

Remark 4. For references on classical extreme value theory (related to m tending to infinity) and its applications, we refer readers to Fisher and Tippett (1928), Gnedenko (1943), (Gumbel 1958), Balkema and de Haan, 1974, Lechner, Simiu, and Heckert, 1993, Goda (2000), Li et al. (2012), Dombry (2015), Zhou (2010), Hosking, Wallis, and Wood (1985), Smith (1985), Grimshaw (2017), Hosking and Wallis (1987), Ferreira and Soares (1998), Laurini and Tawn (2009), Cunnane (1973), and Ferreira and de Haan (2015).

To close this section, we note that while equations (4)–(5) and (9) share similar dynamic structures, setting $A_t > 0$ indicates that the limit form of M_t will not be Q_t . Consequently, M_t and Q_t behave distinctly, and they cannot be approximated by each other. Therefore, it is inappropriate to consider M_t as a generalization of Q_t .

2.3. Stationarity and Ergodicity

To establish the stationary and ergodic conditions for the resulting Markov chain of the stochastic series $\{A_t, B_t\}_{t=1}^n$, we adopt the approach of Chan and Tong (1994) and Zhao, Zhang, and Chen (2018). Essentially, under appropriate conditions, a nonlinear time series model achieves stationarity and ergodicity within the neighborhood of a compact attractor which may have complicated dynamics. Theorem 1 presents the theoretical results.

Theorem 1 (Stationarity and ergodicity). For the AcGB2 model with the parameter vector $\theta \in \Theta$ as defined in (7), the process $\{A_t, B_t\}_{t=1}^n$ is stationary, geometrically ergodic, and bounded.

Remark 5. Considering the relation between M_t and $\{A_t, B_t\}$ in (4), once the process $\{A_t, B_t\}_{t=1}^n$ is stationary and ergodic, $\{M_t\}_{t=1}^n$ also becomes stationary and ergodic.

3. Parameter Estimation and Asymptotic Properties

Let $\theta^0 = (\mu^0, \alpha_0^0, \alpha_1^0, \alpha_2^0, \alpha_3^0, \beta_0^0, \beta_1^0, \beta_2^0, \beta_3^0, p^0, q^0)^\top \in \Theta$ denote the true value of θ . Following (8), the log-likelihood

function of θ , given the observed $\{M_t\}_{t=1}^n$ and true initial values $\{A'_1, B'_1\}$, is

$$\begin{aligned} L_n^\circ(\theta) &= \frac{1}{n} \log \left(\prod_{t=1}^n \frac{1}{B(p, q)} \frac{A'_t (M_t - \mu)^{A'_t p - 1}}{(B'_t)^{A'_t p} \{1 + (\frac{M_t - \mu}{B'_t})^{A'_t}\}^{p+q}} \right) \\ &= \frac{1}{n} \sum_{t=1}^n \left[-\log\{B(p, q)\} + \log(A'_t) - A'_t p \log(B'_t) \right. \\ &\quad \left. + (A'_t p - 1) \log(M_t - \mu) \right. \\ &\quad \left. - (p + q) \log \left\{ 1 + \left(\frac{M_t - \mu}{B'_t} \right)^{A'_t} \right\} \right] \\ &= \frac{1}{n} \sum_{t=1}^n l_t^\circ(\theta), \end{aligned} \quad (10)$$

where $\{A'_t, B'_t\}$ are generated using θ and true initial values $\{A'_1, B'_1\}$, and $l_t^\circ(\theta) = -\log\{B(p, q)\} + \log(A'_t) - A'_t p \log(B'_t) + (A'_t p - 1) \log(M_t - \mu) - (p + q) \log\{1 + (\frac{M_t - \mu}{B'_t})^{A'_t}\}$.

However, the challenge lies in the unknown true initial values $\{A'_1, B'_1\}$ in reality. We thus focus on estimating the parameter θ without knowing $\{A'_1, B'_1\}$. This entails seeking the local maximizer of the log-likelihood function using $\{M_t\}_{t=1}^n$ and arbitrary initial values $\{A_1, B_1\}$. To achieve this, we consider $\{A_t, B_t\}$ generated with θ and arbitrary initial values $\{A_1, B_1\}$, expressing the log-likelihood function as

$$\begin{aligned} L_n(\theta) &= \frac{1}{n} \sum_{t=1}^n \left[-\log\{B(p, q)\} + \log(A_t) - A_t p \log(B_t) \right. \\ &\quad \left. + (A_t p - 1) \log(M_t - \mu) \right. \\ &\quad \left. - (p + q) \log \left\{ 1 + \left(\frac{M_t - \mu}{B_t} \right)^{A_t} \right\} \right] \\ &= \frac{1}{n} \sum_{t=1}^n l_t(\theta). \end{aligned} \quad (11)$$

We will show in [Theorems 2–4](#) that regardless of the true initial value, the local maximizer of $L_n(\theta)$ is consistent, asymptotically normal, and unique.

Theorem 2 (Consistency). Under the conditions in [Theorem 1](#), for a sequence $\{\delta_n\}$ satisfying $\delta_n \downarrow 0$, $n\delta_n \rightarrow +\infty$, and $\frac{1}{\delta_n}(\min_{1 \leq t \leq n} M_t - \mu^\circ) \xrightarrow{P} +\infty$ as $n \rightarrow \infty$, there exists a sequence of local maximizers $\hat{\theta}_n$ of $L_n(\theta)$ for $\{M_t\}_{t=1}^n$ from the AcGB2 model with $\theta \in \Theta$ in (7) and the true parameter θ° in a compact subset of Θ , such that $\|\hat{\theta}_n - \theta^\circ\| \leq \delta_n$ and $\hat{\theta}_n \xrightarrow{P} \theta^\circ$, where $L_n(\theta)$ is the log-likelihood function in (11) with θ and arbitrary initial values $\{A_1, B_1\}$.

For entries θ_j and θ_k in the parameter vector θ , define $m_{\theta_j, \theta_k}(\theta^\circ) = -E_{\theta^\circ}(\frac{\partial^2 l_t^\circ(\theta)}{\partial \theta_j \partial \theta_k} |_{\theta=\theta^\circ})$. Under the conditions of [Theorem 2](#), Lemma 2 states that $E_{\theta^\circ}(\frac{\partial l_t^\circ(\theta)}{\partial \theta} |_{\theta=\theta^\circ}) = \mathbf{0}$, and $\text{Var}_{\theta^\circ}(\frac{\partial l_t^\circ(\theta)}{\partial \theta} |_{\theta=\theta^\circ}) = -E_{\theta^\circ}(\frac{\partial^2 l_t^\circ(\theta)}{\partial \theta \partial \theta} |_{\theta=\theta^\circ})$ is positive definite. Hence, $m_{\theta_j, \theta_k}(\theta^\circ)$ is an element in the matrix $\text{Var}_{\theta^\circ}(\frac{\partial l_t^\circ(\theta)}{\partial \theta} |_{\theta=\theta^\circ})$.

Theorem 3 (Asymptotic normality). Under the conditions in [Theorem 1](#), the local maximizer $\hat{\theta}_n$ in [Theorem 2](#) is asymptotically normal,

$$\sqrt{n}(\hat{\theta}_n - \theta^\circ) \xrightarrow{D} N\left(\mathbf{0}, \left\{ -E_{\theta^\circ} \left(\frac{\partial^2 l_t^\circ(\theta)}{\partial \theta \partial \theta} \Big|_{\theta=\theta^\circ} \right) \right\}^{-1} \right). \quad (12)$$

The proof of [Theorem 3](#), in Appendix A.3, starts with a scalar parameter μ and extends to the vector parameter θ of interest. The uniqueness is given in [Theorem 4](#).

Theorem 4 (Uniqueness). Under the conditions in [Theorems 1](#) and [2](#), there exists an asymptotically unique MLE over the set $\Theta_{*,n} = \{\theta \in \Theta : \mu \leq c \min_{1 \leq t \leq n} M_t + (1 - c)\mu^\circ\}$, where Θ is from (7) and $0 < c < 1$.

We remark that conditions in the set $\Theta_{*,n}$ are technical ones, and the constant c is arbitrary. Theoretically, $M_t \geq \mu^\circ$. When $\min_{1 \leq t \leq n} M_t$ is approaching μ° , the upper bound of μ is μ° .

In addition, although the AcGB2 model settings are similar to those in AcF models, the technical arguments differ due to two key factors: (a) the existence of moments depends on dynamic parameters, and (b) while the Fréchet density and distribution functions are easier to derive, the GB2 distribution function lacks an explicit form, making it more challenging to work with. Furthermore, we provide a complete proof of Lemma 5 of Smith (1985), which has been widely referenced in the literature but was not originally proven in Smith's paper or its subsequent references. Our proof of this widely referenced lemma is included in the proof of [Theorem 2](#).

4. Simulation Experiments

4.1. Large Sample Performance of the AcGB2 Model Estimation

In this section, we simulate data from the AcGB2 model in (4)–(5) using different time lengths n . The true value of the parameter vector $(\mu, \alpha_0, \alpha_1, \alpha_2, \alpha_3, \beta_0, \beta_1, \beta_2, \beta_3, p, q)$ is set to $(-0.00714, -0.318, 0.904, 0.460, 7.180, -1.670, 0.390, -0.675, 9.170, 2.000, 2.000)$, based on an exploratory real data analysis. To compare the estimators across varying data lengths, we conduct 500 replications of the simulation, each with the time lengths of n at 1000, 3000, and 5000, separately. Each simulation yields the MLE of the parameter vector, resulting in 500 estimated sets for each data length. The optimization problem (11) is solved using the `nlmminb` function in R version 4.1.3. We compare the means and Monte Carlo standard deviations of each parameter estimate. The findings are summarized in Table 1.

From Table 1, we observe that all estimated parameters are close to the true values. Longer data lengths yield improved results, indicating that most estimated parameter values are closer to the true ones. Additionally, estimates obtained from longer datasets exhibit smaller standard deviations, except for α_2 and β_2 . This may be because α_2 and β_2 can be combined with α_3 and β_3 when applying Taylor expansions of the exponential functions. We have noted that with sufficiently large data lengths, the MLE becomes accurate and stable.

Figure 1 illustrates a typical scenario depicting the dynamics of A_t and B_t with a simulated sample size of $n = 1000$ and the parameter values specified above.

At first glance, one may perceive A_t and B_t are nonstationary due to some disguised nonstationary patterns. However, we note that our nonlinear time series models exhibit stationarity under the specified conditions, differing from the typical graphical stationary patterns observed in linear time series models. These features occur repeatedly, but that does not imply nonstationarity in the depicted time series in Figure 1. Later, in our real data model fitting and computations, we will see similar patterns in the recovered A_t and B_t . In the figure, we have included a rectangle in the left panel and an ellipse in the right panel, representing clusters of extreme/worst observations. For A_t , smaller values signify higher market risk, while for B_t , larger values indicate higher market risk.

In practice, data that perfectly fit Models (4) and (9) are rarely encountered. To ensure the robustness of our estimation, we introduce uniform noise to (4) and assess the performance, as shown in Table 2. The results indicate that the estimated parameter values closely match those in Table 1, suggesting that the AcGB2 models are applicable to a wide range of data examples.

4.2. Motivating Examples to Justify the Model

In this section, we present illustrative examples to validate modeling the maxima of observations from heterogeneous populations by GB2 distributions, and then demonstrate the inadequate performance of Fréchet models in fitting data generated from GB2 distributions. We have the following Fact 4.1 and Examples 4.1–4.4.

Fact 4.1. Suppose $F(x; t)$ is the cumulative distribution function (C.D.F.) of $\text{GB2}(A_t, B_t, p, q)$, and $F_0(x; t), F_1(x; t), \dots, F_m(x; t)$ are C.D.F.s satisfying $F(x; t) = F_0(x; t) \times F_1(x; t) \times \dots \times F_m(x; t)$. Suppose that for any t , $W_{i,t} = \max(Z_{0,t}, Z_{i,t})$, $i = 1, \dots, m$, and $\{Z_{i,t}\}_{i=0}^m$ are independent; the C.D.F. of $W_{i,t}$ is $F_0(x; t) \times F_i(x; t)$. Then it is easy to check that the C.D.F. of $M_t = \max_{1 \leq i \leq m} W_{i,t}$ is $F(x; t)$.

Example 4.1. For two special cases in Fact 4.1, (1) (cross-sectional independence) with $Z_{0,t} = -\infty$, $F_i(x; t) = \{F(x; t)\}^{\delta_i}$, where $\delta_i > 0$ such that $\delta_1 + \dots + \delta_m = 1$, it follows that the C.D.F. of $M_t^* = \max_{1 \leq i \leq m} W_{i,t}$ is $F(x; t)$, with a particular case being $\delta_i = 1/m$; (2) (complete dependence) $F_0(x; t) = F(x; t)$, $Z_{i,t} = -\infty$, $i = 1, \dots, m$.

Remark 6. We note that C.D.F.s $F_i(x; t)$ in Fact 4.1 may originate from different parametric families, for example, they are related to heterogeneous populations. The condition $F(x; t) = F_0(x; t) \times F_1(x; t) \times \dots \times F_m(x; t)$ is not necessary for $M_t^* = \max_{1 \leq i \leq m} W_{i,t}$ to follow the distribution function $F(x; t)$. Consequently, assuming M_t^* to be GB2-distributed is natural when modeling maxima practically, such as in the case of cross-sectional stock returns in this article, where m represents a fixed, medium-sized value.

Example 4.2. Suppose $\{W_{i,t}\}_{i=1}^\infty$ is an iid random sequence where $\max_{i \in \{i_1, \dots, i_m\}} W_{i,t}$ for any $\{i_1, \dots, i_m\}$ has the distribution function $F(x; t)$. Define $M_t^{(k)} = \max_{m*(k-1)+1 \leq i \leq m*k} W_{i,t}$, $k = 1, 2, 3, \dots$. Then $\max_{1 \leq i < \infty} W_{i,t}$ and $\max_{1 \leq k < \infty} M_t^{(k)}$ are

asymptotically almost surely equal. Hence, the limiting extreme value distributions of $\max_{1 \leq i < \infty} W_{i,t}$ and $\max_{1 \leq k < \infty} M_t^{(k)}$, each with distinct normalizing constants, belong to the same domain of attraction as an extreme value distribution associated with $F(x; t)$.

Note that the iid condition is not necessary to determine the limiting extreme value distribution.

Example 4.3. Consider $\{W_{i,j,t}\}_{i=1,j=1}^\infty$ as an array of random variables where $M_{j,t}^* = \max_{1 \leq i \leq m} W_{i,j,t}$ follows the distribution function $F(x; t)$. Then $\max_{1 \leq i \leq m, 1 \leq j < \infty} W_{i,j,t}$ and $\max_{1 \leq j < \infty} M_{j,t}^*$ are asymptotically almost surely equal. Consequently, the limiting extreme value distributions of $\max_{1 \leq i \leq m, 1 \leq j < \infty} W_{i,j,t}$ and $\max_{1 \leq j < \infty} M_{j,t}^*$, each with different normalizing constants, belong to the same domain of attraction as an extreme value distribution related to $F(x; t)$.

Example 4.4. Using Fact 4.1, the cross-sectional maxima of the following multivariate time series exhibit the assumed GB2 distribution in the article:

$$X_{i,t} = \mu + B_t W_{i,t}^{A_t}, \quad i = 1, \dots, m$$

where B_t and A_t are defined by equations (3)–(5), and $F(x; t)$ represents $\text{GB2}(1, 1, p, q)$.

Remark 7. Examples 4.2–4.4 justify that when m represents a fixed, medium-sized value, the application of extreme value theory is not applicable, necessitating a direct application of $F(x; t)$. The univariate time series depicted in Example 4.4, where A_t s remain constant, is an extended GARCH(1, 1) model, consequently making the multivariate time series in this example an extended multivariate GARCH(1, 1) model. Note that for M_t , defined by (3)–(5), to follow the GB2 distribution, it is not necessary for $X_{i,t}$ to follow the definition in Example 4.4.

Remark 8. In financial econometrics literature, GARCH(1, 1) models are widely used in analyzing stock log-returns and index returns. A well-suited GARCH(1, 1) model for index returns (e.g., S&P 100) does not guarantee an equally good fit for individual returns. While fitting a multivariate GARCH model to multivariate time series (stock returns) is common, the cross-sectional weighted-sum time series of such a multivariate GARCH model may differ from a univariate GARCH model fitted directly using index returns. In the literature, a cross-sectional weighted-sum time series (e.g., index returns) depicts the overall market performance average. These characteristics are also shared by our proposed models, particularly where a cross-sectional maxima time series describes either the worst or best market performance scenarios.

Under the conditions in Example 4.1, considering $m \in \{500, 700, 900\}$, and setting $\delta_i = 1/m$ for $i = 1, \dots, m$, we simulate $\{X_{i,t} : i = 1, \dots, m, 1 \leq t \leq n\}$ with $n = 10,000$ from $\text{GB2}(2, 5, 3, 3)$ and obtain $\{M_t\}_{t=1}^n$ accordingly. We then fit Fréchet distributions, using the C.D.F. $F^*(x) = \exp\{-(x - \mu)/\kappa\}^{-\xi}$ for $x > \mu$, to $\{M_t\}_{t=1}^n$, where the location parameter μ , the scale parameter $\kappa = \exp(\beta_0)$, and the shape parameter $\xi = 1/\exp(\gamma_0)$ are estimated using the MLE. Figure 2 plots

the corresponding p.d.f.s and C.D.F.s for the fitted Fréchet distributions. Observing as m increases, both p.d.f.s and C.D.F.s progressively approach the true lines of GB2(2, 5, 3, 3), indicating reduced gaps between the fitted and true lines. However, it is noteworthy that even with m at 500, the simulation demonstrates a poor performance in approximating the GB2 distribution by the Fréchet distribution. Particularly, around $m = 30$, an even larger and significant gap (not shown in the plot) emerges between the fitted Fréchet distribution and the true GB2 distribution, indicating the inefficiency of models based on Fréchet distributions. This simple example demonstrates the suitability of GB2 distributions as strong candidates for modeling maxima in time series data.

Remark 9. Zhao, Zhang, and Chen (2018) established theory for applying AcF models to real data, when the number of cross-sectional (multivariate) time series tends to infinity within a general factor model framework. However, in our new Fact 4.1 for AcGB2, the dependence structures exhibit much more flexibility compared to the general factor model settings. One notable distinction is that AcGB2 represents a distribution assumption and does not involve the limitations present in AcF models. We note that mean, median, minimum, maximum, and standard deviation are among the most commonly used statistics across various disciplines. In time series contexts, ARMA models are linked to mean auto-regressions, GARCH models to standard deviations, while median dynamics can be characterized in quantile auto-regressions. However, auto-regressions on minimum and maximum values are still in an early developmental phase.

Different from the above examples, much of the literature has focused on analyzing factor structures directly within individual stock return series and modeling idiosyncratic returns after removing the factor components. A simple synthetic experiment was implemented in Zhao, Zhang, and Chen (2018). This approach, along with the widely used GARCH filtering method, can certainly be applied within the AcGB2 framework, as well as in the AcF framework. Such a modeling strategy can inspire numerous new research avenues for interested scholars. Fact 4.1 introduces a new framework that accommodates cross-sectional (in)dependent returns with heterogeneous marginal population distributions, as demonstrated in Example 4.1. In Section 5, we illustrate its broad applicability in dynamically fitting cross-sectional maxima.

5. Real Data Applications

In this section, we apply AcGB2 models to examine the worst performance (i.e., maxima of negative simple-returns) within three typical market datasets. DJIA 30, one of the most commonly followed stock market index, gauges the performance of 30 major companies (e.g., Apple and Disney) listed on U.S. stock exchanges. S&P 100 comprises 101 leading U.S. stocks with exchange-listed options, including one company with dual classes of stocks. Primary dealers serve as trading counterparts of the New York Fed in executing monetary policy. From January 1, 2015, to November 30, 2022, we collected stock prices from Yahoo Finance, ensuring availability across the entire period.

This results in 29 stocks for DJIA 30, 97 stocks for S&P 100, and 22 for primary dealers. The data underwent conversion into negative simple-returns, using $X_{i,t} = -p_{i,t}/p_{i,t-1} + 1$ or an equivalent form for data transformations, where $p_{i,t}$ represents the i th stock price at day t . The names of the 22 primary dealers are provided in Appendix B.

Maxima time series were constructed within each market—DJIA 30, S&P 100, and primary dealers, respectively. Each series is divided into two segments: an in-sample period from January 1, 2015, to December 31, 2019, and an out-of-sample period from January 1, 2020, to November 30, 2022. The AcGB2 model was fitted to each in-sample series.

For the sake of comparison, we also fit the data to the benchmark AcF model and then compare the recovered dynamic scale parameter processes and tail index processes from both models, respectively. In addition, a comparison to the average volatilities derived from GARCH models will also be conducted. Finally, we demonstrate the predictabilities of the fitted models. Among these comparisons, illustrated scale parameter and tail index dynamics are presented in Appendix C.

In our numerical experiments, we initially fitted the original maxima time series of the negative log-returns using both the AcGB2 and AcF models. The results for the AcGB2 model were unsatisfactory. We then applied exponential transformations to the negative log-returns, effectively converting them into simple returns. The validity of this exponential transformation is explained in Remark 10.

Remark 10. An exponential transformation can convert Gumbel distributed data into Fréchet distributed data. Similarly, it can transform EGB2 (exponential GB2) distributed data into GB2 distributed data. Consequently, an exponential transformation can enhance the stability of the estimation process in an extreme value context, akin to how the logarithm transformation achieves normality in various applications. Logarithm-transformed data are frequently employed in fitting regression models, that is, focusing on their central tendency. Analogously, considering exponentially transformed data might yield better results when fitting extreme value distributions. Through such exponential transformation, the dynamics of parameters unveil distinct latent characteristics of the underlying data-generating process, subsequently offering different economic interpretations. We establish an equivalence between the exponentially transformed data and simple-returns as follows.

