Autoregressive Conditional GB2 Models with Applications to Financial Time Series

by

Ning Fan

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Statistics)

at the

UNIVERSITY OF WISCONSIN–MADISON

2020

Date of final oral examination: 12/18/2020

The dissertation is approved by the following members of the Final Oral Committee:

Zhengjun Zhang, Professor, Statistics

Yazhen Wang, Professor, Statistics

Peng Shi, Associate Professor, Business

Nicolas Garcia Trillos, Assistant Professor, Statistics

Chunming Zhang, Professor, Statistics

Acknowledgments

First and foremost, I would like to express my sincere gratitude to my advisor, Professor Zhengjun Zhang. With his helpful advice, guiding and continuous support, I get the opportunity to discover more in statistics such as area of extreme value theory and time series. During tough times of developing the autoregressive conditional GB2 model, he is always by my side to give me encouragement and suggestions and let me keep enthusiasm and curiosity about research, analysis and new challenges. Not only as a advisor but also as a friend, he teaches me how to be a better person and it is my honor to be his Ph.D. student.

Besides, I would like to thank my thesis committee: Professor Yazhen Wang, Professor Peng Shi, Professor Nicolas Garcia Trillos and Professor Chunming Zhang. Their helpful guidance and invaluable ideas are really helpful for me to improve my thesis work.

Moreover, I would like to thank all the professors, faculty members and staff. During the first two years of my Ph.D. study, Professor Chunming Zhang, Dr. Derek Bean, Professor Jun Shao and Professor Yazhen Wang taught me the foundation of mathematical statistics. Then Professor Wei-Yin Loh, Professor Richard J. Chappell, Professor Zhengjun Zhang and professor Chunming Zhang guided me to learn different subareas and methods of statistics like Machine Learning, Regression Analysis, Multivariate Analysis, and Non-parametric Statistics. Professor Michel Newton and Professor Brian Yandell showed me how to apply statistical method in real world problems.

Also, I would also like to acknowledge my academic siblings in Professor Zhang's research group. Zifeng Zhao and Yuqing Xu have provided me suggestions and helps that widen my research. Working with all of them are full of happiness, we help each other during that time we build up friendships and our own research skills.

In adition, I would like to thank all my friends in Madison and all over the world. I am especially grateful to Yi Li, Tianjie Wang, Ziyue Wang, Jianchan Hu, Chengning Zhang and Bin Zhang for their accompany and encouragement. The five years of my Ph.D. pursuit is an important and pressure part of my life and through sharing happiness and sorrows with them this whole Ph.D life is full of enjoy.

Finally, I want to express my sincere thanks to my family. I would like to thank my parents Jianhua Fan and Shupin Huang for their emotional and financial support. They gave me all without any hesitation and their endless love supports me to overcome any challenges during research or daily life.

Contents

Abstract **xi**

1 Introduction **1**

2 Dynamics at the Both Tails Parameter and the Scale Parameter **6**

2.1 Model Specification 7

2.1.1 Proposed Model 7

2.1.2 Properties 8

2.1.3 Parameter Estimation and Asymptotic Properties 9

2.2 Simulation Study 11

2.2.1 Simulation of Different Data Size 11

2.2.2 Compare with the AcF Model 12

2.2.3 Deal with p and q 15

2.3 Real Data Applications 17

2.3.1 DJIA 17

2.3.1.1 Data Description and Model Fitting 17

2.3.1.2 Comparing with the AcF Model 20

2.3.2 S&P 100 24

2.3.2.1 Data Description and Model Fitting 24

2.3.2.2 Comparing with the AcF Model 26

2.3.3 Subgroups of S&P 100 28

2.3.3.1 Data Description and Model Fitting 28
2.3.3.2 Comparing with the AcF Model 32

3 Dynamics at the Right Tail Parameter and the Scale Parameter 38

3.1 Model Specification 39
3.2 Simulation Study 41
3.2.1 Simulation of Different Data Size 41
3.2.2 Compare with the AcF Model 43
3.3 Real Data Application 44
3.3.1 DJIA 44
3.3.2 S&P 100 48
3.3.3 Subgroups of S&P 100 50

4 Discussion 56

5 Appendix: Lemmas and Proofs 60

5.1 Proofs for Model with Parameters a and b Dynamic 61
5.1.1 Proof of Stationarity and Ergodicity 61
5.1.2 Proof of Consistency and Asymptotic Normality 67
5.1.3 Proof of Uniqueness 76
5.1.4 Proof of Lemmas 77
5.1.5 Transformation Between Distributions 83
5.2 Proofs for Model with Parameters q and b Dynamic 84
5.2.1 Proof of Stationarity and Ergodicity 84
5.2.2 Proof of Consistency, Asymptotic Normality, and Uniqueness 86

List of Tables

2.2.1 *Mean of MLE for 100 data sets with length to be 1000, 3000 and 5000. (Totally, there are 300 data sets.)* 11

2.2.2 *Standard deviation of MLE for 100 data sets with length to be 1000, 3000 and 5000. (Totally, there are 300 data sets.)* 12

2.2.3 *Statistical summary of two simulated data sets.* 13

2.2.4 *Estimated parameters of two models* 14

2.2.5 *P-values of K-S tests, Test 1 compares* $\{Y_t\}$ *from AcGB2 model with GB2 (1, 1, 2, 2) distribution; Test 2 compares* $\{Y_t\}$ *from AcF model with Fréchet (1, 1).* 14