Now, suppose $p_{i,t-1}$ and $p_{i,t}$ are the i th stock's prices at time $t - 1$ and t , and $X_{i,t} = \exp\{-\log(p_{i,t}/p_{i,t-1})\}$. Then, $M_t = \max_{1 \leq i \leq m} p_{i,t-1}/p_{i,t}$. Define $V_{i,t} := p_{i,t-1}/p_{i,t}$, we have:

$$\max_{1 \leq i \leq m} p_{i,t-1}/p_{i,t} = \max_{1 \leq i \leq m} V_{i,t} = \mu + B_t Y_t^{1/A_t}, \quad 1 \leq t \leq n, \quad (13)$$

which is equivalent to

$$\min_{1 \leq i \leq m} p_{i,t}/p_{i,t-1} = 1/(\mu + B_t Y_t^{1/A_t}), \quad 1 \leq t \leq n,$$

further equivalent to

$$\min_{1 \leq i \leq m} (p_{i,t}/p_{i,t-1} - 1) = 1/(\mu + B_t Y_t^{1/A_t}) - 1, \quad 1 \leq t \leq n. \quad (14)$$

This models the worst simple-returns across m assets. It is clear that (4) with the exponentially transformed negative log-returns is equivalent to simple-returns in (14). Consequently, we will directly address $V_{i,t}$ in (13) in our computations. It is worth noting that subtracting 1 from both sides of (4) does not alter the model but introduces greater complexity compared to (14). Additionally, due to the appearance of A_t and B_t in the denominator in (14), their dynamic patterns differ from α_t and σ_t in the AcF model. This implies that the maxima of negative log-returns and the minima of simple-returns possess distinct parameter dynamics. Hence, exploring simple-returns in the extreme context may lead to different conclusions and unveil diverse latent market structures.

Throughout the article, the maxima time series represent transformed negative log-returns ($V_{i,t}$) in (13). For example, Figure 3 illustrates the evolution of transformed maxima of $V_{i,t}$ over time. A noticeable discrepancy in extreme values among the three panels is immediately apparent. We have used ellipses to highlight clustered extreme observations, indicative of significant market events. These observations correspond to various periods, including: The 2015–16 stock market selloff, often termed as The Great Fall of China, characterized by a decline in stock prices globally between June 2015 and June 2016. Specific events in December 2018: (i) S&P 500 down 9%, making the worst December since 1931. Dow experiencing an 8.7% decline, where, in a seven-day stretch, it fell by 350 points or more six times. (ii) A market crash on January 3, 2019: Dow decreased almost 2.8% to close at 22,686.22. S&P 500 also decreased by 2.5% to close at 2447.89. Nasdaq closed at 6463.50, losing 3%.

Table 3 summarizes the daily percentage change in prices. The minimum, mean, median, and maximum values from the three markets indicate that they behave differently. We observe that primary dealers experienced more instances where all stock prices rose together (Column $< 0\%$) compared to DJIA 30, with DJIA 30 outperforming S&P 100 in this respect. Stock prices dropping more than 5% or lower occurred frequently, accounting for almost a third of the events among S&P 100 stocks. There were days across all three markets when prices dropped more than 10%. Additionally, extremely large values fluctuations (exceeding 15%–30%) appeared around early 2016 and the middle of 2019 in S&P 100 and primary dealers' data. These observations are also highlighted in the figure via ellipses. These extreme occurrences are reflective of inherent market characteristics (e.g., A_t , B_t , α_t , σ_t in our models), which will be detailed in subsequent subsections.

5.1. Fitting both AcGB2 and AcF to the Maxima

Using the conditional maximum likelihood estimation procedure developed in Section 3, we obtain the log-likelihood values in Table 4, along with the estimated parameter values detailed in Tables 5–7.

Table 4 demonstrates that the AcGB2 models exhibit slightly higher log-likelihood values compared to the AcF models. Note that while the AcGB2 model has 11 parameters, the AcF model consists of 9 parameters. However, the GB2 distribution and the Fréchet distribution are not linked through a parameter limit,

making it inappropriate to infer that AcGB2 models should inherently yield larger likelihood values than the AcF models. This is because although we can arbitrarily choose a distribution with more parameters, it does not necessarily ensure a better fit to the cross-sectional maxima observed in our real data.

From Tables 5–7, it is evident that the majority of parameters in both models are significantly fitted. The autoregressive parameters α_1 and γ_1 exhibit significance across the board. Some parameter values hover close to 1, indicating clustered behavior within the models. Of particular interest is the proximity of the fitted location, intercept, and lag-1 autoregressive parameters: μ , α_0 , α_1 in the AcGB2 model, and μ , γ_0 , γ_1 in the AcF model, respectively. This proximity strongly suggests that both models effectively capture the characteristics of the transformed data ($V_{i,t}$).

To illustrate the fitting efficiency, we conduct 10,000 Kolmogorov-Smirnov (K-S) tests on the recovered $\{Y_t\}$ processes by simulating 10,000 sequences following the GB2 distribution and unit Fréchet sequences, each with length 1257, mirroring the length of the real data. Table 8 presents the number of rejections of the null hypothesis regarding the innovation process being GB2 distributed or following a unit Fréchet distribution at significant levels of 0.01 and 0.05.

The AcGB2 models demonstrate confidence levels within the acceptable range of 99% and 95%. Conversely, the AcF models raise doubts, particularly for DJIA 30 at both confidence levels and S&P 100 at the 95% level. These observations carry substantial implications for predictions, as will be demonstrated in Section 5.3. Table 8 underscores the preference for the AcGB2 model over the AcF model, a trend further substantiated in Sections 5.2 and 5.3.

While Table 8 favors AcGB2 over AcF, readers should exercise caution. Non-rejection of the null hypothesis does not assert the absolute truth of the hypothesis itself. Furthermore, the uncertainty level increases because the test statistics rely on recovered values assumed to follow specific distributions. Nevertheless, the AcGB2 model demonstrates satisfactory performance, whether using estimated parameter values or true values.

We note that the AcGB2 model comprises 11 parameters compared to 9 in the AcF model, making estimation for finite sample sizes trickier and more challenging for AcGB2 models. Additionally, both models represent nonlinear time series models for extreme observations. Therefore, it is unsurprising that some fitted values remain insignificant in real applications and simulation examples.

5.2. Comparison with the Market Average Volatilities

Considering that M_t and Q_t are transformed data of simple-returns in (13), inferences based on the fitted model need a logarithm transformation to return to the original scales. Therefore, the dynamics of the tail parameter $\{A_t\}$ in the AcGB2 model and the tail index parameter $\{\alpha_t\}$ in the AcF model play a role similar to the volatility process in the GARCH model. In this section, we compare the recovered $\{A_t\}$ and $\{\alpha_t\}$ to the GARCH volatilities. We fit a GARCH(1, 1) model with t -distributed innovations having degrees of freedom set to 5 for each negative log-return

series of stocks in DJIA 30, S&P 100, and 22 Primary dealers from January 1, 2015, to December 31, 2019. Daily average volatilities are calculated from all fitted GARCH models within each market, providing insights into market volatility. Figures 4–6 illustrate the reciprocal dynamic tail parameter series ($\{1/A_t\}$ and $\{1/\alpha_t\}$, respectively), together with the daily average volatilities derived from the GARCH models.

In Figures 4–6, the red lines denote the daily average volatilities generated by GARCH models. The left plot's black line represents the recovered reciprocal tail parameter series from the AcGB2 model, while the right plot's black line portrays the recovered reciprocal tail index parameter series from the AcF model. To enhance visualization, all daily average volatilities and recovered reciprocal tail parameter values are standardized by subtracting their means and dividing by their standard deviations. Observing both plots, the black lines exhibit similarities and share comparable patterns with the averaged GARCH volatilities of the market.

To quantify these relationships, we calculate linear correlations between the recovered reciprocal tail parameter values and the GARCH daily average volatilities for both models. The resulting correlations are listed in Table 9. Examining the table, the correlation for DJIA 30 is 6.4% higher, a noteworthy margin in finance terms. For PD 22, the correlation stands at 0.8%, which may be considered less significant. Notably, in the case of SP100, AcF exhibits a 0.2% advantage over AcGB2. These patterns suggest that for intermediate-sized cross-sectional stocks, our AcGB2 model holds advantages over AcF. Specifically, AcGB2 performs better in approximating relatively large values in GARCH average volatilities for DJIA 30 and PD 22, while showing nearly equivalent performance in S&P 100. An intriguing observation is that both AcGB2 and AcF models demonstrate a robust capability in capturing market volatilities within primary dealers' returns.

We observe that the AcF model's recovered α_t values exceed 3 across all three datasets, contradicting with Zhao, Zhang, and Chen (2018). This discrepancy arises due to several factors. First, the transformation alters the dependence structure and consequently impacts the derived latent α_t (A_t) and σ_t (B_t) processes. Second, normal and exponential normal (EGB2 and GB2) distributions possess distinct max-domain of attractions, leading to differing theoretical properties. To illustrate this, consider an analogy: a monthly log-return equals the sum of daily log-returns within that month, while a monthly simple-return does not necessarily equal the sum of daily simple-returns. While either accounting convention (simple-return or log-return) may suffice in practical applications, they can yield markedly different limit properties when assessing risks. Regarding the performance of the AcF model with or without data transformation, it is not appropriate to draw a definitive conclusion as both approaches can yield meaningful results but with different interpretations. For instance, when assessing market risk, stress tests may lead to a preference for either method based on the specific context. We further note that normal distribution and t distribution possess markedly different max-domain of attractions, even though in basic practice, the t distribution can be approximated by the normal distribution.

5.3. Examining Model Predictability

In this section, we delve into assessing the predictability of the AcGB2 and AcF models. The prediction interval spans from January 1, 2020, to November 30, 2022. For instance, considering the AcGB2 model: using the real value of maxima at $t - 1$ along with estimated parameters, we derive the values of A_t and B_t . Subsequently, leveraging A_t , B_t , $\hat{p}$, $\hat{q}$, and $\hat{\mu}$, we simulate 1000 M_t values. Among these values, the 95th percentile is considered as the predicted value at time t . This procedure is iterated to generate predicted values within the time span of January 1, 2020, to November 30, 2022, which are then compared with the actual time series of maxima. The outcomes are illustrated in Figures 7–9.

In Figures 7–9, the blue line represents observed maxima values, the red lines depict predicted maxima, and the magenta lines show the lower 5% predicted percentiles (AcGB2 model on the left and AcF model on the right). Comparing these plots, a notable difference emerges: the predicted values (red lines) in Figure 7 significantly differ from those in Figures 8 and 9. This discrepancy can be attributed to significant market disruptions occurring in March, where the benchmarks performed poorly due to the outbreak and spread of COVID-19. Notably, on March 9, 12, and 18, 2020, the maxima drops were 24.4%, 20%, and 25%, respectively among DJIA 30. Table 10 provides some fundamental statistics regarding model predictability.

Analyzing the percentages in the table, we can see that the AcGB2 fitted models outperform the AcF fitted models. The percentages of violations (Hit functions) wherein actual maxima of the negative log-returns exceed predictions is around 5% for AcGB2 fitted models, compared to approximately 6% for AcF fitted models. Examining the Max.P.V columns further underscores this distinction: the AcF model produces a significantly higher predicted value of 67.9573 compared to AcGB2's 15.5127, over four times larger. This disparity causes smaller values to be masked in the plots shown in Figure 7 due to the higher predictions of the AcF model. If the plots were drawn on the original scales, they would be difficult to interpret. Instead, we provided two truncated plots along with a comparison plot showing the 95% percentiles predicted by the AcGB2 model versus those predicted by the AcF model. It is evident that the AcF model predicts higher extreme values compared to the AcGB2 model. However, higher predicted values from the AcF model do not necessarily indicate better performance, as shown in Table 10. This is because these predictions significantly exceed actual observed values, leading to smaller predictions for other values and, consequently, a greater number of days where $R > P$. Conversely, the AcF models may tend to under-predict market risks for S&P 100 and PD 22, with predicted values of 1.8399 and 1.4251, respectively, against AcGB2's values of 2.0586 and 1.5231. Figure 8's two panels exhibit striking similarity, as do the panels in Figure 9, indicating the potential interchangeability of the AcGB2 model for the AcF model. In Table 10, it is evident that AcGB2 consistently holds advantages over AcF in terms of "Days of $R > P$ " and "Percents".

It is interesting to note that maxima derived from transformed negative simple-returns ($V_{i,t}$) across 22 primary dealers align well with both the AcGB2 and AcF models. This suggests

that the market patterns of primary dealers are robust across different dynamic risk models, such as AcF and AcGB2.

6. Discussions and Conclusions

Accurately modeling and predicting worst-case scenarios across various applications—such as finance, insurance, regional and global climate extremes, economics, social tensions, geopolitical risks, and EEG signals in brain functional research—significantly enhance decision-making processes. These scenarios are often quantified numerically as either maxima or minima values. This work focuses on modeling cross-sectional maxima derived from medium to high-dimensional financial time series. We introduce a novel dynamic “autoregressive conditional generalized beta distribution of second kind” (AcGB2) framework. Using the standard GB2 distribution, we effectively fit highly skewed and long-tailed data, integrating dynamic behavior by introducing time-varying properties to the parameters. We investigate the stationarity and ergodicity of the dynamic AcGB2 model, analyze the conditional maximum likelihood estimators for the parameters, and study their asymptotic properties. Through simulations and real data analysis, we showcase the models’ flexibility, robustness, and goodness of fit.

Both the AcF and AcGB2 models are designed to investigate dynamic and clustered extreme observations, highly relevant to systemic risk and market crises. The impact of clustered extreme observations is more substantial than that of single extremes. The relationship between the AcGB2 model and the AcF model is akin to that between GARCH models with t innovations and those with normal innovations. Notably, when the degrees of freedom of t random variables approach infinity, their limit becomes normally distributed random variables.

In the literature, various tests for model selection and predictability have been developed, such as Vuong (1989) and Diebold and Mariano (1995). However, these tests are not directly applicable to our new models as we focus on dealing with maxima of multivariate time series, tail indexes, and clustered extremes.

Beyond the primary methodological contributions of the AcGB2 model to financial econometrics literature, we propose a new approach to data transformation for real data: the transformation of maxima time series of simple-returns (see Remark 10 for justification), which enhances the reliability of parameter fittings. Another new methodological contribution is the capacity of the tail index processes within the AcGB2 model (as well as the AcF model) to characterize tail risk behaviors in the market and emulate average market volatilities. These processes can aid in constructing new market risk indexes. This article compared the recovered $1/A_t$ with the average of fitted GARCH(1, 1) volatilities of all component time series. Further investigation into its correlation with the fitted GARCH(1, 1) volatility of the index returns directly and a comparison with findings from Hansen and Lunde (2005) could offer additional insights.

When comparing the AcGB2 model and the AcF model, we note that the AcGB2 model incorporates two additional parameters, p and q . To create a more parsimonious AcGB2 model, one can set $p = 1$ and $q = 1$. However, retaining

p and q as unknown variables enhances the performance and flexibility of model fitting. It is crucial to note that since the AcGB2 model and the AcF model lack a hierarchical structure, the notion that more parameters guarantee better performance does not universally apply. Nevertheless, if a model significantly outperforms existing ones in specific scenarios, further developments are warranted. For instance, in a recent study, Lin, Liu, and Zhang (2023) demonstrated a novel EGB2 option pricing formula that surpassed the widely used Black-Scholes formula and other jump diffusion models, demonstrating its potential impact in asset pricing as in financial practice a small percent change can represent a significant large change in total money.

Supplementary Materials

supplementary_JBES-P-2024-0048.pdf: The online supplementary file includes: Appendix A, which provides technical proofs from the article; Appendix B, which lists the 22 Primary Dealers; Appendix C, which illustrates scale parameter and tail index dynamics; and Appendix D, which contains tables and figures from the article. (PDF)

The R and MATLAB implementations, including a README file and demo codes for the S&P 100 data, are available on GitHub at https://github.com/ChunmingZhangUW/AcGB2_JBES.

Acknowledgments

The authors thank the editor, associate editor, and two anonymous referees for their constructive comments and suggestions, which significantly improved the presentation and clarity of the article. The authors also thank Ruodu Wang for insightful discussions that helped address a technical gap in the literature.

Disclosure Statement

The authors report there are no competing interests to declare.

Funding

The research was supported by NSF 71991471, the U.S. NSF grants DMS-2013486 and DMS-1712418, and funding provided by the University of Wisconsin-Madison Office of the Vice Chancellor for Research and Graduate Education through the Wisconsin Alumni Research Foundation.

ORCID

Chunming Zhang  <http://orcid.org/0000-0002-3153-2662>

Zhengjun Zhang  <http://orcid.org/0000-0003-2615-1539>

References

- Bali, T., and Weinbaum, D. (2007), “A Conditional Extreme Value Volatility Estimator based on High-Frequency Returns,” *Journal of Economic Dynamics and Control*, 31, 361–397. [968]
- Balkema, A., and de Haan, L. (1974), “Residual Life Time at Great Age,” *Annals of Probability*, 2, 792–804. [971]
- Bollerslev, T. (1986), “Generalized Autoregressive Conditional Heteroskedasticity,” *Journal of Econometrics*, 31, 307–327. [970]
- Caivano, M., and Harvey, A. (2014), “Time-Series Models with an EGB2 Conditional Distribution,” *Journal of Time Series Analysis*, 35, 558–571. [970]
- Chan, K., and Tong, H. (1994), “A Note on Noisy Chaos,” *Journal of Royal Statistical Society, Series B*, 56, 301–311. [971]