2.2.6 *Estimated parameters for different setting of* p *and* q*: Method 1 with (*p*=2,* q*=2), Method 2 with (*p*=3,* q*=3) and Method 3 with* p*,* q *unknown).* 16

2.2.7 *P-values of K-S tests, Method 1 compares* $\{Y_t\}$ *with GB2 (1, 1, 2, 2) distribution; Method 2 compares* $\{Y_t\}$ *with GB2 (1, 1, 3, 3) and Method 3 compares* $\{Y_t\}$ *with GB2 (1, 1, 2.630, 3.910).* 16

2.2.8 *P-values of K-S tests, method 1 compares* $\{Y_t\}$ *with GB2 (1, 1, 0.368, 0.807) distribution; method 2 compares* $\{Y_t\}$ *with GB2 (1, 1, 10, 10) and method 3 compares* $\{Y_t\}$ *with GB2 (1, 1, 0.436, 0.916), where (0.436, 0.916) is the estimation of (*p*,* q*).* 17

2.3.1 *Statistical summary for maxima of DJIA data.* 19

2.3.2 *Estimated parameters with standard deviations of two models* 20

2.3.3 *Statistical summary for maxima of S&P 100 data.* 24

2.3.4 *Estimated parameters with standard deviations of two models* 26

2.3.5 *Statistical summary for maxima of subgroup of S&P 100 data contains sectors consumer discretionary, consumer staples and communication services.* 31

2.3.6 *Estimated parameters with standard deviations of two models* 32

3.2.1 *Mean of MLE for 100 data sets with length to be 1000, 3000 and 5000.* 42

3.2.2 *Standard deviation of MLE for 100 data sets with length to be 1000, 3000 and 5000.* 42

3.3.1 *Estimated parameters with standard deviations of the AcGB2 model with q and b dynamics for data of DJIA* 45

3.3.2 *Estimated parameters with standard deviation of the AcGB2 model with q and b dynamics for data of S&P 100* 48

3.3.3 *Estimated parameters with standard deviation of the AcGB2 model with q and b dynamics for data of subgroup of S&P 100 contains sectors consumer discretionary, consumer staples and communication services* 52

List of Figures

2.2.1 *Histogram and smooth lines (pink colored) for $\{Y_t\}$, and density curves (red colored) of GB2 (1, 1, 2, 2) (on the left) and Fréchet (1, 1) (on the right).* 14

2.3.1 *Change of maxima over time of DJIA data. The trend is shown by a regression line (blue line) with 95% confidence intervals (gray shade).* 18

2.3.2 *Histogram and density curve for maxima of DJIA data.* 19

2.3.3 *Change of estimated a_t and b_t over time, a_t and b_t are dynamic estimated parameters of the AcGB2 model. The trend is shown by a regression line (blue line) with 95% confidence intervals (gray shade).* 20

2.3.4 *Q-Q plots for the AcGB2 model and the AcF model. The left plot compares the real data with recovered maxima series with the AcGB2 model, while the right plot compares the real data with recovered maxima series with the AcF model.* 21

2.3.5 *Recovered dynamic scale parameter series (black line) with the daily average volatility given by the GARCH models (red line) for two models: the AcGB2 model (left) and the AcF model (right).* 22

2.3.6 *Predicted maxima values from November 25, 2019 to November 09, 2020 (black line) with 95$\%$ CI (range within gray dash lines) against the real maxima values (red line) for two models: the AcGB2 model (left) and the AcF model (right).* . . . 23

2.3.7 *Change of maxima over time of S&P 100 data. The trend is shown by a regression line (blue line) with 95% confidence intervals (gray shade).* 25

2.3.8 *Histogram and density curve for maxima of S&P 100 data.* 25

2.3.9 *Q-Q plots for the AcGB2 model and the AcF model. The left plot compares the real data with recovered maxima series with the AcGB2 model, while the right plot compares the real data with recovered maxima series with the AcF model.* 27

2.3.10 *Recovered dynamic scale parameter series (black line) with the daily average volatility given by the GARCH models (red line) for two models: the AcGB2 model (left) and the AcF model (right).* 28

2.3.11 *Predicted maxima values from May 27, 2020 to November 13, 2020 (black line) with 95% CI (range within gray lines) against the real maxima values (red line) for two models: the AcGB2 model (left) and the AcF model (right).* 29

2.3.12 *Change of maxima over time of subgroup of S&P 100 data contains sectors of consumer discretionary, consumer staples and communication services. The trend is shown by a regression line (blue line) with 95% confidence intervals (gray shade).* . 30

2.3.13 *Histogram and density curve for maxima of subgroup of S&P 100 data contains sectors consumer discretionary, consumer staples and communication services.* . . 31

2.3.14 *Change of estimated a_t and b_t over time for data of subgroup of S&P 100 contains sectors consumer discretionary, consumer staples and communication services. a_t and b_t are dynamic estimated parameters of the AcGB2 model. The trend is shown by a regression line (blue line) with 95% confidence intervals (gray shade).* 32

2.3.15 *Q-Q plots for the AcGB2 model and the AcF model. The left plot compares the real data with recovered maxima series with the AcGB2 model, while the right plot compares the real data with recovered maxima series with the AcF model.* 33

2.3.16 *Recovered dynamic scale parameter series (black line) with the daily average volatility given by the GARCH models (red line) for two models: the AcGB2 model (left) and the AcF model (right).* 34