- Chavez-Demoulin, V., Embrechts, P., and Sardy, S. (2014), "Extreme-Quantile Tracking for Financial Time Series," *Journal of Econometrics*, 188, 44–52. [968]
- Chen, L., and Singh, V. (2017), "Generalized Beta Distribution of the Second Kind for Flood Frequency Analysis," *Entropy*, 19, 254. [970]
- Chiang, T. C., and Li, J. (2015), "Modeling Asset Returns with Skewness, Kurtosis, and Outliers," in *Handbook of Financial Econometrics and Statistics*, eds. C. F. Lee and J. Lee, pp. 2177–2215, New York: Springer. [970]
- Corrado, C. (2001), "Option Pricing based on the Generalized Lambda Distribution," *Journal of Futures Market*, 21, 213–236. [970]
- Cui, Q., and Ma, Y. (2014), "Pricing Synthetic CDO with MGB2 Distribution," *Statistics and Its Interface*, 7, 309–318. [970]
- Cunnane, C. (1973), "A Particular Comparison of Annual Maxima and Partial Duration Series Methods of Flood Frequency Prediction," *Journal of Hydrology*, 18, 257–271. [971]
- Diebold, F., Schuermann, T., and Stroughair, J. (1998), "Pitfalls and Opportunities in the Use of Extreme Value Theory in Risk Management," in *Decision Technologies for Computational Finance*, volume 2 of Advances in Computational Management Science, eds. A. Refenes, J. Moody, and A. Burgess, pp. 3–12, New York: Springer. [968]
- Diebold, F. X., and Mariano, R. S. (1995), "Comparing Predictive Accuracy," *Journal of Business & Economic Statistics*, 13, 253–263. [977]
- Dombry, C. (2015), "Existence and Consistency of the Maximum Likelihood Estimators for the Extreme Value Index Within the Block Maxima Framework," *Bernoulli*, 21, 420–436. [971]
- Engle, R. F. (1982), "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of UK Inflation," *Econometrica*, 50, 987–1007. [970]
- Fan, J., and Yao, Q. (2003), *Nonlinear Time Series: Nonparametric and Parametric Methods*, Springer Series in Statistics, New York: Springer-Verlag. [970]
- Ferreira, A., and de Haan, L. (2015), "On the Block Maxima Method in Extreme Value Theory: PWM Estimators," *Annals of Statistics*, 43, 276–298. [971]
- Ferreira, J., and Soares, C. (1998), "An Application of the Peaks over Threshold Method to Predict Extremes of Significant Wave Height," *Journal of Offshore Mechanics and Arctic Engineering*, 120, 165–176. [971]
- Fisher, R., and Tippett, L. (1928), "Limiting Forms of the Frequency Distribution of the Largest or Smallest Member of a Sample," *Mathematical Proceedings Of The Cambridge Philosophical Society*, 24, 180–190. [971]
- Frees, E., and Valdez, E. (2008), "Hierarchical Insurance Claims Modeling," *Journal of the American Statistical Association*, 103, 1457–1469. [970]
- Gnedenko, B. (1943), "Sur la distribution limite du terme maximum d'une serie aleatoire," *Annals of Mathematics*, 44, 423–453. [971]
- Goda, Y. (2000), *Random Seas and Design of Maritime Structures*, Singapore: World Scientific. [971]
- Grimshaw, S. (2017), "Computing Maximum Likelihood Estimates for the Generalized Pareto Distribution," *Technometrics*, 35, 185–191. [971]
- Gumbel, E. (1958), *Statistics of Extremes*, New York: Colombia University Press. [971]
- Hall, S., Swamy, P., and Tavas, G. (2016), "Time-Varying Coefficient Models: A Proposal for Selecting the Coefficient Driver Sets," *Macroeconomic Dynamics*, 1, 1–17. [970]
- Hansen, P. R., and Lunde, A. (2005), "A Forecast Comparison of Volatility Models: Does Anything Beat a GARCH(1,1)?" *Journal of Applied Econometrics*, 20, 873–889. [977]
- Higbee, J. D., and McDonald, J. B. (2021), "A Comparison of the GB2 and Skewed Generalized Log-t Distributions with an Application in Finance," *Journal of Econometrics*, 240, 105064. [970]
- Hosking, J., and Wallis, J. (1987), "Parameter and Quantile Estimation for the Generalized Pareto Distribution," *Technometrics*, 29, 339–349. [971]
- Hosking, J., Wallis, J., and Wood, E. (1985), "Estimation of the Generalized Extreme-Value Distribution by the Method of Probability-Weighted Moments," *Technometrics*, 27, 251–261. [971]
- Ji, J., and Li, D. (2021), "Application of Autoregressive Tail-Index Model to China's Stock Market," *Statistical Theory and Related Fields*, 5, 31–34. [968]
- Kelly, B., and Jiang, H. (2014), "Tail Risk and Asset Prices," *The Review of Financial Studies*, 27, 2841–2871. [968]
- Kleiber, C., and Kotz, S. (2003), *Statistical Size Distributions in Economics and Actuarial Sciences*, Hoboken, NJ: Wiley. [970]
- Laurini, F., and Tawn, J. A. (2009), "Regular Variation and Extremal Dependence of Garch Residuals with Application to Market Risk Measures," *Econometric Reviews*, 28, 146–169. [971]
- Lechner, J., Simiu, E., and Heckert, N. (1993), "Assessment of 'Peaks over Threshold' Methods for Estimating Extreme Value Distribution Tails," *Structural Safety*, 12, 305–314. [971]
- Li, F., Bicknell, C., Lowry, R., and Li, Y. (2012), "A Comparison of Extreme Wave Analysis Methods with 1994–2010 Offshore Perth Dataset," *Coastal Engineering*, 69, 1–11. [971]
- Lin, H., Liu, L., and Zhang, Z. (2023), "Tail Risk Signal Detection through a Novel egb2 Option Pricing Model," *Mathematics*, 11, 3194. [970,977]
- Lin, H., and Zhang, Z. (2022), "Extreme Co-movements between Infectious Disease Events and Crude Oil Futures Prices: From Extreme Value Analysis Perspective," *Energy Economics*, 5, 106054. [968]
- Lundbergh, S., Terasvirta, T., and Van Dijk, D. (2003), "Time-Varying Smooth Transition Autoregressive Models," *Journal of Business & Economic Statistics*, 21, 104–128. [970]
- Mao, G., and Zhang, Z. (2018), "Stochastic Tail Index Model for High Frequency Financial Data with Bayesian Analysis," *Journal of Econometrics*, 205, 470–487. [968]
- McDonald, J., and Xu, Y. (1995), "A Generalization of the Beta Distribution with Applications," *Journal of Econometrics*, 66, 133–152. [969,970]
- Papalexiou, S., and Koutsoyiannis, D. (2012), "Entropy based Derivation of Probability Distributions: A Case Study to Daily Rainfall," *Advances in Water Resources*, 45, 51–57. [970]
- Rootzen, H., and Finkenstadt, B. (2003), *Extreme Values in Finance, Telecommunications, and the Environment*, New York: Taylor & Francis Inc. [970]
- Smith, R. (1985), "Maximum Likelihood Estimation in a Class of Nonregular Cases," *Biometrika*, 72, 67–90. [971,972]
- Teyssi re, G., and Kirman, A. (2007), *Long Memory in Economics*, Berlin, Heidelberg: Springer. [970]
- Vuong, Q. H. (1989), "Likelihood Ratio Tests for Model Selection and Non-nested Hypotheses," *Econometrica*, 57, 307–333. [977]
- Yang, X., Frees, E., and Zhang, Z. (2011), "A Generalized Beta Copula with Applications in Modeling Multivariate Long-Tailed Data," *Insurance: Mathematics & Economics*, 49, 265–284. [969,970]
- Zhang, X., and Schwaab, B. (2017), "Tail Risk in Government Bond Markets and ECB Asset Purchases," working paper. [968]
- Zhang, Z. (2021a), "On Studying Extreme Values and Systematic Risks with Nonlinear Time Series Models and Tail Dependence Measures," *Statistical Theory and Related Fields*, 5, 1–25. [968]
- (2021b), "Rejoinder of 'On Studying Extreme Values and Systematic Risks with Nonlinear Time Series Models and Tail Dependence Measures,'" *Statistical Theory and Related Fields*, 5, 45–48. [968]
- Zhao, Z., Zhang, Z., and Chen, R. (2018), "Modeling Maxima with Autoregressive Conditional Fr chet Model," *Journal of Econometrics*, 207, 325–351. [968,970,971,974,976]
- Zheng, H., Hao, J., Bai, M., and Zhang, Z. (2019), "Valuation of Guaranteed Unitized Participating Life Insurance Under MEGB2 Distribution," *Discrete Dynamics in Nature and Society*, 2019, 1–16. [970]
- Zhou, C. (2010), "Existence and Consistency of the Maximum Likelihood Estimators for the Extreme Value Index," *Journal of Multivariate Analysis*, 101, 971–983. [971]

SUPPLEMENTARY MATERIAL

A Technical Proofs in the Paper

Notations and definitions The p.d.f. of the ‘GB2(a, b, p, q) distribution’ is:

$$\frac{ax^{ap-1}}{b^{ap}B(p, q)\{1 + (x/b)^a\}^{p+q}}, \quad \text{for } x > 0. \quad (\text{A.1})$$

The p.d.f. of the ‘prime beta(p, q) distribution’ is:

$$\frac{x^{p-1}}{B(p, q)(1+x)^{p+q}}, \quad \text{for } x > 0. \quad (\text{A.2})$$

Recalling from Section 3, the **AcGB2** model involves the parameter vector:

$$\boldsymbol{\theta} = (\mu, \alpha_0, \alpha_1, \alpha_2, \alpha_3, \beta_0, \beta_1, \beta_2, \beta_3, p, q)^\top = (\theta_1, \dots, \theta_{11})^\top$$

in the parameter set Θ defined in (7), where $\boldsymbol{\theta}^\circ = (\mu^\circ, \alpha_0^\circ, \alpha_1^\circ, \alpha_2^\circ, \alpha_3^\circ, \beta_0^\circ, \beta_1^\circ, \beta_2^\circ, \beta_3^\circ, p^\circ, q^\circ)^\top$ denotes the true value of $\boldsymbol{\theta}$. The Euclidean norm of $\boldsymbol{\theta}$ is defined by $\|\boldsymbol{\theta}\|$.

The identifiability of the **AcGB2** model is presented in Lemma 1, with its proof omitted here due to its straightforward nature.

Lemma 1 (Identifiability). *Under the conditions in Theorem 2, the **AcGB2** model is fully identifiable. That is, if $M_t(\boldsymbol{\theta}_1) = M_t(\boldsymbol{\theta}_2)$ almost surely for all t , then $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$, where $M_t(\boldsymbol{\theta})$ denotes M_t generated from (4)–(5) with parameter $\boldsymbol{\theta}$.*

In the following Lemma 2, we establish $E_{\boldsymbol{\theta}^\circ}(\frac{\partial l_t^\circ(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} |_{\boldsymbol{\theta}=\boldsymbol{\theta}^\circ}) = \mathbf{0}$ and the Fisher information matrix. The proof follows the lines of Lemma 4 in Zhao et al. (2018) and is omitted here. We note that due to the moment existing condition involves three parameters in the **AcGB2** model, the proofs of the remaining lemmas and theorems are non-trivial, compared to the proofs in Zhao et al. (2018).

Lemma 2. Assume the conditions in Theorem 2. Then $\mathbb{E}_{\boldsymbol{\theta}^\circ}(\frac{\partial l_t^\circ(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}|_{\boldsymbol{\theta}=\boldsymbol{\theta}^\circ}) = \mathbf{0}$ and

$$\mathbf{H}^\circ = \text{Var}_{\boldsymbol{\theta}^\circ}\left(\frac{\partial l_t^\circ(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^\circ}\right) = -\mathbb{E}_{\boldsymbol{\theta}^\circ}\left(\frac{\partial^2 l_t^\circ(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^\circ}\right),$$

in which $\mathbf{H}^\circ$ is a well-defined positive Fisher information matrix at $\boldsymbol{\theta}^\circ$.

For the sake of presentation, Lemmas 3–8 for proving Theorems 1–4 are relegated to Section A.5.

A.1 Proof of Theorem 1 (stationarity and ergodicity)

The proof utilizes Lemmas 1 and 2 in Appendix A, along with results of Chan and Tong (1994), and follows the lines of Zhao et al. (2018) with appropriate modifications due to the aforementioned moment existing conditions for **AcGB2** model. The lengthy details are excluded here but available upon request from the authors.

A.2 Proof of Theorem 2 (consistency)

Following the approach outlined in Smith (1985) for dealing with non-regular MLE, demonstrating that $L_n(\boldsymbol{\theta})$ has a local maximizer converging to $\boldsymbol{\theta}^\circ$ is equivalent to proving that the function

$$h_n(y, \mathbf{z}) = \frac{1}{\delta_n^2} L_n(\mu^\circ + \delta_n y, \boldsymbol{\psi}^\circ + \delta_n \mathbf{z})$$

has a local maximizer $(\widehat{y}, \widehat{\mathbf{z}})$ within the open set $\|(y, \mathbf{z}^\top)\| < 1$, where $y \in \mathbb{R}$, $\mathbf{z} \in \mathbb{R}^{10}$, and $\boldsymbol{\psi}^\circ = (\alpha_0^\circ, \alpha_1^\circ, \alpha_2^\circ, \alpha_3^\circ, \beta_0^\circ, \beta_1^\circ, \beta_2^\circ, \beta_3^\circ, p^\circ, q^\circ)^\top \in \mathbb{R}^{10}$ (corresponding to $\boldsymbol{\psi} = (\alpha_0, \alpha_1, \alpha_2, \alpha_3, \beta_0, \beta_1, \beta_2, \beta_3, p, q)^\top \in \mathbb{R}^{10}$ and $\boldsymbol{\theta} = (\mu, \boldsymbol{\psi})^\top$).

Given that $\mu^\circ + \delta_n \widehat{y} \xrightarrow{\text{P}} \mu^\circ$ and $\boldsymbol{\psi}^\circ + \delta_n \widehat{\mathbf{z}} \xrightarrow{\text{P}} \boldsymbol{\psi}^\circ$, where $\|(\delta_n \widehat{y}, \delta_n \widehat{\mathbf{z}}^\top)\| \leq \delta_n$, the proof is equivalent to demonstrating that for $\mathbf{x} = (y, \mathbf{z}^\top)^\top \in \mathbb{R}^{10}$,

$$\mathbf{x}^\top \frac{\partial h_n(\mathbf{x})}{\partial \mathbf{x}} = y \frac{\partial h_n(y, \mathbf{z})}{\partial y} + \sum_{j=1}^{10} z_j \frac{\partial h_n(y, \mathbf{z})}{\partial z_j} \quad (\text{A.3})$$

is negative when $\|\mathbf{x}\| = 1$ (i.e., $\|(y, \mathbf{z}^\top)\| = 1$).

Remark 11. *The proof aims to establish that the true parameter values (μ°, ψ°) act as a local maximizer of the likelihood function as the sample size n tends to the infinity. The proof relies on constructing the existence of δ_n such that (μ°, ψ°) falls within an open set to verify the negative value condition of (A.3). This open set converges to the true parameter values (μ°, ψ°) as the sample size tends to infinity.*

We first establish a proof that $h_n(\mathbf{x})$ possesses a local maximizer within $\|\mathbf{x}\| < 1$. Given the continuity and differentiability of $h_n(\mathbf{x})$, it must have at least one (local) maximizer within the closed set $\|\mathbf{x}\| \leq 1$. Assuming $\mathbf{x}^*$ is such a maximizer where $\|\mathbf{x}^*\| = 1$ and lies on the boundary, let us consider a small $\epsilon > 0$ such that $\mathbf{x}^{**} = (1 - \epsilon)\mathbf{x}^*$ resides within $\|\mathbf{x}\| < 1$. Utilizing Taylor expansion, we derive:

$$h_n(\mathbf{x}^{**}) = h_n(\mathbf{x}^*) - \epsilon * \mathbf{x}^{*T} \frac{\partial h_n(\mathbf{x}^*)}{\partial \mathbf{x}^*} + R_n > h_n(\mathbf{x}^*)$$

as $-\epsilon * \mathbf{x}^{*T} \frac{\partial h_n(\mathbf{x}^*)}{\partial \mathbf{x}^*}$ is positive. Therefore, there exists at least one local maximizer within $\|\mathbf{x}\| < 1$.

We now address the term $\frac{\partial h_n(y, \mathbf{z})}{\partial y}$ in (A.3). Considering $\mu = \mu^\circ + \delta_n y$ and $\psi = \psi^\circ + \delta_n \mathbf{z}$, we obtain:

$$\frac{\partial h_n(y, \mathbf{z})}{\partial y} = \frac{1}{\delta_n} \frac{\partial L_n(\mu, \psi^\circ + \delta_n \mathbf{z})}{\partial \mu} \Big|_{\mu=\mu^\circ + \delta_n y}.$$

This function, combined with the Taylor expansion at (μ°, ψ°) , becomes:

$$\begin{aligned} & \frac{1}{\delta_n} \frac{\partial L_n(\mu, \psi^\circ + \delta_n \mathbf{z})}{\partial \mu} \Big|_{\mu=\mu^\circ + \delta_n y} \\ = & \frac{1}{\delta_n} \frac{\partial L_n(\mu, \psi^\circ)}{\partial \mu} \Big|_{\mu=\mu^\circ} + \frac{1}{\delta_n} \frac{\partial^2 L_n(\mu, \psi^c)}{\partial \mu^2} \Big|_{\mu=\mu^c} (\mu - \mu^\circ) \\ & + \frac{1}{\delta_n} \sum_{j=1}^{10} \frac{\partial^2 L_n(\mu, \psi)}{\partial \mu \partial \psi_j} \Big|_{\mu=\mu^c, \psi=\psi^c} (\psi_j - \psi_j^\circ) \\ = & \frac{1}{\delta_n} \frac{\partial L_n(\mu, \psi^\circ)}{\partial \mu} \Big|_{\mu=\mu^\circ} + \frac{\partial^2 L_n(\mu, \psi^c)}{\partial \mu^2} \Big|_{\mu=\mu^c} y + \sum_{j=1}^{10} \frac{\partial^2 L_n(\mu, \psi)}{\partial \mu \partial \psi_j} \Big|_{\mu=\mu^c, \psi=\psi^c} z_j \end{aligned}$$

$$= \frac{1}{\delta_n} \frac{\partial L_n(\mu, \boldsymbol{\psi}^\circ)}{\partial \mu} \Big|_{\mu=\mu^\circ} - \frac{1}{\delta_n} \frac{\partial L_n^\circ(\mu, \boldsymbol{\psi}^\circ)}{\partial \mu} \Big|_{\mu=\mu^\circ} \quad (\text{A.4})$$

$$+ \frac{1}{\delta_n} \frac{\partial L_n^\circ(\mu, \boldsymbol{\psi}^\circ)}{\partial \mu} \Big|_{\mu=\mu^\circ} \quad (\text{A.5})$$

$$+ \frac{\partial^2 L_n(\mu, \boldsymbol{\psi}^c)}{\partial \mu^2} \Big|_{\mu=\mu^c} y + \sum_{j=1}^{10} \frac{\partial^2 L_n(\mu, \boldsymbol{\psi})}{\partial \mu \partial \psi_j} \Big|_{\mu=\mu^c, \boldsymbol{\psi}=\boldsymbol{\psi}^c} z_j, \quad (\text{A.6})$$

where $\boldsymbol{\psi}^c = (\psi_1^c, \dots, \psi_{10}^c)^\top$, $|\mu^c - \mu^\circ| < \delta_n y$, and $|\psi_j^c - \psi_j^\circ| < \delta_n z_j$ for $j = 1, \dots, 10$.

Additionally, $L_n^\circ(\cdot)$ represents the log-likelihood with true initial values $\{A'_1, B'_1\}$.

Next, we aim to demonstrate the convergence of (A.6) as follows: In Part (i), (A.6) $\xrightarrow{P} -y m_{\mu, \mu}(\boldsymbol{\theta}^\circ) - \sum_{j=1}^{10} m_{\mu, \psi_j}(\boldsymbol{\theta}^\circ) z_j$ over $\|(y, \mathbf{z}^\top)\| \leq 1$. In Part (ii), (A.4) $\xrightarrow{P} 0$, and in Part (iii), (A.5) $\xrightarrow{P} 0$. **Part (i).** Starting with the first term in (A.6), we will first prove:

$$\frac{\partial^2 L_n(\mu, \boldsymbol{\psi}^c)}{\partial \mu^2} \Big|_{\mu=\mu^c} \xrightarrow{P} -m_{\mu, \mu}(\boldsymbol{\theta}^\circ), \quad (\text{A.7})$$

where $m_{\theta_j, \theta_k}(\boldsymbol{\theta}^\circ) = -E_{\boldsymbol{\theta}^\circ}(\frac{\partial^2 l_t^\circ(\boldsymbol{\theta})}{\partial \theta_j \partial \theta_k} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^\circ})$. By the law of large numbers, we have $\frac{\partial^2 L_n^\circ(\boldsymbol{\theta})}{\partial \mu^2} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^\circ} \xrightarrow{P} -m_{\mu, \mu}(\boldsymbol{\theta}^\circ)$. According to Lemma 2, $E_{\boldsymbol{\theta}^\circ}(\frac{\partial l_t^\circ(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^\circ}) = \mathbf{0}$ and $\text{Var}_{\boldsymbol{\theta}^\circ}(\frac{\partial l_t^\circ(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^\circ}) = -E_{\boldsymbol{\theta}^\circ}(\frac{\partial^2 l_t^\circ(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^\circ})$.

It suffices to prove:

$$\frac{\partial^2 L_n(\boldsymbol{\theta})}{\partial \mu^2} - \frac{\partial^2 L_n^\circ(\boldsymbol{\theta})}{\partial \mu^2} \xrightarrow{P} 0, \quad (\text{A.8})$$

$$\frac{\partial^2 L_n^\circ(\boldsymbol{\theta})}{\partial \mu^2} - \frac{\partial^2 L_n^\circ(\boldsymbol{\theta})}{\partial \mu^2} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^\circ} \xrightarrow{P} 0, \quad (\text{A.9})$$

for $\|\boldsymbol{\theta} - \boldsymbol{\theta}^\circ\| \leq \delta_n$.

To establish (A.9), a direct calculation from (10) yields:

$$\begin{aligned} \frac{\partial L_n^\circ(\boldsymbol{\theta})}{\partial \mu} &= \frac{1}{n} \sum_{t=1}^n \left\{ (-1)(A'_t p - 1) \frac{1}{M_t - \mu} - (p + q) \frac{1}{1 + (\frac{M_t - \mu}{B'_t})^{A'_t}} \frac{A'_t}{-B'_t} \left(\frac{M_t - \mu}{B'_t} \right)^{A'_t - 1} \right\}, \\ \frac{\partial^2 L_n^\circ(\boldsymbol{\theta})}{\partial \mu^2} &= \frac{1}{n} \sum_{t=1}^n \left[(-1)(A'_t p - 1)(M_t - \mu)^{-2} \right. \\ &\quad + (p + q) \frac{A'_t (1 - A'_t)}{(B'_t)^2} \left(\frac{M_t - \mu}{B'_t} \right)^{A'_t - 2} \left\{ 1 + \left(\frac{M_t - \mu}{B'_t} \right)^{A'_t} \right\}^{-1} \\ &\quad \left. + (p + q) \frac{(A'_t)^2}{(B'_t)^2} \left(\frac{M_t - \mu}{B'_t} \right)^{2A'_t - 2} \left\{ 1 + \left(\frac{M_t - \mu}{B'_t} \right)^{A'_t} \right\}^{-2} \right]. \end{aligned}$$

Then,

$$\frac{\partial^2 L_n^\circ(\boldsymbol{\theta})}{\partial \mu^2} - \frac{\partial^2 L_n^\circ(\boldsymbol{\theta})}{\partial \mu^2} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^\circ} = I_1 + I_2 + I_3, \quad (\text{A.10})$$

with the first term

$$I_1 = \frac{1}{n} \sum_{t=1}^n (-1) \left[(A'_t p - 1)(M_t - \mu)^{-2} - (A'^{\circ}_t p^{\circ} - 1)(M_t - \mu^{\circ})^{-2} \right],$$

the second term I_2 equal to

$$\begin{aligned} & (p + q) \frac{1}{n} \sum_{t=1}^n \left[\frac{A'_t (1 - A'_t)}{(B'_t)^2} \left(\frac{M_t - \mu}{B'_t} \right)^{A'_t - 2} \left\{ 1 + \left(\frac{M_t - \mu}{B'_t} \right)^{A'_t} \right\}^{-1} \right] \\ & - (p^{\circ} + q^{\circ}) \frac{1}{n} \sum_{t=1}^n \left[\frac{A'^{\circ}_t (1 - A'^{\circ}_t)}{(B'^{\circ}_t)^2} \left(\frac{M_t - \mu^{\circ}}{B'^{\circ}_t} \right)^{A'^{\circ}_t - 2} \left\{ 1 + \left(\frac{M_t - \mu^{\circ}}{B'^{\circ}_t} \right)^{A'^{\circ}_t} \right\}^{-1} \right], \end{aligned}$$

and the third term I_3 equal to

$$\begin{aligned} & (p + q) \frac{1}{n} \sum_{t=1}^n \left[\frac{(A'_t)^2}{(B'_t)^2} \left(\frac{M_t - \mu}{B'_t} \right)^{2A'_t - 2} \left\{ 1 + \left(\frac{M_t - \mu}{B'_t} \right)^{A'_t} \right\}^{-2} \right] \\ & - (p^{\circ} + q^{\circ}) \frac{1}{n} \sum_{t=1}^n \left[\frac{(A'^{\circ}_t)^2}{(B'^{\circ}_t)^2} \left(\frac{M_t - \mu^{\circ}}{B'^{\circ}_t} \right)^{2A'^{\circ}_t - 2} \left\{ 1 + \left(\frac{M_t - \mu^{\circ}}{B'^{\circ}_t} \right)^{A'^{\circ}_t} \right\}^{-2} \right]. \end{aligned}$$

Here,

$\{A_t, B_t\}$ are computed with $\boldsymbol{\theta}$ and arbitrary initial values $\{A_1, B_1\}$;

$\{A'_t, B'_t\}$ are computed with $\boldsymbol{\theta}$ and true initial values $\{A'_1, B'_1\}$;

$\{A^{\circ}_t, B^{\circ}_t\}$ are computed with $\boldsymbol{\theta}^{\circ}$ and arbitrary initial values $\{A_1, B_1\}$;

while $\{A'^{\circ}_t, B'^{\circ}_t\}$ are computed with $\boldsymbol{\theta}^{\circ}$ and true initial values $\{A'_1, B'_1\}$.

Next, we examine the limiting behavior of I_1 , I_2 , and I_3 individually.

First, starting with I_1 in (A.10),

$$|I_1| \leq \frac{1}{n} \sum_{t=1}^n |(A'_t p - 1)(M_t - \mu)^{-2} - (A'_t p - 1)(M_t - \mu^{\circ})^{-2}| \quad (\text{A.11})$$

$$+ \frac{1}{n} \sum_{t=1}^n |(A'_t p - 1)(M_t - \mu^{\circ})^{-2} - (A'^{\circ}_t p^{\circ} - 1)(M_t - \mu^{\circ})^{-2}|. \quad (\text{A.12})$$

We have (A.11) = $\frac{1}{n} \sum_{t=1}^n |A'_t p - 1| \cdot |(M_t - \mu)^{-2} - (M_t - \mu^{\circ})^{-2}|$. According to Lemma 4,

(A.11) $\xrightarrow{P} 0$. For (A.12) = $\frac{1}{n} \sum_{t=1}^n |(M_t - \mu^{\circ})^{-2}| \cdot |A'_t p - A'^{\circ}_t p^{\circ}|$. Here,

$$\begin{aligned} |A'_t p - A'^{\circ}_t p^{\circ}| & \leq |A'_t p - A'^{\circ}_t p| + |A'_t p - A'^{\circ}_t p^{\circ}| \\ & = p |A'_t - A'^{\circ}_t| + A'^{\circ}_t |p - p^{\circ}|. \end{aligned} \quad (\text{A.13})$$

By Lemma 5, (A.13) $\xrightarrow{P} 0$. Consequently, by Lemma 3, (A.12) $\xrightarrow{P} 0$. Hence, $I_1 \xrightarrow{P} 0$.

Second, for I_2 in (A.10),

$$|I_2| \leq \left| (p+q) \frac{1}{n} \sum_{t=1}^n \left[\frac{A'_t(1-A'_t)}{(B'_t)^2} \left(\frac{M_t - \mu}{B'_t} \right)^{A'_t-2} \left\{ 1 + \left(\frac{M_t - \mu}{B'_t} \right)^{A'_t} \right\}^{-1} \right] \right. \quad (\text{A.14})$$

$$\left. - (p^\circ + q^\circ) \frac{1}{n} \sum_{t=1}^n \left[\frac{A'_t(1-A'_t)}{(B'_t)^2} \left(\frac{M_t - \mu}{B'_t} \right)^{A'_t-2} \left\{ 1 + \left(\frac{M_t - \mu}{B'_t} \right)^{A'_t} \right\}^{-1} \right] \right|$$

$$+ \left| (p^\circ + q^\circ) \frac{1}{n} \sum_{t=1}^n \left[\frac{A'_t(1-A'_t)}{(B'_t)^2} \left(\frac{M_t - \mu}{B'_t} \right)^{A'_t-2} \left\{ 1 + \left(\frac{M_t - \mu}{B'_t} \right)^{A'_t} \right\}^{-1} \right] \right. \quad (\text{A.15})$$

$$\left. - (p^\circ + q^\circ) \frac{1}{n} \sum_{t=1}^n \left[\frac{A'^\circ_t(1-A'^\circ_t)}{(B'^\circ_t)^2} \left(\frac{M_t - \mu^\circ}{B'^\circ_t} \right)^{A'^\circ_t-2} \left\{ 1 + \left(\frac{M_t - \mu^\circ}{B'^\circ_t} \right)^{A'^\circ_t} \right\}^{-1} \right] \right|.$$

For (A.14), we have

$$(\text{A.14}) = |(p+q) - (p^\circ + q^\circ)| \cdot |I_{2,1}|,$$

where $I_{2,1} = \frac{1}{n} \sum_{t=1}^n \left[\frac{A'_t(1-A'_t)}{(B'_t)^2} \left(\frac{M_t - \mu}{B'_t} \right)^{A'_t-2} \left\{ 1 + \left(\frac{M_t - \mu}{B'_t} \right)^{A'_t} \right\}^{-1} \right]$. Note $|(p+q) - (p^\circ + q^\circ)| \rightarrow 0$,

since $|(p+q) - (p^\circ + q^\circ)| \leq |p - p^\circ| + |q - q^\circ|$. Also, we have $|I_{2,1}| \leq \left| \frac{1}{n} \sum_{t=1}^n \left[\frac{A'_t(1-A'_t)}{(B'_t)^2} \left(\frac{M_t - \mu^\circ}{B'_t} \right)^{A'_t-2} \right] \right|$.

By Lemma 3, $|I_{2,1}|$ is finite. So (A.14) $\xrightarrow{P} 0$. For (A.15),

$$|(\text{A.15})|/(p^\circ + q^\circ)$$

$$\leq \frac{1}{n} \sum_{t=1}^n \left| \frac{A'_t(1-A'_t)}{(B'_t)^2} \right| \cdot \left| \left[\left(\frac{M_t - \mu}{B'_t} \right)^{A'_t-2} \left\{ 1 + \left(\frac{M_t - \mu}{B'_t} \right)^{A'_t} \right\}^{-1} \right. \right. \quad (\text{A.16})$$

$$\left. - \left(\frac{M_t - \mu^\circ}{B'_t} \right)^{A'_t-2} \left\{ 1 + \left(\frac{M_t - \mu^\circ}{B'_t} \right)^{A'_t} \right\}^{-1} \right] \right|$$

$$+ \frac{1}{n} \sum_{t=1}^n \left| \frac{A'_t(1-A'_t)}{(B'_t)^2} \right| \cdot \left| \left[\left(\frac{M_t - \mu^\circ}{B'_t} \right)^{A'_t-2} \left\{ 1 + \left(\frac{M_t - \mu^\circ}{B'_t} \right)^{A'_t} \right\}^{-1} \right. \right. \quad (\text{A.17})$$

$$\left. - \left(\frac{M_t - \mu^\circ}{B'^\circ_t} \right)^{A'^\circ_t-2} \left\{ 1 + \left(\frac{M_t - \mu^\circ}{B'^\circ_t} \right)^{A'^\circ_t} \right\}^{-1} \right] \right|$$

$$+ \frac{1}{n} \sum_{t=1}^n \left| \frac{A'_t(1-A'_t)}{(B'_t)^2} - \frac{A'^\circ_t(1-A'^\circ_t)}{(B'^\circ_t)^2} \right| \left(\frac{M_t - \mu^\circ}{B'^\circ_t} \right)^{A'^\circ_t-2} \left\{ 1 + \left(\frac{M_t - \mu^\circ}{B'^\circ_t} \right)^{A'^\circ_t} \right\}^{-1} \quad (\text{A.18})$$

We next analyze the three parts: (A.16), (A.17) and (A.18) separately. For (A.16), we have

$$(\text{A.16}) \leq I_{3,1} + I_{3,2}, \quad (\text{A.19})$$

where

$$\begin{aligned}
I_{3,1} &= \frac{1}{n} \sum_{t=1}^n \left| \frac{A'_t(1-A'_t)}{(B'_t)^2} \right| \cdot \left| \frac{\left(\frac{M_t-\mu}{B'_t}\right)^{A'_t-2}}{1 + \left(\frac{M_t-\mu}{B'_t}\right)^{A'_t}} - \frac{\left(\frac{M_t-\mu}{B'_t}\right)^{A'_t-2}}{1 + \left(\frac{M_t-\mu^\circ}{B'_t}\right)^{A'_t}} \right|, \\
I_{3,2} &= \frac{1}{n} \sum_{t=1}^n \left| \frac{A'_t(1-A'_t)}{(B'_t)^2} \right| \cdot \left| \frac{\left(\frac{M_t-\mu}{B'_t}\right)^{A'_t-2} - \left(\frac{M_t-\mu^\circ}{B'_t}\right)^{A'_t-2}}{1 + \left(\frac{M_t-\mu^\circ}{B'_t}\right)^{A'_t}} \right|.
\end{aligned}$$

Note that

$$\begin{aligned}
I_{3,1} &\leq \frac{1}{n} \sum_{t=1}^n \left| \frac{A'_t(1-A'_t)}{(B'_t)^2} \right| \cdot \left| \frac{\left(\frac{M_t-\mu}{B'_t}\right)^{A'_t-2}}{\left(\frac{M_t-\mu}{B'_t}\right)^{A'_t}} - \frac{\left(\frac{M_t-\mu}{B'_t}\right)^{A'_t-2}}{\left(\frac{M_t-\mu^\circ}{B'_t}\right)^{A'_t}} \right| \\
&= \frac{1}{n} \sum_{t=1}^n |(1-A'_t)A'_t| (M_t-\mu)^{A'_t-2} \left| \frac{(M_t-\mu^\circ)^{A'_t} - (M_t-\mu)^{A'_t}}{(M_t-\mu)^{A'_t}(M_t-\mu^\circ)^{A'_t}} \right| \\
&= \frac{1}{n} \sum_{t=1}^n |(1-A'_t)A'_t| (M_t-\mu)^{-2} (M_t-\mu^\circ)^{-A'_t} |(M_t-\mu^\circ)^{A'_t} - (M_t-\mu)^{A'_t}|.
\end{aligned}$$

By Lemma 4, Lemma 7, and Hölder inequality, $I_{3,1} \xrightarrow{P} 0$.

Since $I_{3,2} = \frac{1}{n} \sum_{t=1}^n |A'_t(1-A'_t)(B'_t)^{-A'_t}| \frac{1}{1 + \left(\frac{M_t-\mu^\circ}{B'_t}\right)^{A'_t}} |(M_t-\mu)^{A'_t-2} - (M_t-\mu^\circ)^{A'_t-2}|$, by Lemma 4 and Lemma 3, $I_{3,2} \xrightarrow{P} 0$. So (A.16) $\xrightarrow{P} 0$. For (A.17),

$$(A.17) \leq I_{4,1} + I_{4,2}, \quad (A.20)$$

where

$$\begin{aligned}
I_{4,1} &= \frac{1}{n} \sum_{t=1}^n \left| \frac{A'_t(1-A'_t)}{(B'_t)^2} \right| \cdot \left| \frac{\left(\frac{M_t-\mu^\circ}{B'_t}\right)^{A'_t-2}}{1 + \left(\frac{M_t-\mu^\circ}{B'_t}\right)^{A'_t}} - \frac{\left(\frac{M_t-\mu^\circ}{B'_t}\right)^{A'_t-2}}{1 + \left(\frac{M_t-\mu^\circ}{B'_t}\right)^{A'_t}} \right|, \\
I_{4,2} &= \frac{1}{n} \sum_{t=1}^n \left| \frac{A'_t(1-A'_t)}{(B'_t)^2} \right| \cdot \left| \frac{\left(\frac{M_t-\mu^\circ}{B'_t}\right)^{A'_t-2} - \left(\frac{M_t-\mu^\circ}{B'_t}\right)^{A'_t-2}}{1 + \left(\frac{M_t-\mu^\circ}{B'_t}\right)^{A'_t}} \right|.
\end{aligned}$$

and

$$\begin{aligned}
I_{4,1} &\leq \frac{1}{n} \sum_{t=1}^n \left| \frac{A'_t(1-A'_t)}{(B'_t)^2} \right| \cdot \left| \frac{\left(\frac{M_t-\mu^\circ}{B'_t}\right)^{A'_t-2}}{\left(\frac{M_t-\mu^\circ}{B'_t}\right)^{A'_t}} - \frac{\left(\frac{M_t-\mu^\circ}{B'_t}\right)^{A'_t-2}}{\left(\frac{M_t-\mu^\circ}{B'_t}\right)^{A'_t}} \right| \\
&= \frac{1}{n} \sum_{t=1}^n |A'_t(1-A'_t)(B'_t)^{-A'_t}| (M_t-\mu^\circ)^{A'_t-2} \left| \frac{\left(\frac{M_t-\mu^\circ}{B'_t}\right)^{A'_t} - \left(\frac{M_t-\mu^\circ}{B'_t}\right)^{A'_t}}{\left(\frac{M_t-\mu^\circ}{B'_t}\right)^{A'_t} \left(\frac{M_t-\mu^\circ}{B'_t}\right)^{A'_t}} \right| \\
&= \frac{1}{n} \sum_{t=1}^n |A'_t(1-A'_t)(B'_t)^{A'_t}| (M_t-\mu^\circ)^{-2-A'_t} |(B'_t)^{-A'_t} (M_t-\mu^\circ)^{A'_t} - (B'_t)^{-A'_t} (M_t-\mu^\circ)^{A'_t}| \\
&\leq \frac{1}{n} \sum_{t=1}^n |A'_t(1-A'_t)(B'_t)^{A'_t}| (M_t-\mu^\circ)^{-2-A'_t} |(B'_t)^{-A'_t}| \cdot |(M_t-\mu^\circ)^{A'_t} - (M_t-\mu^\circ)^{A'_t}|
\end{aligned}$$

$$\begin{aligned}
& + \frac{1}{n} \sum_{t=1}^n |A'_t(1 - A'_t)(B'_t)^{A'_t}| (M_t - \mu^\circ)^{-2-A'_t} |(B'_t)^{-A'_t} - (B'_t)^{-A'_t}| \cdot |(M_t - \mu^\circ)^{A'_t}| \\
& = \frac{1}{n} \sum_{t=1}^n |A'_t(1 - A'_t)| (M_t - \mu^\circ)^{-2-A'_t} |(M_t - \mu^\circ)^{A'_t} - (M_t - \mu^\circ)^{A'_t}| \\
& \quad + \frac{1}{n} \sum_{t=1}^n |A'_t(1 - A'_t)(B'_t)^{A'_t}| (M_t - \mu^\circ)^{A'_t-2-A'_t} |(B'_t)^{-A'_t} - (B'_t)^{-A'_t}|.
\end{aligned}$$

By Lemma 7, Lemma 5 and Lemma 3, $I_{4,1} \xrightarrow{P} 0$. Also, $I_{4,2} \xrightarrow{P} 0$ follows from

$$\begin{aligned}
I_{4,2} &= \frac{1}{n} \sum_{t=1}^n \left| \frac{A'_t(1 - A'_t)}{(B'_t)^2} \right| \frac{1}{1 + \left(\frac{M_t - \mu^\circ}{B'_t} \right)^{A'_t}} (B'_t)^{2-A'_t} |(M_t - \mu^\circ)^{A'_t-2} - (M_t - \mu^\circ)^{A'_t-2}| \\
& \quad + \frac{1}{n} \sum_{t=1}^n \left| \frac{A'_t(1 - A'_t)}{(B'_t)^2} \right| \frac{1}{1 + \left(\frac{M_t - \mu^\circ}{B'_t} \right)^{A'_t}} (M_t - \mu^\circ)^{A'_t-2} |(B'_t)^{2-A'_t} - (B'_t)^{2-A'_t}|,
\end{aligned}$$

and Lemma 7, Lemma 5 and Lemma 3. So (A.17) $\xrightarrow{P} 0$. Also, (A.18) $\xrightarrow{P} 0$ follows from

$$(A.18) \leq \frac{1}{n} \sum_{t=1}^n \left| \frac{A'_t(1 - A'_t)}{(B'_t)^2} - \frac{A'_t(1 - A'_t)}{(B'_t)^2} \right| \cdot \left(\frac{M_t - \mu^\circ}{B'_t} \right)^{-2}, \quad (A.21)$$

Lemma 5 and Lemma 3. Putting all things together, we have $I_2 \xrightarrow{P} 0$.

Third, for I_3 in (A.10), we can prove $I_3 \xrightarrow{P} 0$ using arguments similar to those for I_2 .

So, we have proved (A.9). Similar to (A.9), by Lemma 8 and Lemma 6 (instead of Lemma 7 and Lemma 5), we can prove (A.8).

For the second term in (A.6), we can use similar arguments as above to prove

$$\left. \frac{\partial^2 L_n(\mu, \psi)}{\partial \mu \partial \psi_j} \right|_{\mu=\mu^\circ, \psi=\psi^\circ} \xrightarrow{P} -m_{\mu, \psi_j}(\boldsymbol{\theta}^\circ), \quad j = 1, \dots, 10. \quad (A.22)$$

Combining (A.7) and (A.22), we obtain (A.6) $\xrightarrow{P} -m_{\mu, \mu}(\boldsymbol{\theta}^\circ)y - \sum_{j=1}^{10} m_{\mu, \psi_j}(\boldsymbol{\theta}^\circ)z_j$, over

$\|(y, \mathbf{z}^\top)\| \leq 1$. **Part (ii).** Next we want to prove (A.4) $\xrightarrow{P} 0$. Note

$$(A.4) = \frac{1}{\delta_n} \frac{1}{n} \sum_{t=1}^n \left[(1 - A_t^\circ p) \frac{1}{M_t - \mu^\circ} - (1 - A_t^\circ p) \frac{1}{M_t - \mu^\circ} \right] \quad (A.23)$$

$$\begin{aligned}
& + \frac{1}{\delta_n n} \sum_{t=1}^n (p + q) \left[\frac{A_t^\circ}{B_t^\circ} \left(\frac{M_t - \mu^\circ}{B_t^\circ} \right)^{A_t^\circ-1} \left\{ 1 + \left(\frac{M_t - \mu^\circ}{B_t^\circ} \right)^{A_t^\circ} \right\}^{-1} \right. \\
& \quad \left. - \frac{A_t^\circ}{B_t^\circ} \left(\frac{M_t - \mu^\circ}{B_t^\circ} \right)^{A_t^\circ-1} \left\{ 1 + \left(\frac{M_t - \mu^\circ}{B_t^\circ} \right)^{A_t^\circ} \right\}^{-1} \right]. \quad (A.24)
\end{aligned}$$

For (A.23), note that $|(A.23)| = \frac{1}{\delta_n} \frac{1}{n} \sum_{t=1}^n |A_t^\circ - A_t^\circ| \frac{p}{M_t - \mu^\circ}$. By Lemma 6, $|(A.23)| \leq \frac{C}{\delta_n n} \sum_{t=1}^n U_{r,b}^{t-1} \frac{1}{M_t - \mu^\circ}$. This gives (A.23) $\xrightarrow{P} 0$, since $\sum_{t=1}^n U_{r,b}^{t-1} \frac{1}{M_t - \mu^\circ}$ is finite.

To justify (A.24) $\xrightarrow{P} 0$, we observe that

$$|(A.24)| \leq \frac{p+q}{n\delta_n} \sum_{t=1}^n \frac{A_t^{\prime o}}{B_t^{\prime o}} \left| \frac{\left(\frac{M_t - \mu^o}{B_t^{\prime o}}\right)^{A_t^{\prime o} - 1}}{1 + \left(\frac{M_t - \mu^o}{B_t^{\prime o}}\right)^{A_t^{\prime o}}} - \frac{\left(\frac{M_t - \mu^o}{B_t^o}\right)^{A_t^o - 1}}{1 + \left(\frac{M_t - \mu^o}{B_t^o}\right)^{A_t^o}} \right| \quad (A.25)$$

$$+ \frac{p+q}{n\delta_n} \sum_{t=1}^n \left| \frac{A_t^{\prime o}}{B_t^{\prime o}} - \frac{A_t^o}{B_t^o} \right| \frac{\left(\frac{M_t - \mu^o}{B_t^o}\right)^{A_t^o - 1}}{1 + \left(\frac{M_t - \mu^o}{B_t^o}\right)^{A_t^o}}. \quad (A.26)$$

We now analyze (A.25) and (A.26) separately. For (A.25), we have

$$(A.25) \leq I_{5,1} + I_{5,2} + I_{5,3} + I_{5,4}, \quad (A.27)$$

where

$$\begin{aligned} I_{5,1} &= \frac{p+q}{n\delta_n} \sum_{t=1}^n A_t^{\prime o} B_t^{\prime o} (M_t - \mu^o)^{-1-A_t^o} |(M_t - \mu^o)^{A_t^o} - (M_t - \mu^o)^{A_t^{\prime o}}|, \\ I_{5,2} &= \frac{p+q}{n\delta_n} \sum_{t=1}^n A_t^{\prime o} B_t^{\prime o} (B_t^o)^{A_t^o} (M_t - \mu^o)^{A_t^{\prime o} - 1 - A_t^o} |(B_t^o)^{-A_t^o} - (B_t^{\prime o})^{-A_t^{\prime o}}|, \\ I_{5,3} &= \frac{p+q}{n\delta_n} \sum_{t=1}^n A_t^{\prime o} B_t^{\prime o} \frac{1}{1 + \left(\frac{M_t - \mu^o}{B_t^o}\right)^{A_t^o}} (B_t^{\prime o})^{1-A_t^{\prime o}} |(M_t - \mu^o)^{A_t^{\prime o} - 1} - (M_t - \mu^o)^{A_t^o - 1}|, \\ I_{5,4} &= \frac{p+q}{n\delta_n} \sum_{t=1}^n A_t^{\prime o} B_t^{\prime o} \frac{1}{1 + \left(\frac{M_t - \mu^o}{B_t^o}\right)^{A_t^o}} (M_t - \mu^o)^{A_t^o - 1} |(B_t^{\prime o})^{1-A_t^{\prime o}} - (B_t^o)^{1-A_t^o}|. \end{aligned}$$

For $I_{5,1}$, by the mean value theorem:

$$I_{5,1} \leq \frac{p+q}{n\delta_n} \sum_{t=1}^n |A_t^{\prime o} B_t^{\prime o}| (M_t - \mu^o)^{-1-A_t^o} (M_t - \mu^o)^{A_t^*} |\log(M_t - \mu^o)| \cdot |A_t^o - A_t^{\prime o}|, \quad (A.28)$$

where A_t^* is located between A_t^o and $A_t^{\prime o}$. By Lemma 6, (A.28) gives

$$I_{5,1} \leq \frac{C(p+q)}{n\delta_n} \sum_{t=1}^n |A_t^{\prime o} B_t^{\prime o}| (M_t - \mu^o)^{-1-A_t^o} (M_t - \mu^o)^{A_t^*} |\log(M_t - \mu^o)| U_{r,b}^{t-1}.$$

Since $\sum_{t=1}^n A_t^{\prime o} B_t^{\prime o} (M_t - \mu^o)^{-1-A_t^o} (M_t - \mu^o)^{A_t^*} |\log(M_t - \mu^o)| U_{r,b}^{t-1}$ is finite, as $n\delta_n \rightarrow \infty$, we have that $I_{5,1} \xrightarrow{P} 0$. For $I_{5,2}$,

$$I_{5,2} \leq \frac{C(p+q)}{n\delta_n} \sum_{t=1}^n |A_t^{\prime o} B_t^{\prime o} (B_t^o)^{A_t^o}| (M_t - \mu^o)^{A_t^{\prime o} - 1 - A_t^o} U_{r,b}^{t-1}.$$

Since $\sum_{t=1}^n |A_t^{\prime o} B_t^{\prime o} (B_t^o)^{A_t^o}| (M_t - \mu^o)^{A_t^{\prime o} - 1 - A_t^o} U_{r,b}^{t-1}$ is finite, we have $I_{5,2} \xrightarrow{P} 0$. For $I_{5,3}$ and $I_{5,4}$, they are similar to $I_{5,1}$ and $I_{5,2}$. We have that $I_{5,3} \xrightarrow{P} 0$, and $I_{5,4} \xrightarrow{P} 0$. Using (A.27), we get (A.25) $\xrightarrow{P} 0$.