2.3.17 *Predicted maxima values from May 27, 2020 to November 13, 2020 (black line) with 95% CI (range within gray lines) against the real maxima values (red line) for two models: the AcGB2 model (left) and the AcF model (right).* 35

2.3.18 *Recovered dynamic scale parameter series (black line) with the daily average volatility given by the GARCH models (red line) for two models: the AcGB2 model (left) and the AcF model (right).* 36

2.3.19 *Recovered dynamic scale parameter series (black line) with the daily average volatility given by the GARCH models (red line) for two models: the AcGB2 model (left) and the AcF model (right).* 37

3.2.1 *Q-Q plots for the AcGB2 model with q and b dynamics and the AcF model. The left plot is generated with the method of simulating with the AcF model and fitting with the AcGB2 model with q and b dynamics, while the right plot is generated with the method of simulating with the AcGB2 model with q and b dynamics and fitting with the AcF model.* 44

3.3.1 *Change of estimated q_t and b_t over time for data of DJIA, q_t and b_t are dynamic estimated parameters of the AcGB2 model. The trend is shown by a regression line (blue line) with 95% confidence intervals (gray shade).* 46

3.3.2 *Q-Q plots for the AcGB2 model with q and b dynamics for data of DJIA.* 46

3.3.3 *Recovered dynamic scale parameter series (black line) with the daily average volatility given by the GARCH models (red line) for the AcGB2 model with q and b dynamics.* 47

3.3.4 *Predicted maxima values from November 25, 2019 to November 09, 2020 (black line) with 95% CI (range within gray lines) against the real maxima values (red line) for the AcGB2 model with q and b dynamics.* 48

3.3.5 *Change of estimated q_t and b_t over time for data of S&P 100, q_t and b_t are dynamic estimated parameters of the AcGB2 model. The trend is shown by a regression line (blue line) with 95% confidence intervals (gray shade).* 49

3.3.6 *Q-Q plots for the AcGB2 model with q and b dynamics for data of S&P 100.* 50

3.3.7 *Recovered dynamic scale parameter series (black line) with the daily average volatility given by the GARCH models (red line) for the AcGB2 model with q and b dynamics.* 51

3.3.8 *Predicted maxima values from May 27, 2020 to November 13, 2020 (black line) with 95% CI (range within gray lines) against the real maxima values (red line) for the AcGB2 model with q and b dynamics.* 52

3.3.9 *Change of estimated q_t and b_t over time for data of subgroup of S&P 100 contains sectors consumer discretionary, consumer staples and communication services. q_t and b_t are dynamic estimated parameters of the AcGB2 model with q and b dynamics. The trend is shown by a regression line (blue line) with 95% confidence intervals (gray shade).* 53

3.3.10 *Q-Q plots for the AcGB2 model with q and b dynamics for data of subgroup of S&P 100 contains sectors consumer discretionary, consumer staples and communication services.* 53

3.3.11 *Recovered dynamic scale parameter series (black line) with the daily average volatility given by the GARCH models (red line) for the AcGB2 model with q and b dynamics.* 54

3.3.12 *Predicted maxima values from May 27, 2020 to November 13, 2020 (black line) with 95% CI (range within gray lines) against the real maxima values (red line) for the AcGB2 model with q and b dynamics.* 55

Abstract

In order to model the time varying behavior of maxima in financial time series, we introduce a novel dynamic generalized beta distribution of second kind (GB2) framework. The proposed autoregressive conditional GB2 (AcGB2) model will be fitted to the high skewed and long tailed financial time series data. The time dependence among the maxima is characterized in the parameter dynamics of the GB2 distribution. One merit property of GB2 distributions is that many distributions including extreme value distributions can be approximated by GB2 distributions with different parameter values. Incorporating the dynamics on the parameters results in the time series of maxima possessing distribution dynamics, e.g., observations can be at extreme levels and at less extreme levels (intermediate extremes). As a result, the newly proposed modeling framework has greater flexibility than extreme value distributions and can better be fitted to real data. The proposed thesis work proves the existence of stationary and ergodic solution of the new time series model under mild conditions of the parameters. Statistical estimation method (conditional maximum likelihood estimation) will be developed. The consistency, the asymptotic normality, and the uniqueness of estimators will be derived. The thesis uses simulation examples to demonstrate the model properties and the efficiency of statistical inference method. Real data analysis will be applied to two data sets, the maxima of negative logarithm returns from 30 stocks in Dow Johns Industrial Average (DJIA), and the maxima of negative logarithm returns from 101 stocks in S&P 100. The thesis also compares the AcGB2 method inference with the existing method AcF in terms of their performance in real applications.

Chapter 1

Introduction

It is widely noticed that many statistical models cannot be directly fit to the high skewed and long tailed data. The skewed data will have mean different dramatically from median and there will be a long tail that extends to the left or right. For example, it is common knowledge that the financial data are heavy-tailed. Since the data of tailed region will act like outliers for many statistical models and affect the performance of the models, efficiently treating high skewed and long tailed data is required in model fitting and inference.