As for (A.26),

$$(A.26) \leq \frac{p+q}{n\delta_n} \sum_{t=1}^n \left| \frac{A'_t}{B'_t} - \frac{A_t^\circ}{B_t^\circ} \right| \left(\frac{M_t - \mu^\circ}{B_t^\circ} \right)^{-1} \leq \frac{C(p+q)}{n\delta_n} \sum_{t=1}^n \left(\frac{M_t - \mu^\circ}{B_t^\circ} \right)^{-1} U_{r,b}^{t-1}.$$

Since $\sum_{t=1}^n (\frac{M_t - \mu^\circ}{B_t^\circ})^{-1} U_{r,b}^{t-1}$ is finite, we have (A.26) $\xrightarrow{P} 0$. **Part (iii).** Recall that

$$(A.5) = \frac{1}{\delta_n} \frac{\partial L_n^\circ(\mu, \psi^\circ)}{\partial \mu} \Big|_{\mu=\mu^\circ} = \frac{1}{\delta_n n} \sum_{t=1}^n \frac{\partial l_t^\circ(\mu, \psi^\circ)}{\partial \mu} \Big|_{\mu=\mu^\circ}. \quad (A.29)$$

(A.29) together with $E_{\theta^\circ}(\frac{\partial l_t^\circ(\theta)}{\partial \theta} |_{\theta=\theta^\circ}) = \mathbf{0}$ and $\delta_n n \rightarrow \infty$ imply (A.5) $\xrightarrow{P} 0$.

Combining results of (A.4), (A.5) and (A.6) obtained in Parts (i)–(iii) above, we obtain

$$\frac{\partial h_n(y, \mathbf{z})}{\partial y} \xrightarrow{P} -m_{\mu, \mu}(\theta^\circ) y - \sum_{j=1}^{10} m_{\mu, \psi_j}(\theta^\circ) z_j, \quad (A.30)$$

over $\|(y, \mathbf{z}^\top)\| \leq 1$. Similarly, for the other term $\frac{\partial h_n(y, \mathbf{z})}{\partial z_j}$ in (A.3), we can show

$$\frac{\partial h_n(y, \mathbf{z})}{\partial z_j} \xrightarrow{P} -m_{\mu, \psi_j}(\theta^\circ) y - \sum_{k=1}^{10} m_{\psi_j, \psi_k}(\theta^\circ) z_k, \quad (A.31)$$

over $\|(y, \mathbf{z}^\top)\| \leq 1$. (A.30) and (A.31) combined with (A.3) yield

$$\mathbf{x}^\top \frac{\partial h_n(\mathbf{x})}{\partial \mathbf{x}} \xrightarrow{P} -y^2 m_{\mu, \mu}(\theta^\circ) - 2y \sum_{j=1}^{10} m_{\mu, \psi_j}(\theta^\circ) z_j - \sum_{j=1}^{10} \sum_{k=1}^{10} m_{\psi_j, \psi_k}(\theta^\circ) z_j z_k, \quad (A.32)$$

over $\|(y, \mathbf{z}^\top)\| \leq 1$. It is easy to see that the limit in the right-side of (A.32) is

$$\begin{aligned} & -\mathbf{x}^\top \begin{pmatrix} m_{\mu, \mu}(\theta^\circ) & m_{\mu, \psi_1}(\theta^\circ) & \cdots & m_{\mu, \psi_{10}}(\theta^\circ) \\ m_{\psi_1, \mu}(\theta^\circ) & m_{\psi_1, \psi_1}(\theta^\circ) & \cdots & m_{\psi_1, \psi_{10}}(\theta^\circ) \\ \vdots & \vdots & \ddots & \vdots \\ m_{\psi_{10}, \mu}(\theta^\circ) & m_{\psi_{10}, \psi_1}(\theta^\circ) & \cdots & m_{\psi_{10}, \psi_{10}}(\theta^\circ) \end{pmatrix} \mathbf{x} \\ & = -\mathbf{x}^\top \left\{ -E_{\theta^\circ} \left(\frac{\partial^2 l_t^\circ(\theta)}{\partial \theta \partial \theta^\top} \Big|_{\theta=\theta^\circ} \right) \right\} \mathbf{x}, \end{aligned}$$

where $\mathbf{x} = (y, \mathbf{z}^\top)^\top$. From Lemma 2, $-E_{\theta^\circ}(\frac{\partial^2 l_t^\circ(\theta)}{\partial \theta \partial \theta^\top} |_{\theta=\theta^\circ}) = \text{Var}_{\theta^\circ}(\frac{\partial l_t^\circ(\theta)}{\partial \theta} |_{\theta=\theta^\circ})$ is positive definite, we have that for $\|\mathbf{x}\| = \|(y, \mathbf{z}^\top)\| = 1$, $\mathbf{x}^\top \frac{\partial h_n(\mathbf{x})}{\partial \mathbf{x}} < 0$. $\square$

A.3 Proof of Theorem 3 (asymptotic normality)

To show (12), we start from a scalar estimator $\hat{\mu}_n$. By the mean value theorem, we have:

$\frac{\partial L_n(\boldsymbol{\theta})}{\partial \mu} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}'_n} - \frac{\partial L_n(\boldsymbol{\theta})}{\partial \mu} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^\circ} = (\hat{\mu}_n - \mu^\circ) \frac{\partial^2 L_n(\boldsymbol{\theta})}{\partial \mu^2} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}$, where $\hat{\boldsymbol{\theta}}'_n = (\hat{\mu}_n, \beta_0^\circ, \beta_1^\circ, \beta_2^\circ, \beta_3^\circ, \alpha_0^\circ, \alpha_1^\circ, \alpha_2^\circ, \alpha_3^\circ, p^\circ, q^\circ)^\top$ (i.e., $\hat{\boldsymbol{\theta}}'_n$ is nearly identical to $\boldsymbol{\theta}^\circ$ except μ° replaced by $\hat{\mu}_n$), and $\boldsymbol{\theta}^*$ is between $\hat{\boldsymbol{\theta}}'_n$ and $\boldsymbol{\theta}^\circ$. Since $\hat{\mu}_n$ is the local maximizer of $L_n(\cdot)$ with parameters $(\beta_0^\circ, \beta_1^\circ, \beta_2^\circ, \beta_3^\circ, \alpha_0^\circ, \alpha_1^\circ, \alpha_2^\circ, \alpha_3^\circ, p^\circ, q^\circ)$ given, $\frac{\partial L_n(\boldsymbol{\theta})}{\partial \mu} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}'_n} = 0$. We have

$$\sqrt{n}(\hat{\mu}_n - \mu^\circ) = -\frac{\sqrt{n} \frac{\partial L_n(\boldsymbol{\theta})}{\partial \mu} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^\circ}}{\frac{\partial^2 L_n(\boldsymbol{\theta})}{\partial \mu^2} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}} = -I_1/I_2.$$

The denominator

$$I_2 \xrightarrow{P} E_{\boldsymbol{\theta}^\circ} \left(\frac{\partial^2 l_t^\circ(\boldsymbol{\theta})}{\partial \mu^2} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^\circ} \right),$$

by Theorem 2. The numerator

$$\begin{aligned} I_1 &= \sqrt{n} \left\{ \frac{1}{n} \sum_{t=1}^n \frac{\partial l_t^\circ(\boldsymbol{\theta})}{\partial \mu} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^\circ} \right\} = \sqrt{n} \left\{ \frac{1}{n} \sum_{t=1}^n \frac{\partial l_t^\circ(\boldsymbol{\theta}^\circ)}{\partial \mu} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^\circ} - E \left(\frac{\partial l_t^\circ(\boldsymbol{\theta})}{\partial \mu} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^\circ} \right) \right\} \\ &\xrightarrow{\mathcal{D}} N \left(0, \left\{ -E_{\boldsymbol{\theta}^\circ} \left(\frac{\partial^2 l_t^\circ(\boldsymbol{\theta})}{\partial \mu^2} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^\circ} \right) \right\} \right). \end{aligned}$$

So, $-I_1/I_2 \xrightarrow{\mathcal{D}} N(0, V)$, with $V = \{-E_{\boldsymbol{\theta}^\circ}(\frac{\partial^2 l_t^\circ(\boldsymbol{\theta})}{\partial \mu^2} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^\circ})\}^{-1}$. The proof can be extended to the vector $\boldsymbol{\theta}$ for $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^\circ) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \{-E_{\boldsymbol{\theta}^\circ}(\frac{\partial^2 l_t^\circ(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^\circ})\}^{-1})$. $\square$

A.4 Proof of Theorem 4 (uniqueness)

Recall $\Theta_{*,n} = \{\boldsymbol{\theta} \in \Theta : \mu \leq c \min_{1 \leq t \leq n} M_t + (1-c)\mu^\circ\}$, with Θ in (7) and $0 < c < 1$. To prove the uniqueness of the MLE, we apply Theorem 2.6 in Makelainen et al. (1981), which states as follows. Let $l(\boldsymbol{\theta})$ be the logarithm of a twice continuously differentiable likelihood function, with $\boldsymbol{\theta}$ varying in a connected open subset $\Theta \subset \mathbb{R}^p$. Suppose that

- (i) the gradient vector ∇l vanishes in at least one point $\boldsymbol{\theta} \in \Theta$, and that
- (ii) the Hessian matrix of $l(\boldsymbol{\theta})$ is negative definite at every point $\boldsymbol{\theta} \in \Theta$.

Then, (a) $l(\boldsymbol{\theta})$ is a strictly concave function of $\boldsymbol{\theta}$, (b) there is a unique maximum likelihood estimate $\boldsymbol{\theta} \in \Theta$, and (c) $l(\boldsymbol{\theta})$ has no other maxima or minima or other stationary points in Θ .

It suffices for us to prove conditions (i) and (ii). For notational simplicity, define sets

$$\begin{aligned}\Theta^a &= \{\boldsymbol{\theta} \in \Theta_{*,n} : \|\boldsymbol{\theta} - \boldsymbol{\theta}^o\| \geq a\}, \\ \Theta^{a^c} &= \{\boldsymbol{\theta} \in \Theta_{*,n} : \|\boldsymbol{\theta} - \boldsymbol{\theta}^o\| < a\}, \\ \Theta^{a,\mu^o} &= \{\boldsymbol{\theta} \in \Theta_{*,n} : \|\boldsymbol{\theta} - \boldsymbol{\theta}^o\| \geq a, \mu > \mu^o\}, \\ \Theta^{a,\mu^{oc}} &= \{\boldsymbol{\theta} \in \Theta_{*,n} : \|\boldsymbol{\theta} - \boldsymbol{\theta}^o\| \geq a, \mu \leq \mu^o\}, \\ \Theta^{a^c,\mu^o} &= \{\boldsymbol{\theta} \in \Theta_{*,n} : \|\boldsymbol{\theta} - \boldsymbol{\theta}^o\| < a, \mu > \mu^o\}, \\ \Theta^{a^c,\mu^{oc}} &= \{\boldsymbol{\theta} \in \Theta_{*,n} : \|\boldsymbol{\theta} - \boldsymbol{\theta}^o\| < a, \mu \leq \mu^o\}.\end{aligned}$$

Then $\Theta^a = \Theta^{a,\mu^o} \cup \Theta^{a,\mu^{oc}}$ and $\Theta^{a^c} = \Theta^{a^c,\mu^o} \cup \Theta^{a^c,\mu^{oc}}$.

First, we check conditions (i). It can be proved by Lemmas 4 and 8 that $\sup_{\boldsymbol{\theta} \in \Theta^a} |L_n(\boldsymbol{\theta}) - L_n^o(\boldsymbol{\theta})| \xrightarrow{P} 0$, then

$$\sup_{\boldsymbol{\theta} \in \Theta^a} L_n(\boldsymbol{\theta}) = \sup_{\boldsymbol{\theta} \in \Theta^a} L_n^o(\boldsymbol{\theta}) + o_P(1) = \max \left\{ \sup_{\boldsymbol{\theta} \in \Theta^{a,\mu^o}} L_n^o(\boldsymbol{\theta}), \sup_{\boldsymbol{\theta} \in \Theta^{a,\mu^{oc}}} L_n^o(\boldsymbol{\theta}) \right\} + o_P(1). \quad (\text{A.33})$$

Since $\sup_{\boldsymbol{\theta} \in \Theta^{a,\mu^o}} |L_n^o(\boldsymbol{\theta}) - L_n^o(\mu^o, \boldsymbol{\psi})| \xrightarrow{P} 0$, we have

$$\sup_{\boldsymbol{\theta} \in \Theta^{a,\mu^o}} L_n(\boldsymbol{\theta}) = \sup_{\boldsymbol{\theta} \in \Theta^{a,\mu^o}} L_n^o(\mu^o, \boldsymbol{\psi}) + o_P(1), \quad (\text{A.34})$$

which, applied to (A.33), gives

$$\begin{aligned}& \max \left\{ \sup_{\boldsymbol{\theta} \in \Theta^{a,\mu^o}} L_n^o(\boldsymbol{\theta}), \sup_{\boldsymbol{\theta} \in \Theta^{a,\mu^{oc}}} L_n^o(\boldsymbol{\theta}) \right\} + o_P(1) \\ &= \max \left\{ \sup_{\boldsymbol{\theta} \in \Theta^{a,\mu^o}} L_n^o(\mu^o, \boldsymbol{\psi}), \sup_{\boldsymbol{\theta} \in \Theta^{a,\mu^{oc}}} L_n^o(\boldsymbol{\theta}) \right\} + o_P(1).\end{aligned} \quad (\text{A.35})$$

Since $\min_{1 \leq t \leq n} M_t \downarrow \mu^o$, $\{\boldsymbol{\psi} : \boldsymbol{\psi} \in \Theta^{a,\mu^o}\} \subseteq \{\boldsymbol{\psi} : \boldsymbol{\psi} \in \Theta^{a/2,\mu^{oc}}\}$. This implies

$$\sup_{\boldsymbol{\theta} \in \Theta^{a,\mu^o}} L_n^o(\mu^o, \boldsymbol{\psi}) \leq \sup_{\boldsymbol{\theta} \in \Theta^{a/2,\mu^{oc}}} L_n^o(\boldsymbol{\theta}). \quad (\text{A.36})$$

Also, we have

$$\sup_{\boldsymbol{\theta} \in \Theta^{a,\mu^{oc}}} L_n^o(\boldsymbol{\theta}) \leq \sup_{\boldsymbol{\theta} \in \Theta^{a/2,\mu^{oc}}} L_n^o(\boldsymbol{\theta}), \quad (\text{A.37})$$

so

$$\max \left\{ \sup_{\boldsymbol{\theta} \in \Theta^{a, \mu^o}} L_n^o(\mu^o, \boldsymbol{\psi}), \sup_{\boldsymbol{\theta} \in \Theta^{a, \mu^{oc}}} L_n^o(\boldsymbol{\theta}) \right\} \leq \sup_{\boldsymbol{\theta} \in \Theta^{a/2, \mu^{oc}}} L_n^o(\boldsymbol{\theta}). \quad (\text{A.38})$$

From (A.33), (A.35) and (A.38), we have $\sup_{\boldsymbol{\theta} \in \Theta^a} L_n(\boldsymbol{\theta}) \leq \sup_{\boldsymbol{\theta} \in \Theta^{a/2, \mu^{oc}}} L_n^o(\boldsymbol{\theta}) + o_P(1)$.

Following the lines of the proof of Proposition 4 in Zhao et al. (2018) together with the proof of Proposition 2 in Dombry (2015), we have $\sup_{\boldsymbol{\theta} \in \Theta^a} L_n(\boldsymbol{\theta}) \leq L_n^o(\boldsymbol{\theta}^o) + o_P(1)$; moreover, using $L_n^o(\boldsymbol{\theta}^o) = L_n(\boldsymbol{\theta}^o) + o_P(1)$, we can conclude

$$\sup_{\boldsymbol{\theta} \in \Theta^a} L_n(\boldsymbol{\theta}) \leq L_n(\boldsymbol{\theta}^o) + o_P(1),$$

which is equivalent to that the global maximizer of $L_n(\boldsymbol{\theta})$ ($\widehat{\boldsymbol{\theta}}$ satisfies $\frac{\partial L_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} = \mathbf{0}$) over $\Theta_{*,n}$ is located within Θ^{a^c} . Thus condition (i) is proved.

Second, for condition (ii), relating Lemma 4 and Lemma 8 to Lemma 7 and Lemma 13 in Zhao et al. (2018), we have

$$\sup_{\boldsymbol{\theta} \in \Theta^{a^c}} \left| \frac{\partial^2}{\partial \theta_j \partial \theta_k} L_n(\boldsymbol{\theta}) - \frac{\partial^2}{\partial \theta_j \partial \theta_k} L_n^o(\boldsymbol{\theta}) \right| \xrightarrow{P} 0.$$

For Θ^{a^c, μ^o} , we have

$$\sup_{\boldsymbol{\theta} \in \Theta^{a^c, \mu^o}} \left| \frac{\partial^2}{\partial \theta_j \partial \theta_k} L_n^o(\boldsymbol{\theta}) - \frac{\partial^2}{\partial \theta_j \partial \theta_k} L_n^o(\mu^o, \boldsymbol{\psi}) \right| \xrightarrow{P} 0.$$

For $\Theta^{a^c, \mu^{oc}}$, we have

$$\sup_{\boldsymbol{\theta} \in \Theta^{a^c, \mu^{oc}}} \left| \frac{\partial^2}{\partial \theta_j \partial \theta_k} L_n^o(\boldsymbol{\theta}) - E_{\boldsymbol{\theta}^o} \left(\frac{\partial^2}{\partial \theta_j \partial \theta_k} l_1(\boldsymbol{\theta}) \right) \right| \xrightarrow{P} 0$$

by ergodicity and the uniform law of large numbers. Since $E_{\boldsymbol{\theta}^o}(\frac{\partial^2 l_1(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}}|_{\boldsymbol{\theta}=\boldsymbol{\theta}^o})$ is negative definite, we can find an a^* such that $E_{\boldsymbol{\theta}^o}(\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}} l_1(\boldsymbol{\theta}))$ is negative definite for $\boldsymbol{\theta} \in \Theta^{a^*, \mu^{oc}}$. We can conclude that the Hessian matrix $\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}} L_n(\boldsymbol{\theta})$ is negative definite over $\Theta^{a^*, c}$, and thus condition (ii) is proved. $\square$

A.5 Proofs of Lemmas

Lemma 3. Assume the conditions in Theorem 2. Let A_U and A_L (B_U and B_L) be the upper bound and lower bound of $\{A_t\}$ ($\{B_t\}$). Suppose $\{c_t\}$ is a bounded arbitrary non-zero sequence of real numbers. Denote $I_c^+ = \{t : c_t > 0\}$, and $I_c^- = \{t : c_t < 0\}$. Denote the upper and lower bounds of $\{c_t/A_t\}$ in I_c^+ and I_c^- as A_U^{c+} and A_L^{c+} , A_U^{c-} and A_L^{c-} respectively; and the upper and lower bounds of $\{B_t^{c_t}\}$ in I_c^+ and I_c^- as B_U^{c+} and B_L^{c+} , B_U^{c-} and B_L^{c-} respectively. Then we have

- (i) $\frac{1}{n} \sum_{t=1}^n (M_t - \mu^0)^{c_t}$ is finite as $n \rightarrow \infty$ if $p > -A_L^{c-}$, or $q > A_U^{c+}$, or both;
- (ii) $\frac{1}{n} \sum_{t=1}^n (M_t - \mu)^{c_t}$ is finite as $n \rightarrow \infty$, if $p > -A_L^{c-}$, or $q > A_U^{c+}$, or both;
- (iii) For any positive k , $\frac{1}{n} \sum_{t=1}^n \{\log(M_t - \mu^0)\}^k \xrightarrow{P} E_{\theta^0} \{\log(M_t - \mu^0)\}^k$ as $n \rightarrow \infty$; and $\frac{1}{n} \sum_{t=1}^n \{\log(M_t - \mu)\}^k \xrightarrow{P} E_{\theta^0} \{\log(M_t - \mu)\}^k$ as $n \rightarrow \infty$.

Proof. First, for part (i) and $t \in I_c^+$, we have

$$\begin{aligned}
 B_L^{c+} \min(Y_t^{A_L^{c+}}, Y_t^{A_U^{c+}}) &\leq B_L^{c+} Y_t^{c_t/A_t} \\
 &\leq (M_t - \mu^0)^{c_t} = B_t^{c_t} Y_t^{c_t/A_t} \\
 &\leq B_U^{c+} Y_t^{c_t/A_t} \leq B_U^{c+} \max\{Y_t^{A_L^{c+}}, Y_t^{A_U^{c+}}\},
 \end{aligned}$$

and then $\frac{1}{n} \sum_{t \in I_c^+} (M_t - \mu^0)^{c_t}$ is finite as $n \rightarrow \infty$ if $E(Y_t^a) < \infty$ where $a = A_U^{c+}$, which is equivalent to $q > A_U^{c+}$. Analogously, for part (i) and $t \in I_c^-$, we have

$$\begin{aligned}
 B_L^{c-} \min(Y_t^{A_L^{c-}}, Y_t^{A_U^{c-}}) &\leq B_L^{c-} Y_t^{c_t/A_t} \\
 &\leq (M_t - \mu^0)^{c_t} = B_t^{c_t} Y_t^{c_t/A_t} \\
 &\leq B_U^{c-} Y_t^{c_t/A_t} \leq B_U^{c-} \max\{Y_t^{A_L^{c-}}, Y_t^{A_U^{c-}}\},
 \end{aligned}$$

and then $\frac{1}{n} \sum_{t \in I_c^-} (M_t - \mu^0)^{c_t}$ is finite as $n \rightarrow \infty$ if $E(Y_t^a) < \infty$ where $a = A_L^{c-}$, which is equivalent to $p > -A_L^{c-}$.

Second, for part (ii), since $\frac{1}{\delta_n}(\min_{1 \leq t \leq n} M_t - \mu^o) \xrightarrow{P} +\infty$, there exists $0 < c < 1$ such that as $n \rightarrow \infty$,

$$P\left(c\left(\min_{1 \leq t \leq n} M_t - \mu^o\right) > \delta_n\right) \rightarrow 1, \quad \text{and } P(c(M_t - \mu^o) > \delta_n) \rightarrow 1. \quad (\text{A.39})$$

We have four cases.

Case 1 ($t \in I_c^-, \mu > \mu^o$): $\frac{1}{n} \sum_{t \in I_c^-} (M_t - \mu)^{c_t} = \frac{1}{n} \sum_{t \in I_c^-} (M_t - \mu^o + \mu^o - \mu)^{c_t} < \frac{1}{n} \sum_{t \in I_c^-} (M_t - \mu^o - \delta_n)^{c_t}$. Therefore, $\frac{1}{n} \sum_{t \in I_c^-} (M_t - \mu)^{c_t} < \frac{1}{n} \sum_{t \in I_c^-} \{M_t - \mu^o - c(M_t - \mu^o)\}^{c_t} = \frac{1}{n} \sum_{t \in I_c^-} \{(1 - c)(M_t - \mu^o)\}^{c_t}$, which is finite as $n \rightarrow \infty$ by part (i).

Case 2 ($t \in I_c^-, \mu < \mu^o$): $\frac{1}{n} \sum_{t \in I_c^-} (M_t - \mu)^{c_t} = \frac{1}{n} \sum_{t \in I_c^-} (M_t - \mu^o + \mu^o - \mu)^{c_t} < \frac{1}{n} \sum_{t \in I_c^-} (M_t - \mu^o)^{c_t}$, which is finite as $n \rightarrow \infty$ by part (i).

Case 3 ($t \in I_c^+, \mu > \mu^o$): $\frac{1}{n} \sum_{t \in I_c^+} (M_t - \mu)^{c_t} = \frac{1}{n} \sum_{t \in I_c^+} (M_t - \mu^o + \mu^o - \mu)^{c_t} < \frac{1}{n} \sum_{t \in I_c^+} (M_t - \mu^o)^{c_t}$, which is finite as $n \rightarrow \infty$ by part (i).

Case 4 ($t \in I_c^+, \mu < \mu^o$): $\frac{1}{n} \sum_{t \in I_c^+} (M_t - \mu)^{c_t} = \frac{1}{n} \sum_{t \in I_c^+} (M_t - \mu^o + \mu^o - \mu)^{c_t} < \frac{1}{n} \sum_{t \in I_c^+} (M_t - \mu^o + \delta_n)^{c_t}$. By (A.39), $\frac{1}{n} \sum_{t \in I_c^+} (M_t - \mu)^{c_t} < \frac{1}{n} \sum_{t \in I_c^+} \{M_t - \mu^o + c(M_t - \mu^o)\}^{c_t} = \frac{1}{n} \sum_{t \in I_c^+} \{(1 + c)(M_t - \mu^o)\}^{c_t}$, which is finite as $n \rightarrow \infty$ by part (i).

Third, for part (iii), we have $|\log(M_t - \mu^o)|^k = |\log(B_t) + 1/A_t \log(Y_t)|^k \leq 2^{k-1}(B_U + 1/A_L |\log(Y_t)|^k)$ by the convexity of x^k , $k \geq 1$. Since $\log(Y_t)$ follows an EGB2 distribution, its k th moment exists for $k \geq 1$ from Wu et al. (2000). By the law of large numbers, ergodicity and stationarity of our model and pointwise ergodicity theorem in Birkhoff (1931), the proof is completed. $\square$

Lemma 4. Assume the conditions in Theorem 2. Let c_t and $c'_t = c_t - 1$ (treat c'_t as c_t) be similarly defined in Lemma 3. Then we have

$$\left| \frac{1}{n} \sum_{t=1}^n (M_t - \mu)^{c_t} - \frac{1}{n} \sum_{t=1}^n (M_t - \mu^o)^{c_t} \right| \leq O_P(\delta_n)$$

with $|\mu - \mu^o| < \delta_n$, $\delta_n \downarrow 0$, $n\delta_n \rightarrow +\infty$, and $\frac{1}{\delta_n}(\min_{1 \leq t \leq n} M_t - \mu^o) \xrightarrow{P} +\infty$.

Proof. Define $\Delta_n = \frac{1}{n} \sum_{t=1}^n (M_t - \mu)^{c_t} - \frac{1}{n} \sum_{t=1}^n (M_t - \mu^\circ)^{c_t}$. Five cases will be discussed for $|\Delta_n|$.

Case 1 ($c_t < 0$, $\mu > \mu^\circ$):

$$\begin{aligned}
& |\Delta_n| \\
& \leq \frac{1}{n} \sum_{t=1}^n |(M_t - \mu)^{c_t} - (M_t - \mu^\circ)^{c_t}| \\
& = \frac{1}{n} \sum_{t=1}^n (M_t - \mu)^{c_t-1} (M_t - \mu^\circ) |(M_t - \mu)(M_t - \mu^\circ)^{-1} - \{(M_t - \mu)(M_t - \mu^\circ)^{-1}\}^{1-c_t}| \\
& \leq \frac{1}{n} \sum_{t=1}^n (M_t - \mu)^{c_t-1} (M_t - \mu^\circ) (1 - c_t) |1 - (M_t - \mu)(M_t - \mu^\circ)^{-1}| \\
& = \frac{1}{n} \sum_{t=1}^n (1 - c_t) (\mu - \mu^\circ) (M_t - \mu)^{c_t-1},
\end{aligned}$$

where the second inequality uses the fact that $x - x^{m+1} \leq (m+1)(1-x)$ with $m > 0$ and $0 < x < 1$. By Lemma 3, $|\Delta_n| \leq O_P(\delta_n)$.

Case 2 ($c_t < 0$, $\mu < \mu^\circ$): The derivations are similar to those in Case 1.

Case 3 ($c_t > 0$, $\mu > \mu^\circ$):

$$\begin{aligned}
& |\Delta_n| \\
& \leq \frac{1}{n} \sum_{t=1}^n |(M_t - \mu)^{c_t} - (M_t - \mu^\circ)^{c_t}| \\
& = \frac{1}{n} \sum_{t=1}^n (M_t - \mu)^{-1} (M_t - \mu^\circ)^{1+c_t} |(M_t - \mu)(M_t - \mu^\circ)^{-1} - \{(M_t - \mu)(M_t - \mu^\circ)^{-1}\}^{1+c_t}| \\
& \leq \frac{1}{n} \sum_{t=1}^n (M_t - \mu)^{-1} (M_t - \mu^\circ)^{1+c_t} (1 + c_t) |1 - (M_t - \mu)(M_t - \mu^\circ)^{-1}| \\
& = \frac{1}{n} \sum_{t=1}^n (1 + c_t) (\mu - \mu^\circ) (M_t - \mu)^{-1} (M_t - \mu^\circ)^{c_t}.
\end{aligned}$$

By Cauchy-Schwartz inequality and Lemma 3, $|\Delta_n| \leq O_P(\delta_n)$.

Case 4 ($c_t > 0$, $\mu < \mu^\circ$): The proof is similar to that in Case 3.

Case 5 (c_t has positive and negative values): The claim holds by considering I_c^+ and I_c^- as in Lemma 3.

□

Lemma 5. Assume the conditions in Theorem 2. If $\|(\beta_0, \beta_1, \beta_2, \beta_3) - (\beta_0^\circ, \beta_1^\circ, \beta_2^\circ, \beta_3^\circ)\| < \delta_n$ where $\delta_n \downarrow 0$, then $|B'_t - B_t^{\circ}| \leq O(\delta_n)$; similarly, if $\|(\alpha_0, \alpha_1, \alpha_2, \alpha_3) - (\alpha_0^\circ, \alpha_1^\circ, \alpha_2^\circ, \alpha_3^\circ)\| < \delta_n$ where $\delta_n \downarrow 0$, then $|A'_t - A_t^{\circ}| \leq O(\delta_n)$.

Proof. We only prove the result for A_t ; the proof for B_t is similar and omitted. Denote the upper bound of α_1 and β_1 as $C_b < 1$ in Θ .

By the auto-regressive formula (5), we have:

$$\begin{aligned} \log(A_t) &= \alpha_0 + \alpha_1 \log(A_{t-1}) + \alpha_2 \exp(-\alpha_3 M_{t-1}) \\ &= \alpha_0 \sum_{i=1}^{t-1} \alpha_1^{i-1} + \alpha_1^{t-1} \log(A_1) + \alpha_2 \sum_{i=1}^{t-1} \alpha_1^{i-1} \exp(-\alpha_3 M_{t-i}), \end{aligned}$$

and

$$\begin{aligned} &|\log(A'_t) - \log(A_t^{\circ})| \\ \leq &\left| \alpha_0 \sum_{i=1}^{t-1} \alpha_1^{i-1} - \alpha_0^\circ \sum_{i=1}^{t-1} (\alpha_1^\circ)^{i-1} \right| \end{aligned} \tag{A.40}$$

$$+ \left| \alpha_1^{t-1} \log(A'_1) - (\alpha_1^\circ)^{t-1} \log(A_1) \right| \tag{A.41}$$

$$+ \left| \alpha_2 \sum_{i=1}^{t-1} \alpha_1^{i-1} \exp(-\alpha_3 M_{t-i}) - \alpha_2^\circ \sum_{i=1}^{t-1} (\alpha_1^\circ)^{i-1} \exp(-\alpha_3^\circ M_{t-i}) \right|. \tag{A.42}$$

We analyze (A.40), (A.41) and (A.42) one by one next. For (A.40),

$$\begin{aligned} \text{(A.40)} &= \left| \alpha_0 \sum_{i=1}^{t-1} \alpha_1^{i-1} - \alpha_0 \sum_{i=1}^{t-1} (\alpha_1^\circ)^{i-1} + \alpha_0 \sum_{i=1}^{t-1} (\alpha_1^\circ)^{i-1} - \alpha_0^\circ \sum_{i=1}^{t-1} (\alpha_1^\circ)^{i-1} \right| \\ &\leq |\alpha_0| \cdot \left| \sum_{i=1}^{t-1} \alpha_1^{i-1} - \sum_{i=1}^{t-1} (\alpha_1^\circ)^{i-1} \right| + |\alpha_0 - \alpha_0^\circ| \cdot \left| \sum_{i=1}^{t-1} (\alpha_1^\circ)^{i-1} \right|. \end{aligned}$$

Since

$$\begin{aligned} &|\alpha_0| \cdot \left| \sum_{i=1}^{t-1} \alpha_1^{i-1} - \sum_{i=1}^{t-1} (\alpha_1^\circ)^{i-1} \right| \\ \leq &|\alpha_0| \left(\left| \sum_{i=1}^{\infty} \alpha_1^{i-1} - \sum_{i=1}^{\infty} (\alpha_1^\circ)^{i-1} \right| + \left| \alpha_1^{t-1} \sum_{i=1}^{\infty} \alpha_1^{i-1} - (\alpha_1^\circ)^{t-1} \sum_{i=1}^{\infty} (\alpha_1^\circ)^{i-1} \right| \right) \\ = &|\alpha_0| \left(\left| \frac{1}{1-\alpha_1} - \frac{1}{1-\alpha_1^\circ} \right| + \left| \frac{\alpha_1^{t-1}}{1-\alpha_1} - \frac{(\alpha_1^\circ)^{t-1}}{1-\alpha_1^\circ} \right| \right) \end{aligned}$$

$$\begin{aligned}
&\leq |\alpha_0| \{ \delta_n / (1 - C_b)^2 + o(\delta_n) \} \\
&= O(\delta_n),
\end{aligned}$$

and

$$|\alpha_0 - \alpha_0^o| \cdot \left| \sum_{i=1}^{t-1} (\alpha_1^o)^{i-1} \right| < \delta_n \left| \sum_{i=1}^{t-1} (\alpha_1^o)^{i-1} \right| = O(\delta_n)$$

we obtain (A.40) $\leq O(\delta_n)$. For (A.41), we see that

$$(A.41) = |\alpha_1^{t-1} - (\alpha_1^o)^{t-1}| \cdot |\log(A'_1)| \leq |\alpha_1 - \alpha_1^o| \cdot |\log(A'_1)| = O(\delta_n).$$

For (A.42),

$$(A.42) = \left| \alpha_2 \sum_{i=1}^{t-1} \alpha_1^{i-1} \exp(-\alpha_3 M_{t-i}) - \alpha_2^o \sum_{i=1}^{t-1} (\alpha_1^o)^{i-1} \exp(-\alpha_3 M_{t-i}) \right| \quad (A.43)$$

$$+ \left| \alpha_2^o \sum_{i=1}^{t-1} \alpha_1^{i-1} \exp(-\alpha_3 M_{t-i}) - \alpha_2^o \sum_{i=1}^{t-1} (\alpha_1^o)^{i-1} \exp(-\alpha_3 M_{t-i}) \right| \quad (A.44)$$

$$+ \left| \alpha_2^o \sum_{i=1}^{t-1} (\alpha_1^o)^{i-1} \exp(-\alpha_3 M_{t-i}) - \alpha_2^o \sum_{i=1}^{t-1} (\alpha_1^o)^{i-1} \exp(-\alpha_3^o M_{t-i}) \right|, \quad (A.45)$$

where

$$(A.43) = |\alpha_2 - \alpha_2^o| \cdot \left| \sum_{i=1}^{t-1} \alpha_1^{i-1} \exp(-\alpha_3 M_{t-i}) \right| = O(\delta_n),$$

$$(A.44) \leq \alpha_2^o \sum_{i=1}^{t-1} |\alpha_1^{i-1} - (\alpha_1^o)^{i-1}| \exp(-\alpha_3 M_{t-i}) = O(\delta_n),$$

$$(A.45) \leq \alpha_2^o \sum_{i=1}^{t-1} (\alpha_1^o)^{i-1} |\exp(-\alpha_3 M_{t-i}) - \exp(-\alpha_3^o M_{t-i})|.$$

By the mean value theorem: $|\exp(-\alpha_3 M_{t-i}) - \exp(-\alpha_3^o M_{t-i})| = |(-M_{t-i}) \exp(-\alpha_3^* M_{t-i}) (\alpha_3 - \alpha_3^o)| = M_{t-i} \exp(-\alpha_3^* M_{t-i}) |\alpha_3 - \alpha_3^o|$, where α_3^* is between α_3 and α_3^o . It follows that $(A.45) \leq \alpha_2^o \sum_{i=1}^{t-1} (\alpha_1^o)^{i-1} M_{t-i} \exp(-\alpha_3^* M_{t-i}) |\alpha_3 - \alpha_3^o| = O(\delta_n)$.

Thus, we conclude $|\log(A'_t) - \log(A_t^o)| \leq O(\delta_n)$, which implies $|A'_t - A_t^o| \leq O(\delta_n)$. $\square$

Lemma 6. *Under the conditions in Theorem 2, there exists a constant C such that $|A_t - A'_t| \leq C U_{r,b}^{t-1}$ and $|B_t - B'_t| \leq C U_{r,b}^{t-1}$, where $U_{r,b}$ is the upper bound of β_1 and α_1 .*

Proof. We only prove the situation for A_t , and the proof for B_t is similar.

We start from $\log(A_2)$ and $\log(A'_2)$:

$$\begin{cases} \log(A_2) &= \alpha_0 + \alpha_1 \log(A_1) + \alpha_2 \exp(-\alpha_3 x_1), \\ \log(A'_2) &= \alpha_0 + \alpha_1 \log(A'_1) + \alpha_2 \exp(-\alpha_3 x_1). \end{cases}$$

We have $|\log(A_2) - \log(A'_2)| = \alpha_1 |\log(A_1) - \log(A'_1)|$. Then $|\log(A_3) - \log(A'_3)| = \alpha_1 |\log(A_2) - \log(A'_2)| = \alpha_1^2 |\log(A_1) - \log(A'_1)|$. Repeating the auto-regressive process, we will have $|\log(A_t) - \log(A'_t)| = \alpha_1^{t-1} |\log(A_1) - \log(A'_1)| \leq U_{r,b}^{t-1} C_1$ for some constant C_1 .

So, we can conclude $|A_t - A'_t| \leq C U_{r,b}^{t-1}$ for some C . $\square$

Lemma 7. Assume the conditions in Theorem 2 and Lemma 4, and $|A'_t - A_t^o| \leq O(\delta_n)$.

Then $\frac{1}{n} \sum_{t=1}^n |(M_t - \mu)^{-A'_t} - (M_t - \mu)^{-A_t^o}| = O_P(\delta_n)$.

Proof. By the mean value theorem: $\frac{1}{n} \sum_{t=1}^n |(M_t - \mu)^{-A'_t} - (M_t - \mu)^{-A_t^o}| = \frac{1}{n} \sum_{t=1}^n |A'_t - A_t^o| (M_t - \mu)^{-A_t^*} |\log(M_t - \mu)| = O_P(\delta_n)$ by Lemma 3, where A_t^* is between A'_t and A_t^o . $\square$

Lemma 8. Under the conditions in Theorem 2 and Lemma 4, $\frac{1}{n} \sum_{t=1}^n |(M_t - \mu)^{-A_t} - (M_t - \mu)^{-A'_t}| \xrightarrow{P} 0$.

Proof. By the mean value theorem:

$$\begin{aligned} & \frac{1}{n} \sum_{t=1}^n |(M_t - \mu)^{-A_t} - (M_t - \mu)^{-A'_t}| \\ &= \frac{1}{n} \sum_{t=1}^n |A_t - A'_t| (M_t - \mu)^{-A_t^*} |\log(M_t - \mu)| \\ &\leq \frac{C}{n} \sum_{t=1}^n (M_t - \mu)^{-A_t^*} |\log(M_t - \mu)| U_{r,b}^{t-1}, \end{aligned}$$

where the last step is according to Lemma 6. Since $\sum_{t=1}^n (M_t - \mu)^{-A_t^*} |\log(M_t - \mu)| U_{r,b}^{t-1}$ is finite with probability one as $U_{r,b}$ is between 0 and 1, and $U_{r,b}^{t-1}$ (together with the coefficients) forms a convergent geometric series in the summation when $n \rightarrow \infty$ and accordingly for large t . We have

$$\sum_{t=1}^n (M_t - \mu)^{-A_t^*} |\log(M_t - \mu)| U_{r,b}^{t-1} = \sum_{t=1}^n \left[(M_t - \mu)^{-A_t^*} U_{r,b}^{\frac{t-1}{3}} \right] \left[|\log(M_t - \mu)| U_{r,b}^{\frac{t-1}{3}} \right] U_{r,b}^{\frac{t-1}{3}}.$$

It is easy to check $[(M_t - \mu)^{-A_t^*} U_{r,b}^{\frac{t-1}{3}}]$, $[|\log(M_t - \mu)| U_{r,b}^{\frac{t-1}{3}}]$, $U_{r,b}^{\frac{t-1}{3}}$ decay either exponentially or super exponentially, $[(M_t - \mu)^{-A_t^*} U_{r,b}^{\frac{t-1}{3}}] [|\log(M_t - \mu)| U_{r,b}^{\frac{t-1}{3}}] U_{r,b}^{\frac{t-1}{3}}$ decays exponentially fast for large t , and the probabilities decrease fast enough for the sum to grow slower than n , then $\frac{C}{n} \sum_{t=1}^n (M_t - \mu)^{-A_t^*} |\log(M_t - \mu)| U_{r,b}^{t-1} \xrightarrow{P} 0$. $\square$

A.6 Transformation between distributions

In this section, we show connections between distributions.