In the financial market literature, extreme events for asset returns, insurance and many other areas with highly skewed observations are widely studied in the filed of risk management. In the extreme value theory, there are two fundamental approaches: the block maxima (BM) method and the peak-over-threshold (POT) method. Back to Fisher and Tippett (1928) and Gnedenko (1943), the extreme value distribution (e.g., Gumbel distribution, Weibull distribution and Fréchet distribution) started to be applied to analyze maxiama. The BM method divides the observation period into periods of equal size and find the maximum observation in each period, then statistical methods for the extreme value distributions are applied to those observations (see Gumbel (1958)). The BM method not only retains high observations but also some lower observations, while the POT method uses all relevant high observations and is often considered as more powerful. For the POT method, the observations are selected by the rule that they need to exceed a certain high threshold. It assumes that the exceedances over high threshold converge to a generalized Pareto distribution (GPD) (see Balkema and de Haan (1974) and Lechner et al. (1993)). The step of selecting an appropriate threshold value is critical in POT method, since the estimated extreme events are sensitive to the changes of the value. The methods of selecting a suitable threshold value were discussed by Goda (2000) and Li et al. (2012). The estimation approaches and properties based on these two methods were widely studied: Dombry (2015) and Zhou (2010)) used the maximum likelihood (ML) estimators for BM mehod; Hosking et al. (1985) used the the probability weighted moment (PWM) estimators for BM method; Smith (1985) and Grimshaw (2017) applied MLE for POT method; Hosking and Wallis (1987) applied PWM for POT method. Applications of these two methods in many areas were also tried and deeply researched (e.g., Ferreira and Soares (1998) and Laurini and Tawn (2009)). The relationship and comparisons of the two methods were also widely discussed. Based

on simulated data, Cunnane (1973) showed that while the number of exceedances larger than the number of blocks POT is more efficient than BM; Ferreira and de Haan (2015) showed BM outperform POT through theoretical comparison for the reason that BM tends to have smaller asymptotic variances of quantile estimators and extreme value than POT. The BM method and POT method can perform well when the length of data is long enough and the conditions of extreme value theorem are satisfied. For some time series data, the required conditions for the extreme value theorem to hold may not hold. As a result, new time series models are needed.

To provide an excellent description for long-tailed and highly skewed data, the generalized beta distribution of second kind (GB2) has drawn much attention. In the literature, McDonald and Xu (1995) first introduced and discussed the application of generalized beta distribution. The GB2 distribution is a four-parameter distribution with one scale parameter b, and three shape parameters a, p, and q (a controls both tails, p controls left tail and q controls right tail), allowing the distribution to form many different shapes including J-shaped, bell-shaped, long-tailed, light-tailed, right-skewed and left-skewed. Corrado (2001) applied the GB2 distribution to option pricing. The GB2 distribution has also been applied in actuarial science by Kleiber and Kotz (2003). Frees and Valdez (2008) used the four parameter distribution GB2 to insurance claims. Papalexiou and Koutsoyiannis (2012) showed that the GB2 distribution is appropriate for worldwide rainfall data, and Chen and Singh (2017) showed that it fits well for flood frequency analysis. The GB2 distribution was extended to MGB2 copula in order to describe multivariate data by Yang et al. (2011). The applications of MGB2 copula included pricing synthetic CDO (Cui and Ma (2014)) and modeling guaranteed unitized participating life insurance (Zheng et al. (2019)). The reason that using GB2 to model long-tailed and highly skewed data is that it is connected with many other distributions including extreme value distributions. These distributions can be treated as the limiting or special cases of GB2. For example, when fixing the second shape parameter p to be 1, we have the Burr type 12 distribution; when letting the third shape parameter q converge to infinity, we have the generalized gamma distribution. McDonald and Xu (1995) summarized the relationship between GB2 family distributions in the form of distribution trees, in which common distributions such as the generalized gamma (GG), Burr type 3, Burr type 12, Dagum, lognormal, Weibull, gamma, Lomax,

F, Fisk or Rayleigh, chi-square, half-normal, half-Student's t, exponential, asymmetric log-Laplace, log-Laplace, power function, and the log-logistic are included. By the reason that the GB2 distribution family includes other extreme distributions as special or limiting cases, we can apply them to long-tailed and highly skewed data. Moreover, comparing with other extreme distributions, the GB2 distribution is a more flexible family. As such, while dealing with a data set that it is not highly skewed, the GB2 model which can provide sufficient flexibility while fitting a large variety of data sets should outperform other extreme distributions.

While dealing with extreme events in time series, another challenge appears. The volatility clustering behavior is commonly observed, which means the behavior and structure of the extreme events may also change as time goes by. For non-stationary time series, the volatility exists and changes for different periods; even for stationary time series there exists volatility clustering. Also need to mention that the data in financial time series are correlated and heteroskedastic. However, applying the standard extreme value theory to financial filed is restricted to identically independent distributed series (see Diebold et al. (1998)). To accommodate the dynamics of extreme events in financial time series, recently there are multiple studies with different approaches. Chavez-Demoulin et al. (2014) proposed a non-parametric extension of the classical Peak-Over-Threshold method to fit the time varying volatility. Kelly and Jiang (2014) provided a panel estimation approach to capture the time varying tail risk of returns for individual firms. Zhang and Schwaab (2017) used the score of the predictive log-likelihood to drive the tail shape dynamics for Generalized Pareto Distribution (GPD) and proposed a novel observation-driven model to study the sovereign bond yields at a high frequency. Since volatility dynamics can be directly related to the changing parameters of extremal distributions, another way to properly model the extreme events in financial time series is not sticking to a distribution with fix parameters. In other words, by introducing dynamics to the parameters of distribution (make the parameters vary according to time), we can also capture the dynamic behavior of time series data. Bali and Weinbaum (2007) detected time series variation in extreme value distributions and introduced a conditional extreme value volatility estimator using high-frequency returns. Zhao et al. (2018) and Mao and Zhang (2018) applied Fréchet distribution with time-varying parameters to model time-varying behavior of maxima in financial time series.