- (i) The GB2($a = 1, b = 1, p, q$) and the prime beta(p, q) distributions are equivalent.

Proof. (A.1) and (A.2) verify the equivalence. $\square$

- (ii) If $Y_t \sim$ the prime beta(p, q) distribution, then for constants A_t and B_t , $B_t Y_t^{1/A_t} \sim$ GB2(A_t, B_t, p, q).

Proof. For $Y_t \sim$ the prime beta(p, q) distribution, An application of the Jacobian theorem for the transformed variable $B_t Y_t^{1/A_t}$ and the p.d.f. in (A.2) completes the proof. $\square$

A.7 Some limit properties of GB2

Considering $a > 0$, $p > 0$, $q > 0$, $x > 0$, and $t > 0$, we derive the following expressions and limits: (i) From

$$f_X(tx; a, b, p, q) = \frac{a}{B(p, q)tx} (tx/b)^{ap} \{1 + (tx/b)^a\}^{-(p+q)},$$

we get

$$\frac{f_X(tx; a, b, p, q)}{f_X(x; a, b, p, q)} = \frac{1}{t} t^{ap} \left\{ \frac{1 + (\frac{tx}{b})^a}{1 + (\frac{x}{b})^a} \right\}^{-(p+q)} = t^{ap-1} \left\{ \frac{\frac{1}{(\frac{x}{b})^a} + t^a}{\frac{1}{(\frac{x}{b})^a} + 1} \right\}^{-(p+q)},$$

yielding

$$\begin{aligned}\lim_{x \rightarrow \infty} \frac{f_X(tx; a, b, p, q)}{f_X(x; a, b, p, q)} &= t^{ap-1} t^{-a(p+q)} = t^{-aq-1}, \\ \lim_{x \rightarrow 0} \frac{f_X(tx; a, b, p, q)}{f_X(x; a, b, p, q)} &= t^{ap-1} \lim_{x \rightarrow 0} \left\{ \frac{1 + (\frac{tx}{b})^a}{1 + (\frac{x}{b})^a} \right\}^{-(p+q)} = t^{ap-1}.\end{aligned}$$

(ii) Analogously, for $a > 0$, from

$$f_X(tx; -a, b, p, q) = \frac{ax^{-ap-1}t^{-ap-1}}{B(p, q)b^{-ap}} \left\{ 1 + \left(\frac{b}{tx} \right)^a \right\}^{-(p+q)},$$

we obtain

$$\frac{f_X(tx; -a, b, p, q)}{f_X(x; -a, b, p, q)} = t^{-ap-1} \left\{ \frac{1 + (\frac{b}{tx})^a}{1 + (\frac{b}{x})^a} \right\}^{-(p+q)},$$

leading to

$$\begin{aligned}\lim_{x \rightarrow \infty} \frac{f_X(tx; -a, b, p, q)}{f_X(x; -a, b, p, q)} &= t^{-ap-1}, \\ \lim_{x \rightarrow 0} \frac{f_X(tx; -a, b, p, q)}{f_X(x; -a, b, p, q)} &= t^{-ap-1} \lim_{x \rightarrow 0} \left\{ \frac{x^a + (\frac{b}{t})^a}{x^a + b^a} \right\}^{-(p+q)} = t^{-ap-1} t^{ta(p+q)} = t^{aq-1}.\end{aligned}$$

(iii) Moreover, for $a > 0$,

$$\lim_{q \rightarrow \infty} f_X(x; a, q^{1/a}\beta, p, q) = \frac{ax^{ap-1}e^{-(x/\beta)^a}}{\beta^{ap}\Gamma(p)}, \quad (\text{A.46})$$

where $B(p, q) \approx \Gamma(p)q^{-p}$.

For $a < 0$, the limits can be derived similarly; see Eqn. (2). $\square$

B List of 22 Primary Dealers

- Bank of Montreal, Chicago Branch
- Bank of Nova Scotia, New York Agency
- BNP Paribas Securities Corp.
- Barclays Capital Inc.

- BofA Securities, Inc.
- Citigroup Global Markets Inc.
- Credit Suisse AG, New York Branch
- Daiwa Capital Markets America Inc.
- Deutsche Bank Securities Inc.
- Goldman Sachs & Co. LLC
- HSBC Securities (USA) Inc.
- Jefferies LLC
- J.P. Morgan Securities LLC
- Mizuho Securities USA LLC
- Morgan Stanley & Co. LLC
- NatWest Markets Securities Inc.
- Nomura Securities International, Inc.
- RBC Capital Markets, LLC
- Societe Generale, New York Branch
- TD Securities (USA) LLC
- UBS Securities LLC.
- Wells Fargo Securities, LLC

C Illustrating scale parameter and tail index dynamics

In this section, we explore the inherent characteristics within markets' most severe scenarios by examining the recovered scale dynamics $\{B_t\}$ (σ_t) and the tail index (shape parameter) dynamics $\{A_t\}$ (α_t) derived from the fitted **AcGB2** models and AcF models, respectively. These dynamics are plotted in Figures 10–12.

Upon examining the left panels of the figures, it is evident that the **AcGB2**-generated A_t and the AcF-generated α_t exhibit similar patterns and clusters of extreme values, depicted by rectangles. However, they showcase different value ranges: approximately 18 to 80 (3.2 to 16) in Figure 10 (DJIA 30); 20 to 65 (3 to 8.2) in Figure 11 (S&P 100); and 8 to 32 (3.8 to 16) in Figure 12 (Primary dealers), corresponding to A_t (α_t), respectively. These different parameter ranges indicate that **AcGB2** and AcF are distinct models, even though they share the same dynamic structures. As such, they can uncover both similar and unique market patterns. These observations suggest that the **AcGB2** model effectively captures hidden features in tail behavior, complementing the AcF model. It is noteworthy that the lower bounds of α_t significantly surpass 2, indicating that transformed maxima ensure the existence of second moments—a guarantee absent in maxima without such transformation.

In the right panels of the figures, the patterns of **AcGB2**-generated B_t and AcF-generated σ_t diverge, a divergence explicable through the fitted models. This discrepancy arises from the first-order autoregressive parameters in AcF fitted models, which are close to 1, whereas those in **AcGB2** fitted models deviate from this trend. In Figure 10, the rectangles, certain points above horizontal lines in Figure 11, and the points with ellipses in Figure 12 correspond to extreme clusters observed in the left panels and in Figure 3. Their values approximately range from 0.1705 to 0.179 (0.1185 to 0.13) in Figure 10 (DJIA 30); 0.201 to 0.213 (0.112 to 0.133) in Figure 11 (S&P 100); and 0.105 to 0.125 (0.14 to

0.172) in Figure 12 (Primary dealers), corresponding to B_t (σ_t) respectively. Notably, in Figure 12 for primary dealers, the largest value in the upper right panel is smaller than the smallest value in the lower right panel.

The first rectangle and/or ellipse corresponds to the 2015-16 stock market selloff, also known as the Great Fall of China, a period marked by a global decline in stock prices. The second rectangle and/or ellipse represents the market's downturn in December 2018, driven by concerns over America's trade war with China and rising interest rates. As such, it may be safe to say these rectangles and ellipses serve as indicators of market instability, symbolizing the most severe periods within the worst scenarios.

D Tables and Figures in the Paper

θ	θ°	$n = 1000$		$n = 3000$		$n = 5000$	
		Mean	S.D.	Mean	S.D.	Mean	S.D.
μ	-0.00714	-0.0063	0.002	-0.0067	0.00097	-0.007	0.00077
α_0	-0.318	-0.223	0.1752	-0.2822	0.127	-0.3734	0.13395
α_1	0.904	0.9062	0.0364	0.9111	0.0158	0.9038	0.01378
α_2	0.460	0.3817	0.1873	0.4219	0.12782	0.5133	0.13739
α_3	7.180	17.0594	11.8678	9.393	4.10611	6.7484	2.04245
β_0	-1.670	-1.7335	0.7051	-1.6371	0.18912	-1.6325	0.18154
β_1	0.390	0.4163	0.1812	0.4023	0.0299	0.3909	0.02878
β_2	0.675	0.5349	0.2105	0.6721	0.1376	0.7045	0.14357
β_3	9.170	19.7378	12.5339	10.9751	3.21432	9.5999	3.0086
p	2.000	2.2598	1.8664	1.9312	0.579	2.0029	0.34125
q	2.000	2.3594	1.745	1.9897	0.52327	2.0522	0.2354

Table 1: Mean and standard deviation of MLE for 500 sets of simulated data with lengths $n \in \{1000, 3000, 5000\}$.

θ	θ°	$n = 1000$		$n = 3000$		$n = 5000$	
		Mean	S.D.	Mean	S.D.	Mean	S.D.
μ	-0.00714	-0.0049	0.002	-0.0056	0.0011	-0.0057	0.0004
α_0	-0.318	-0.2393	0.217	-0.2757	0.1235	-0.3313	0.0786
α_1	0.904	0.914	0.037	0.9095	0.0119	0.9011	0.0078
α_2	0.460	0.4002	0.2158	0.4299	0.1153	0.4926	0.0856
α_3	7.180	16.4856	12.0644	8.9954	4.4724	7.4553	1.2429
β_0	-1.670	-1.7005	0.685	-1.539	0.1851	-1.7084	0.129
β_1	0.390	0.4227	0.1843	0.4008	0.0287	0.3864	0.0201
β_2	0.675	0.5493	0.2115	0.7665	0.156	0.6513	0.1659
β_3	9.170	21.6664	15.468	8.0449	2.7684	10.2608	2.427
p	2.000	2.1568	1.5376	1.8774	0.5021	2.0721	0.327
q	2.000	2.1873	1.4392	1.7617	0.3984	1.9832	0.3119

Table 2: *Robust check of Table 1: Mean and standard deviation of MLE for 100 sets of simulated data with lengths $n \in \{1000, 3000, 5000\}$.*

Market	< 0%	> 5%	> 10%	> 15%	> 20%	> 30%	min	mean	median	max
**	<1	> 1.051	> 1.105	> 1.162	> 1.221	> 1.350				
DJIA 30	2.47%	8.11%	0.88%	0	0	0	−0.0211	0.0238	0.0191	0.1386
S&P 100	0.40%	28.48%	5.25%	0.88%	0.32%	0.08%	−0.0131	0.0439	0.0345	0.3210
PD 22	4.53%	6.92%	0.48%	0.16%	0.16%	0.08%	−0.0157	0.0218	0.0175	0.3216

Table 3: *Descriptive statistics of price changes.* ‘**’ corresponds to the values in Figure 3, representing maxima of $(V_{i,t})$. ‘PD 22’ denotes stocks from 22 primary dealers. Note that we are considering exponentially transformed negative log-returns $(V_{i,t})$. In the first row, $< 0\%$ means all stocks experienced price increases. In the next row, < 1 corresponds to the exponential transformation, i.e., $< \exp(0) = 1$. Similarly, $> 5\%$ represents price drops of more than 5% (and $\exp(0.05) = 1.051$ in the next row). The other numbers are interpreted similarly.

Model\Market	DJIA 30	S&P 100	PD 22
AcGB2	2.7978	2.3146	2.7239
AcF	2.7862	2.3002	2.7208

Table 4: *Log-likelihood values of the fitted AcGB2 and AcF models.* ‘PD 22’ represents stocks from 22 primary dealers.

Fitting Model	μ	α_0	α_1	α_2	α_3	β_0
AcGB2	0.8375	-1.8683	0.8884	15.0000	1.8281	-1.5000
	(0.0253)	(0.1102)	(0.0008)	(0.5400)	(0.3163)	(0.1835)
	β_1	β_2	β_3	p	q	
	0.1000	5.6562	4.1862	0.7770	0.2352	
	(0.0109)	(33.2229)	(7.3629)	(0.1759)	(0.0974)	
	μ	γ_0	γ_1	γ_2	γ_3	β_0
AcF	0.89408	-2.363	0.84027	19.329	1.9104	-0.1261
	(0.019866)	(2.3043)	(0.0046103)	(18.805)	(2.0047)	(2.2438)
	β_1	β_2	β_3			
	0.81849	0.94259	1.2718			
	(0.032036)	(2.838)	(11.489)			

Table 5: *Dow Jones 30 - Estimated parameters with standard deviations for two models.*

The standard deviations are recorded in parentheses.

Fitting model	μ	α_0	α_1	α_2	α_3	β_0
AcGB2	0.8147	-0.5593	0.7381	12.5600	2.0000	-1.0000
	(0.0178)	(0.0828)	(0.0029)	(0.6366)	(0.1507)	(0.0656)
	β_1	β_2	β_3	p	q	
	0.3398	6.0000	4.6027	0.9500	0.1849	
	(0.0042)	(15.9230)	(3.1104)	(0.2104)	(0.1056)	
	μ	γ_0	γ_1	γ_2	γ_3	β_0
AcF	0.9131	-0.6258	0.7357	19.7870	2.7588	-0.3271
	(0.0109)	(1.0738)	(0.0145)	(34.5290)	(2.2809)	(0.3714)
	β_1	β_2	β_3			
	0.7664	3.9422	2.9942			
	(0.0137)	(19.7100)	(6.8272)			

Table 6: *S&P 100 - Estimated parameters with standard deviations for two models. Standard deviations are recorded in parentheses.*

Fitting model	μ	α_0	α_1	α_2	α_3	β_0
AcGB2	0.9000	-0.6635	0.9203	15.0000	2.7403	-1.6917
	(0.0153)	(0.0339)	(0.0003)	(0.6899)	(0.3624)	(0.4220)
	β_1	β_2	β_3	p	q	
	0.1009	2.0000	1.8340	0.9900	0.3455	
	(0.0667)	(2.6524)	(2.4709)	(0.2289)	(0.1435)	
	μ	γ_0	γ_1	γ_2	γ_3	β_0
AcF	0.8693	-0.7221	0.9144	19.5100	2.9812	-0.7798
	(0.0192)	(0.0533)	(0.0003)	(2.0794)	(0.5698)	(2.2496)
	β_1	β_2	β_3			
	0.3809	1.5592	1.2808			
	(0.1671)	(2.8407)	(6.9248)			

Table 7: 22 Primary dealers - Estimated parameters with standard deviations for two models. Standard deviations are recorded in parentheses.

Market	DJIA 30		S&P 100		PD 22	
Model\Sig. level	0.01	0.05	0.01	0.05	0.01	0.05
AcGB2	10	63	4	58	6	98
AcF	180	1142	101	677	7	102

Table 8: Number of rejections of assumed models using 10000 K-S tests.

Model \ Market	DJIA 30	S&P 100	PD 22
AcGB2	0.6182	0.4593	0.7309
AcF	0.5544	0.4613	0.7237

Table 9: *Correlations between A_t in **AcGB2** (σ_t in AcF) with GARCH-fitted average volatilities. The ‘average volatilities’ represent the averaging GARCH volatilities among all 28 fitted GARCH(1,1) models within 28 DJIA stocks, similarly computed for S&P 100 and PD 22.*

Market \ Model	AcGB2			AcF		
	Days of $R > P$	Percents	Max.P.V	Days of $R > P$	Percents	Max.P.V
DJIA 30	44	0.0599	15.5127	50	0.0680	67.9573
S&P 100	36	0.0490	2.0586	42	0.0571	1.8389
PD 22	43	0.0585	1.5231	50	0.0680	1.4251

Table 10: *Prediction Statistics. ‘Days of $R > P$ ’ represent the days where the actual maxima of the transformed negative simple-returns exceed the predicted values, i.e., underpredicted. ‘Percents’ are calculated as ‘Days of $R > P$ ’ divided by the prediction length (days). ‘Max.P.V’ denotes the maximum predicted value.*

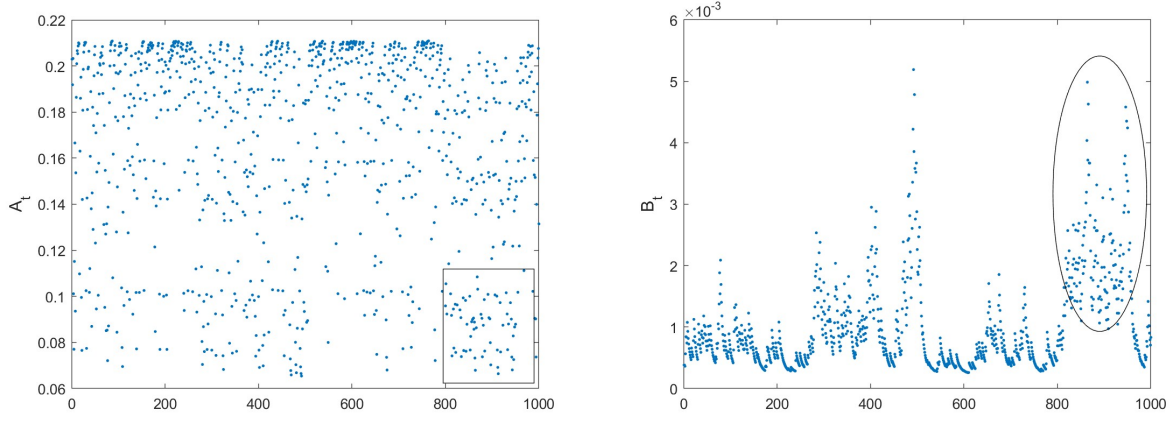


Figure 1: Simulated dynamics of A_t (on the left) and B_t (on the right) over time.

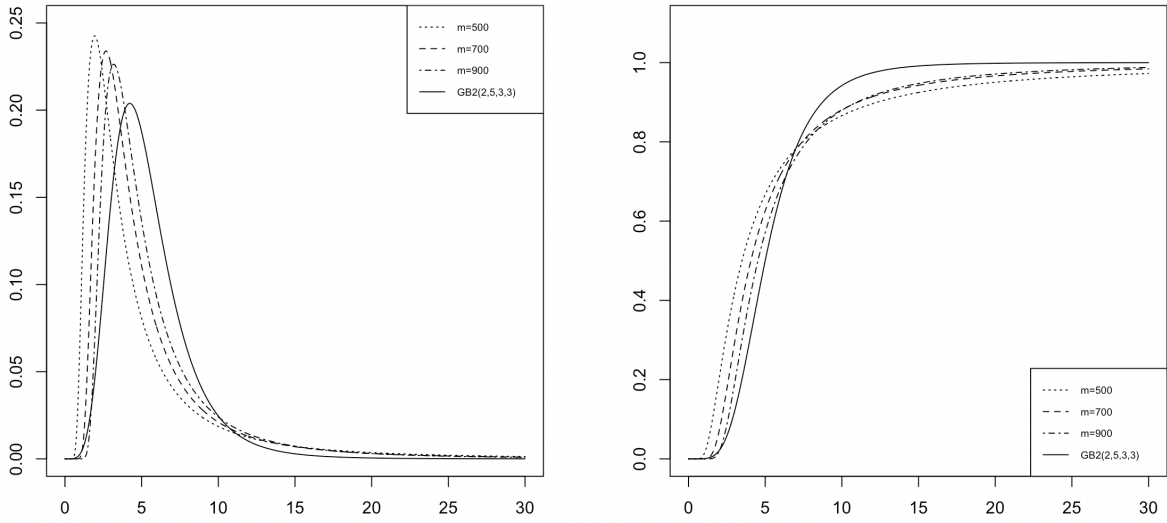


Figure 2: Comparing the distributions of the fitted Fréchet distribution with the true $GB2(2, 5, 3, 3)$ distribution. The left panel displays p.d.f.s, and the right panel displays C.D.F.s, with the dotted line for $m = 500$, dashed line for $m = 700$, dash-dotted line for $m = 900$, and a solid line for $GB2(2, 5, 3, 3)$ in each panel.