Taking into account the discussions and researches outlined above, in this thesis, we introduce a novel dynamic GB2 framework for modeling the time-varying behavior of maxima in financial time series. With the variations of parameters of GB2 through time, we can capture the time-varying behavior of the time series data. In summary, GB2 not only can model the high-skewed and long tailed financial data, accurately, it is also more flexible.

The rest of the paper is organized as follows. Chapter 2 introduces the autoregressive conditional GB2 (AcGB2) model with the two tail shape parameter a and the scale parameter b to be dynamic. The properties of stationarity and ergodicity are studied. Also the parameter estimation method is discussed. Then the simulation study is performed with purposes including analyzing the effect of the sample size to estimation accuracy, demonstrating the flexibility of the AcGB2 model by comparing with other models and discussing more about two non-time-varying parameters of the GB2 distribution. After that the real data application is conducted where stocks returns of companies involved in Dow Jones Industrial Average (DJIA), companies involved in S&P 100 or subgroups of S&P 100 are analyzed with our model and the goodness of fit is tested. Then in Chapter 3, different pair of parameters of the GB2 distribution is chosen to be time-varying and they are the right tail shape parameter q and the scale parameter b. Also, estimation methods, simulation studies and real data analysis are included. Conclusions and discussions are provided in Chapter 4. Proofs of theorems can be found in the appendix of Chapter 5.

Chapter 2

Dynamics at the Both Tails Parameter and the Scale Parameter

2.1 Model Specification

2.1.1 Proposed Model

The data we are trying to analyze is a time series of maxima X_t ($X_t = \max_{1\leq i\leq m} X_{it}$, where $\{X_{it}\}_{i=1}^{m}$ are m time series). The conventional time series analysis methods contain ARMA model or ARIMA model (see Shuway and Stoffer (2017)). However when dealing with maxima of financial time series, where the data is highly skewed, long tailed and often exhibits the volatility clustering property (large changes tend to cluster together), the conventional methods do not fit well or provide sufficient economic explanation (see Teyssière and Kirman (2007) and Rootzen and Finkenstadt (2003)). To fit maxima and capture the relationship among them, we propose a novel dynamic generalized beta distribution of second kind (GB2) framework. In our setting we consider model maxima time series X_t by GB2 with parameters (μ, a_t, b_t, p, q):

$$X_t = \mu + b_t Y_t^{\frac{1}{a_t}},$$

where μ is location parameter, a_t is a shape parameter that controls both tails and b_t is a scale parameter; $\{Y_t\}$ is a sequence of $i.i.d.$ random variables of beta prime distribution with parameters p and q (It is equivalent to GB2 (a = 1, b = 1, p, q). The proof can be found in 5.1.5.

Remark 2.1.1. If we have Y_t follow GB2 (a = 1, b = 1, p, q), then it can be proved that $X_t - \mu$ (= $b_t Y_t^{\frac{1}{a_t}}$) follows GB2 (a_t , b_t, p, q). The proof can be found in 5.1.5.

Since ($X_t - \mu$) follows GB2 (a_t, b_t, p, q), we can find p.d.f of X_t:

$$f_{X_t}(x_t) = \frac{1}{B(p,q)} \frac{a_t (x_t-\mu)^{a_t p-1}}{b_t^{a_t p}(1+(\frac{x_t-\mu}{b_t})^{a_t})^{p+q}}, X_t > \mu, \tag{2.1.1}$$

where $B(p, q)$ is a beta function with parameters (p, q).

The time-varying parameters are assumed to have the dynamic pattern:

$$\log a_t = \alpha_0 + \alpha_1 \log a_{t-1} + \eta_1(X_{t-1}), \tag{2.1.2}$$

$$\log b_t = \beta_0 + \beta_1 \log b_{t-1} + \eta_2(X_{t-1}), \tag{2.1.3}$$

where α_0, β_0, $\mu \in \mathbb{R}$, $0 \leq \alpha_1, \beta_1 < 1$ $\alpha_1 \neq \beta_1$ and $\alpha_2, \beta_2 > 0$; $\eta_1(X_{t-1})$ and $\eta_2(X_{t-1})$ are two continuous monotone functions of X_{t-1}.

The reason that we add these terms to the dynamic equation is that, the time-varying parameters a_t and b_t depend not only on their own history but also on the history of X_t. The choice of the two functions is not unique. Considering the boundedness, monotonicity, differentiability and interpretable, in this paper we choose the simple exponential function which is widely applied (e.g. Lundbergh et al. (2003), Hall et al. (2016), and Zhao et al. (2018)). Similar to Zhao et al. (2018), we set η functions to be exponential functions, we have:

$$\log a_t = \alpha_0 + \alpha_1 \log a_{t-1} + \alpha_2 \exp(-\alpha_3 X_{t-1}), \tag{2.1.4}$$

$$\log b_t = \beta_0 + \beta_1 \log b_{t-1} - \beta_2 \exp(-\beta_3 X_{t-1}), \tag{2.1.5}$$

or

$$\log a_t = \alpha_0 + \alpha_1 \log a_{t-1} + \alpha_2 \exp(-\alpha_3(\mu + b_{t-1} Y_{t-1}^{\frac{1}{a_{t-1}}})), \tag{2.1.6}$$

$$\log b_t = \beta_0 + \beta_1 \log b_{t-1} - \beta_2 \exp(-\beta_3(\mu + b_{t-1} Y_{t-1}^{\frac{1}{a_{t-1}}})), \tag{2.1.7}$$

where $\alpha_3, \beta_3 > 0$.

2.1.2 Properties

When a noise has infinite support, usually we assume that strongly non-linear time series models would not be stationary. However, for finite support, the result can be different. To show that the Markov chain of $\{a_t, b_t\}$ satisfies stationary and ergodic conditions, we follow the idea of Chan and Tong (1994) (see also Zhao et al. (2018)). They showed that under appropriate conditions,

a strongly non-linear time series model is stationary and ergodic in a neighborhood of a compact attractor which may have complicated dynamics.

Theorem 2.1.1. *(Stationarity and ergodicity) For AcGB2 with* α_0, β_0, $\mu \in \mathbb{R}$, $0 \leq \alpha_1, \beta_1 < 1$, $\alpha_1 \neq \beta_1$ *and* $\alpha_2, \alpha_3, \beta_2, \beta_3, p, q > 0$, *the process* $\{a_t, b_t\}$ *is stationary and geometrically ergodic.*

The proof of Theorem 2.1.1 can be found in 5.1.1.

Remark 2.1.2. Since we have relation between X_t and a_t, b_t that $X_t = \mu + b_t Y_t^{\frac{1}{a_t}}$, once the process $\{a_t, b_t\}$ is stationary and ergodic, $\{X_t\}$ is stationary and ergodic.

2.1.3 Parameter Estimation and Asymptotic Properties

The parameters we try to estimate are (μ, α_0, α_1, α_2, α_3, β_0, β_1, β_2, β_3, p, q). Define $\theta = (\mu, \alpha_0, \alpha_1, \alpha_2, \alpha_3, \beta_0, \beta_1, \beta_2, \beta_3, p, q)$. Let $\theta^0 = (\mu^0, \alpha_0^0, \alpha_1^0, \alpha_2^0, \alpha_3^0, \beta_0^0, \beta_1^0, \beta_2^0, \beta_3^0, p^0, q^0)$ be the true estimator. Then, the log-likelihood function with $\{X_t\}_{t=1}^n$ and true initial value a_1', b_1' given is:

$$
\begin{aligned}
L_n^0(\theta) &= \frac{1}{n} \log \prod_{t=1}^{n} \frac{1}{B(p,q)} \frac{{a_t}'(x_t - \mu)^{{a_t}'p-1}}{({b_t}')^{a_t p}(1 + (\frac{x_t - \mu}{{b_t}'})^{{a_t}'})^{p+q}} \\
&= \frac{1}{n} \sum_{t=1}^{n} \Bigg[- \log(B(p,q)) + \log({a_t}') - {a_t}' p \log({b_t}') + ({a_t}'p - 1)\log(x_t - \mu) \\
&\quad - (p+q)\log[1 + (\frac{x_t - \mu}{{b_t}'})^{{a_t}'}] \Bigg],
\end{aligned}
$$

where a_t, b_t are generated with θ and arbitrary initial values a_1, b_1; a_t', b_t' are generated with θ and true initial values a_1', b_1'; a_t^0, b_t^0 are generated with θ^0 and arbitrary initial values a_1, b_1, while $a_t'^0$, $b_t'^0$ are generated with θ^0 and true initial values a_1', b_1'. For convenience, let $l_t^0(\theta) = -\log(B(p,q)) + \log({a_t}') - {a_t}' p \log({b_t}') + ({a_t}'p - 1)\log(x_t - \mu) - (p+q)\log[1 + (\frac{x_t - \mu}{{b_t}'})^{{a_t}'}]$. However the difficulty is that the true initial values are unknown in reality. So we focus on estimating the parameter θ without knowing a_1' and b_1', which means we want to find the local maximizer of log-likelihood

function with $\{X_t\}_{t=1}^n$ and arbitrary initial value a_1, b_1. Let's define this log-likelihood function as:

$$\begin{aligned} L_n(\theta) &= \frac{1}{n}\sum_{t=1}^{n} l_t(\theta) \\ &= \frac{1}{n}\sum_{t=1}^{n}\Bigg[-\log(B(p,q)) + \log(a_t) - a_t p\log(b_t) + (a_t p - 1)\log(x_t - \mu) \\ &\quad - (p+q)\log[1 + (\frac{x_t-\mu}{b_t})^{a_t}]\Bigg]. \end{aligned}$$

We will show in the Theorems 2.1.2, Theorems 2.1.3 and Theorems 2.1.4 that the local maximizer is consistent, asymptotic normal, and unique, and it does not matter that the true initial value is not given.