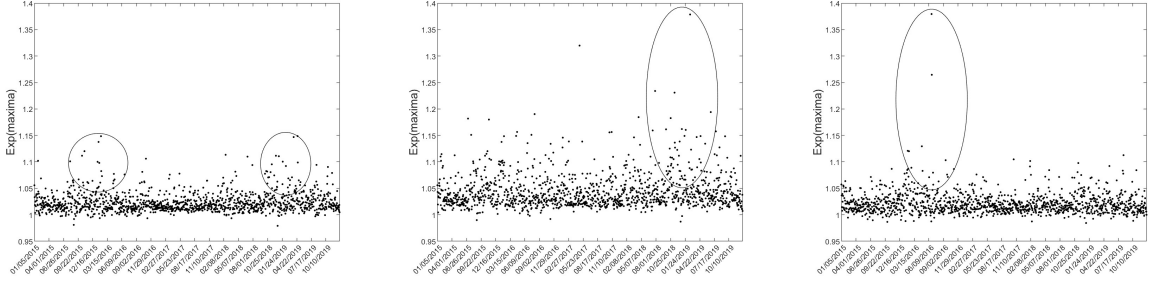


Figure 3: *Changes in transformed maxima over 01/01/2015-12/31/2019 of negative simple-returns of 29 stocks from DJIA 30 (left panel), 97 stocks from S&P 100 (middle panel), and 22 stocks from primary dealers (right panel). A value of 1.3794 corresponds to a log-price change by $\log(1.3794) = 32.16\%$ or a price drop by $1 - 1/1.3794 = 27.5\%$ for one of the stocks in the market.*

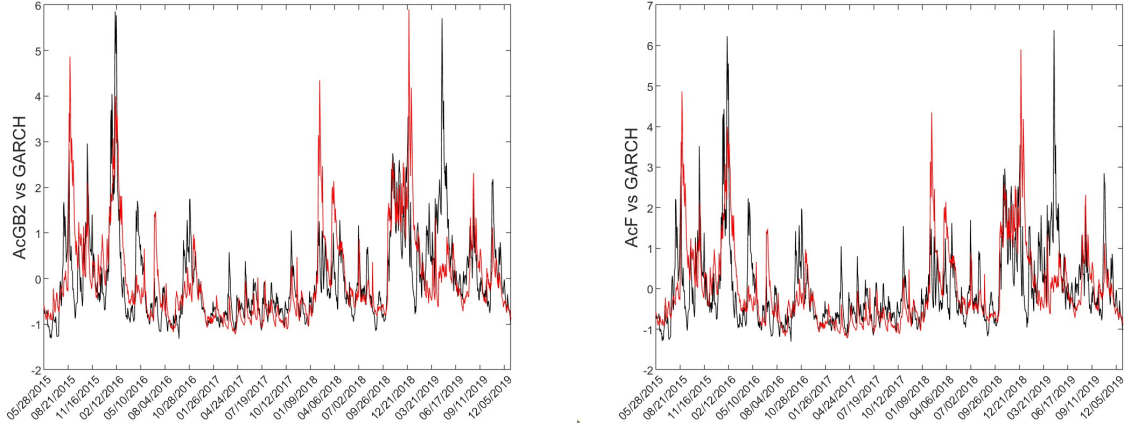


Figure 4: DJIA 30: *Recovered reciprocals of dynamic tail parameter series (black line) alongside the daily average volatilities provided by the GARCH models (red line) for two models: the **AcGB2** model (left) and the **AcF** model (right).*

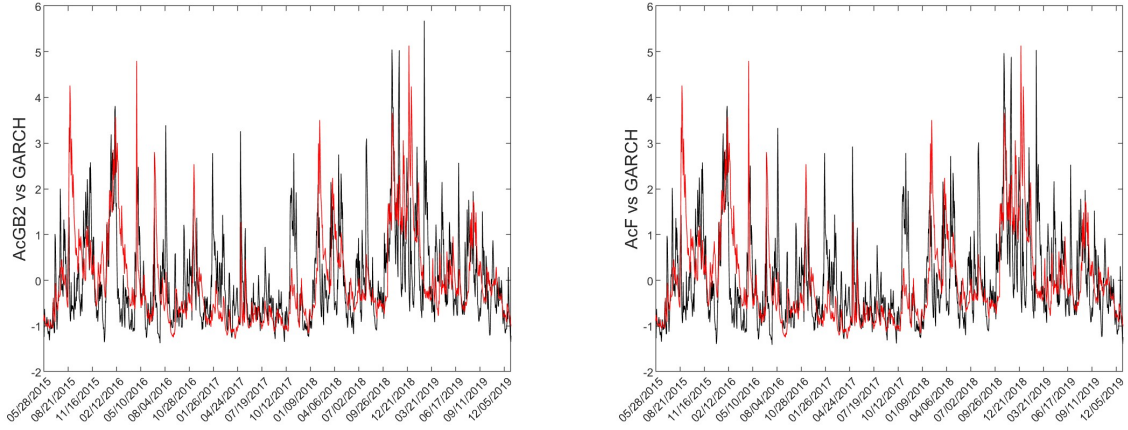


Figure 5: *S&P 100: Recovered dynamic scale parameter series (black line) alongside the daily average volatility provided by the GARCH models (red line) for two models: the **AcGB2** model (left) and the AcF model (right).*

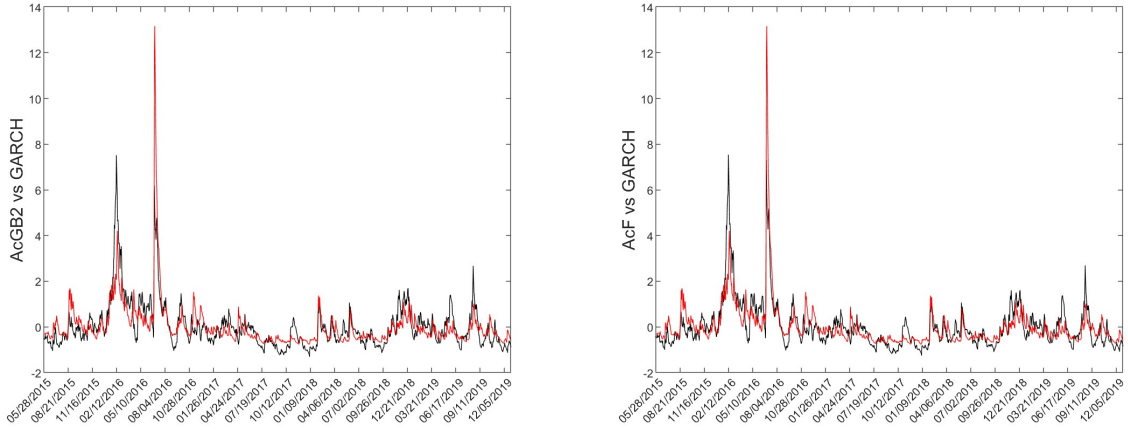


Figure 6: *Primary Dealers: Recovered dynamic scale parameter series (black line) alongside the daily average volatility provided by the GARCH models (red line) for two models: the **AcGB2** model (left) and the AcF model (right).*

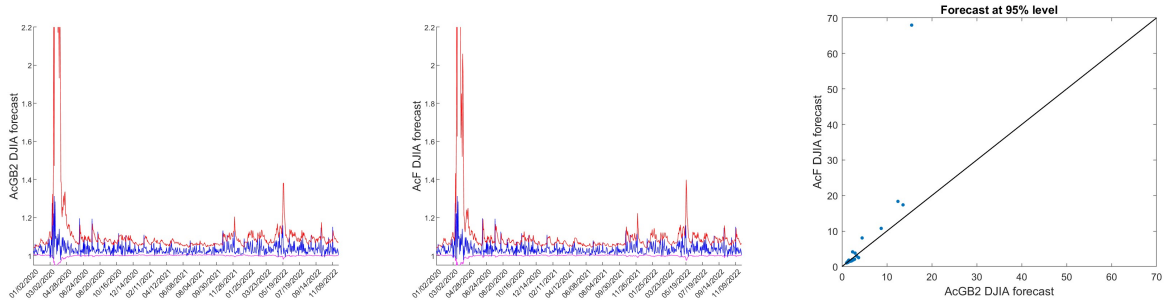


Figure 7: *Dow Jones 30: Predicted maxima values at the 95th and 5th percentiles from 1/1/2020 to 11/30/2022 (red line and magenta) against the real maxima values (blue line) for two models: the **AcGB2** model (left) and the AcF model (middle). The right panel illustrates AcF model predicted maxima values at the 95th percentiles against **AcGB2** model predicted maxima values at the 95th percentiles.*

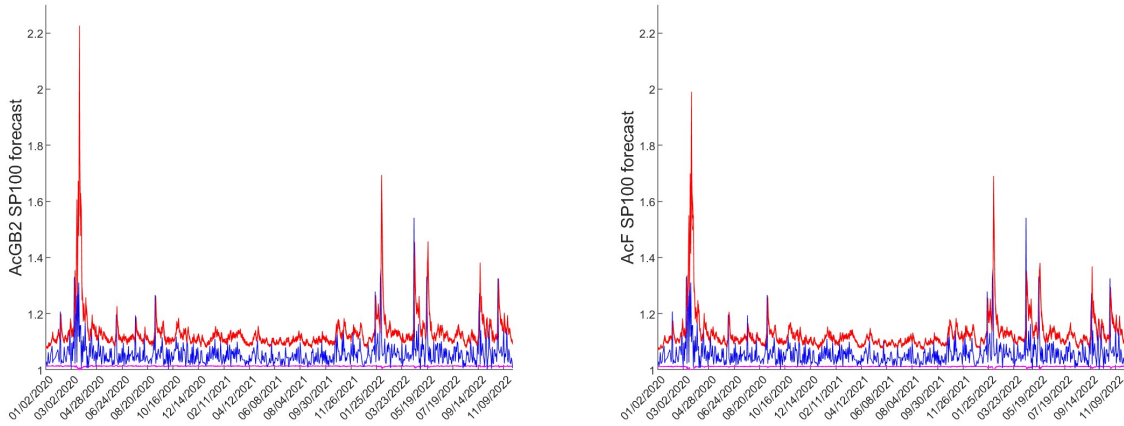


Figure 8: *S&P 100: Predicted maxima values at the 95th and 5th percentiles from 1/1/2020 to 11/30/2022 (red line and magenta) against the real maxima values (blue line) for two models: the **AcGB2** model (left) and the AcF model (right).*

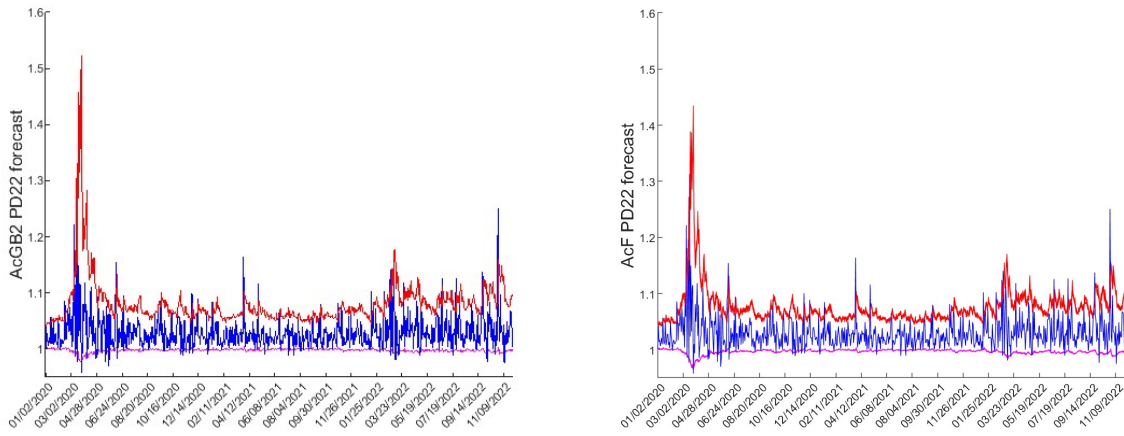


Figure 9: PD 22: *Predicted maxima values at the 95th and 5th percentiles from 1/1/2020 to 11/30/2022 (red line and magenta) against the real maxima values (blue line) for two models: the **AcGB2** model (left) and the AcF model (right).*

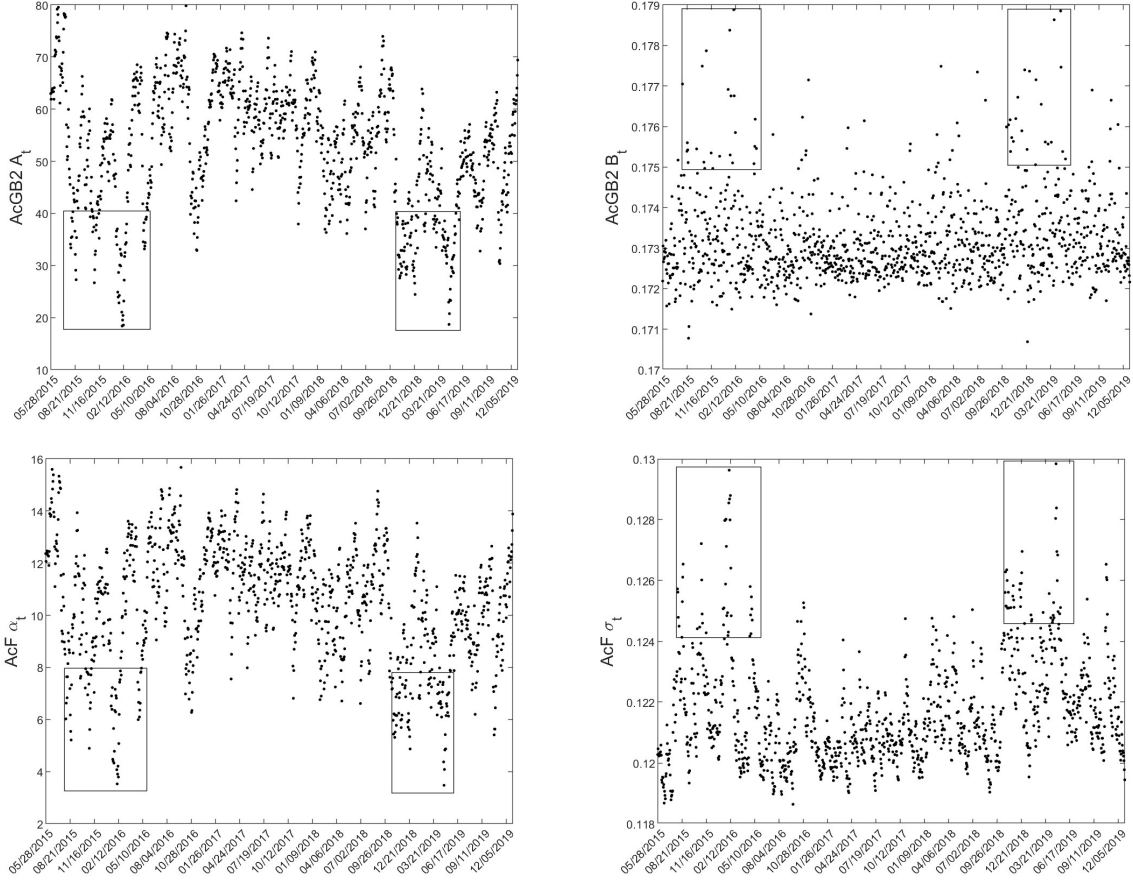


Figure 10: DJIA 30: (1) *Dynamics patterns of estimated A_t (upper left) and B_t (upper right) between 01/01/2015 and 12/31/2019, representing the estimated dynamic parameters of the **AcGB2** model.* (2) *Dynamics patterns of estimated α_t (bottom left) and σ_t (bottom right) over time, reflecting the estimated dynamic parameters of the AcF model.*

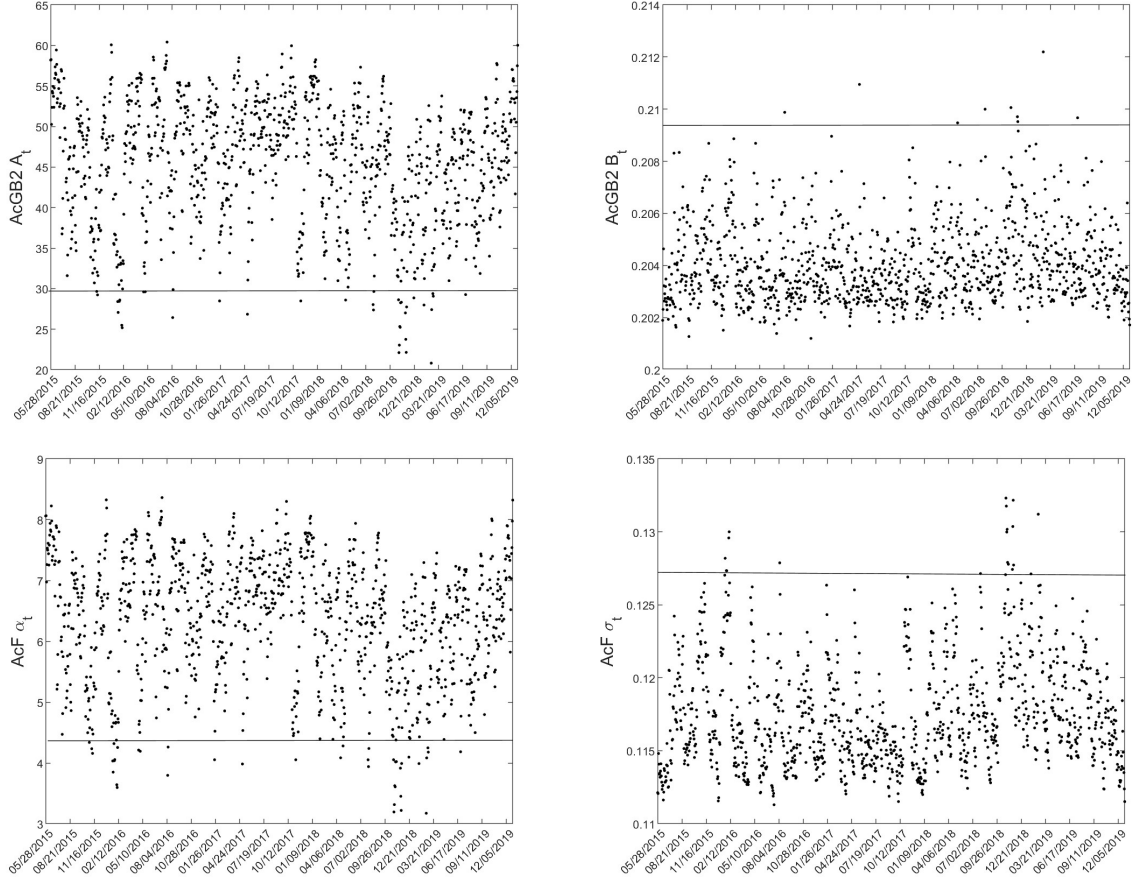


Figure 11: *S&P 100*: (1) *Dynamics patterns of estimated A_t (upper left) and B_t (upper right) between 01/01/2015 and 12/31/2019, representing the estimated dynamic parameters of the **AcGB2** model.* (2) *Dynamics patterns of estimated α_t (bottom left) and σ_t (bottom right) over time, reflecting the estimated dynamic parameters of the AcF model.*

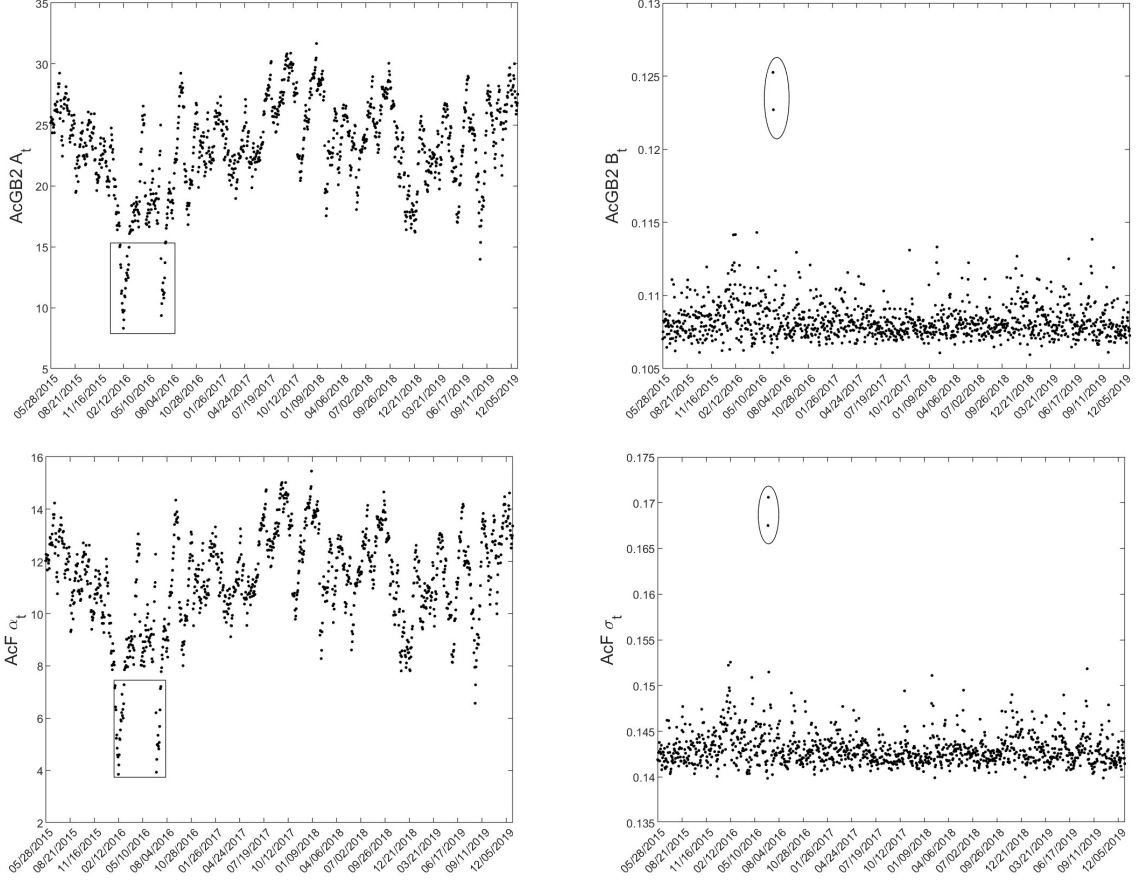


Figure 12: *Primary dealers: (1) Dynamics patterns of estimated A_t (upper left) and B_t (upper right) between 01/01/2015 and 12/31/2019, representing the estimated dynamic parameters of the **AcGB2** model. (2) Dynamics patterns of estimated α_t (bottom left) and σ_t (bottom right) over time, reflecting the estimated dynamic parameters of the AcF model.*

References

- Birkhoff, G. D. (1931). Proof of the ergodic theorem. *Proceedings of the National Academy of Sciences* 17 (12), 656–660.
- Chan, K. and H. Tong (1994). A note on noisy chaos. *Journal of Royal Statistical Society - Series B* 56 (2), 301–311.
- Dombry, C. (2015). Existence and consistency of the maximum likelihood estimators for the extreme value index within the block maxima framework. *Bernoulli* 21 (1), 420–436.
- Makelainen, T., K. Schmidt, and G. Styan (1981). On the existence and uniqueness of the maximum likelihood estimate of a vector-valued parameter in fixed-size samples. *Annals of Statistics* 9, 758–767.
- Smith, R. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika* 72 (1), 67–90.
- Wu, J.-W., W.-L. Hung, and H.-M. Lee (2000). Some moments and limit behaviors of the generalized logistic distribution with applications. *Proc. Natl. Sci. Counc. ROC(A)* 24 (1), 7–14.
- Zhao, Z., Z. Zhang, and R. Chen (2018). Modeling maxima with autoregressive conditional Fréchet model . *Journal of Econometrics* 207 (2), 325–351.