Theorem 2.1.2. *(Consistency) There exists a local maximizer $\widehat{\theta_n}$ of $L_n(\theta)$ for $\{X_t\}_{t=1}^n$ from AcGB2 with α_0, β_0, $\mu \in \mathbb{R}$, $0 \le \alpha_1, \beta_1 < 1$, $\alpha_1 \ne \beta_1$, $\alpha_2, \alpha_3, \beta_2, \beta_3, p, q > 0$ and true parameter θ^0 within a compact space, such that $\widehat{\theta_n} \xrightarrow{p} \theta^0$, $\|\widehat{\theta_n} - \theta^0\| \le \delta_n$, for $\delta_n \searrow 0$, $n\delta_n \to +\infty$, $\frac{1}{\delta_n}(\min\{X_i\}_{i=1}^t - \mu^0) \to +\infty$, where $L_n(\theta)$ is log-likelihood function with θ and arbitrary initial value a_1, b_1.*

The proof of Theorem 2.1.2 can be found in 5.1.2. It follows from the idea of Smith (1985). To prove that a continuously differentiable real-valued function has a local maximum x, we can show the scale product of x and the gradient vector of the function is negative whenever $\mid x \mid = 1$.

Theorem 2.1.3. *(Asymptotic Normality) The local maximizer $\widehat{\theta_n}$ of $L_n(\theta)$ for $\{X_t\}_{t=1}^n$ from AcGB2 with α_0, β_0, $\mu \in \mathbb{R}$, $0 \le \alpha_1, \beta_1 < 1$, $\alpha_1 \ne \beta_1$, $\alpha_2, \alpha_3, \beta_2, \beta_3, p, q > 0$ and true parameter θ^0 within a compact space, such that $\widehat{\theta_n} \to \theta^0$, $\|\widehat{\theta_n} - \theta^0\| \le \delta_n$, for $\delta_n \searrow 0$, $n\delta_n \to +\infty$, $\frac{1}{\delta_n}(\min\{X_i\}_{i=1}^t - \mu^0) \to +\infty$, is asymptotic normal. Equivalently, $\sqrt{n}(\widehat{\theta_n} - \theta^0) \xrightarrow{d} N\left(0, \frac{1}{-E_{\theta^0}(\frac{\partial^2 l_t^0(\theta^0)}{\partial\theta\partial\theta^T})}\right)$.*

The proof can be found in 5.1.2, the prof starts with a single parameter μ, then extends to the vector of parameters θ that we are interested.

Remark 2.1.3. Define $m_{\theta_i\theta_j} = -E_{\theta^0}(\frac{\partial^2 l_t^0(\theta^0)}{\partial\theta_i\partial\theta_j})$. Since $E_{\theta^0}(\frac{\partial l_t^0(\theta^0)}{\partial\theta}) = 0$, $Var_{\theta^0}(\frac{\partial l_t^0(\theta^0)}{\partial\theta}) = E_{\theta^0}(\frac{\partial l_t^0(\theta^0)}{\partial\theta}) - E_{\theta^0}(\frac{\partial^2 l_t^0(\theta^0)}{\partial\theta\partial\theta^T}) = -E_{\theta^0}(\frac{\partial^2 l_t^0(\theta^0)}{\partial\theta\partial\theta^T})$. So $m_{\theta_i\theta_j}$ is a value in matrix $Var_{\theta^0}(\frac{\partial l_t^0(\theta^0)}{\partial\theta})$.

Theorem 2.1.4. *(Uniqueness) Under conditions in Theorem 2.1.1, there is an asymptotic unique MLE over M_n, where $M_n = \{\theta \in \Theta | \mu \leq c \min_{1\leq t\leq n} X_t + (1-c)\mu_0\}$; Θ is a compact set of $\{\theta | \alpha_0, \beta_0, \mu \in \mathbb{R}, 0 \leq \alpha_1, \beta_1 < 1, \alpha_1 \neq \beta_1, \alpha_2, \alpha_3, \beta_2, \beta_3, p, q > 0\}$; $0 < c < 1$.*

The proof can be found in 5.1.3.

2.2 Simulation Study

2.2.1 Simulation of Different Data Size

In this section, we simulate data from the AcGB2 model with a and b dynamic with different sample sizes and compare the results. We simulate with the parameters θ = (μ, α_0, α_1, α_2, α_3, β_0, β_1, β_2, β_3, p, q) to be (-0.00714, -0.318, 0.904, 0.460, 7.180, -1.670, 0.390, -0.675, 9.170, 2, 2). We fit the AcGB2 model and get MLE. To compare the estimators for different data length, we simulate 100 data sets each time with length T to be 1000, 3000 and 5000 respectively. Then for each simulation we get one estimator set, so we will have 100 estimator sets for each data length. We compare the means and standard deviations. The results are shown below.

$Parameter$	μ	α_0	α_1	α_2	α_3	β_0
$True\,Value$	-0.00714	-0.318	0.904	0.460	7.180	-1.670
$Mean\,of$ $T = 1000$	-0.00577	-0.324	0.908	0.481	10.316	-1.672
$Mean\,of$ $T = 3000$	-0.00669	-0.335	0.902	0.481	8.408	-1.654
$Mean\,of$ $T = 5000$	-0.00712	-0.315	0.903	0.462	7.890	-1.705
$Parameter$	β_1	β_2	β_3	p	q	
$True\,Value$	0.390	-0.675	9.170	2.000	2.000	
$Mean\,of$ $T = 1000$	0.403	-0.632	12.977	1.715	1.893	
$Mean\,of$ $T = 3000$	0.400	-0.662	10.819	1.989	2.018	
$Mean\,of$ $T = 5000$	0.385	-0.668	10.425	2.000	1.988	

Table 2.2.1: *Mean of MLE for 100 data sets with length to be 1000, 3000 and 5000. (Totally, there are 300 data sets.)*

$Parameter$	μ	α_0	α_1	α_2	α_3	β_0
$S.D.\ of$ $T = 1000$	1.45E-03	1.56E-01	3.13E-02	1.74.E-01	6.75E+00	7.50E-01
$S.D.\ of$ $T = 3000$	3.13E-03	1.37E-01	1.92E-02	1.43E-01	5.00E+00	2.03E-01
$S.D.\ of$ $T = 5000$	7.31E-04	7.66E-02	9.90E-03	8.01E-02	1.92E+00	1.39E-01
$Parameter$	β_1	β_2	β_3	p	q	
$S.D.\ of$ $T = 1000$	2.02E-01	2.05E-01	5.65E+00	6.53E-01	4.67E-01	
$S.D.\ of$ $T = 3000$	2.60E-02	1.46E-01	2.87E+00	6.95E-01	5.81E-01	
$S.D.\ of$ $T = 5000$	2.73E-02	1.17E-01	2.96E+00	2.84E-01	2.33E-01	

Table 2.2.2: *Standard deviation of MLE for 100 data sets with length to be 1000, 3000 and 5000. (Totally, there are 300 data sets.)*

From Table 2.2.1 we can see all sets of estimated values close to the true parameters. For data with larger length, the results are better which means most of the estimated parameter values are closer to true ones. Based on Table 2.2.2, estimated values for the data with larger length will have smaller standard deviations. We have observed that when the data length is large enough, the MLE is accurate and stable. To verify the conclusion that the MLE is appropriate we may need further work and test. These will be shown latter.

2.2.2 Compare with the AcF Model

In this section we compare our AcGB2 model with autoregressive conditional Fréchet (AcF) model. The AcF model is one of the time-varying generalized extreme value distribution model that can be used to fit time series data of maxima. In the AcF model the Fréchet distribution is chosen and the two parameters of it is set to be dynamic. The model of AcF is shown below:

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

where $\{Q_t\}$ is maxima and $\{Y_t\}$ is a sequence of $i.i.d.$ unit Fréchet random variables. More details can be found in Zhao et al. (2018). The way to fulfill the comparison is: first, we simulate data use the AcF model, then apply the AcGB2 model and estimate parameters. To check that whether the model fit well, we recover $\{Y_t\}$ based on the equation $Y_t = (\frac{X_t - \widehat{\mu}}{b_t})^{a_t}$ (since we have $X_t = \mu + b_t Y_t^{\frac{1}{a_t}}$ in our model), where $\{X_t\}$ is from the simulation, $\widehat{\mu}$ is one of the estimated parameter, and $\{a_t\}$, $\{b_t\}$ can be recovered by the autoregressive equations:

$$\begin{cases} \log a_t = \widehat{\alpha_0} + \widehat{\alpha_1} \log a_{t-1} + \widehat{\alpha_2} \exp(-\widehat{\alpha_3}(X_{t-1})), \\ \log b_t = \widehat{\beta_0} + \widehat{\beta_1} \log b_{t-1} - \widehat{\beta_2} \exp(-\widehat{\beta_3}(X_{t-1})), \end{cases}$$

with estimated parameters for (α_0, α_1, α_2, α_3, β_0, β_1, β_2, β_3). According to the model, $\{Y_t\}$ should follow GB2 (a = 1, b = 1, p, q). For the reason that we need $\{Y_t\}$ to be fixed distribution with known parameters to fulfill the following test, we assume p = 2 and q = 2 to be known. More discussions about how to deal with p and q is covered in the next section. We test the goodness of fit with Kolmogorov-Smirnov (K-S) test which has the null hypothesis that two samples are drawn from the same continuous distribution. The K-S test is used to decide if a sample comes from a population with a specific distribution based on the empirical distribution function (ECDF) (check Chakravarti et al. (1967) for details). Then for comparison, the next step is to simulate data with the AcGB2 model, then apply the AcF model to estimate parameters. Similarly, we recover $\{Y_t\}$, but test whether it follows standard Fréchet distribution (with parameters (1, 1)). Table 2.2.3 shows the statistical summary of two simulated data from the AcF model and the AcGB2 model respectively. Through the table, we can see that for both data sets the mean is bigger than the median, so both data sets are skewed. The results of estimated parameters of the two models are shown in Table 2.2.4.

$Simulation\ Method$	Min	$1st\ Quantile$	$Median$	$Mean$	$3rd\ Quantile$	Max
$AcGB2$	-0.00658	0.0114	0.0187	0.0255	0.0290	0.300
AcF	-0.00841	0.0111	0.0189	0.0224	0.0287	0.178

Table 2.2.3: *Statistical summary of two simulated data sets.*

We draw the histograms and smooth lines (pink colored) of recovered $\{Y_t\}$ for both models

Fitting Model	μ	α_0	α_1	α_2	α_3	β_0	β_1	β_2	β_3
AcGB2	-0.00886	-0.273	0.857	0.489	9.790	-2.840	0.0502	-0.661	6.370
	μ	γ_0	γ_1	γ_2	γ_3	β_0	β_1	β_2	β_3
AcF	-0.0602	-0.0332	0.867	0.383	13.200	-1.420	0.429	-0.0932	22.700

Table 2.2.4: *Estimated parameters of two models* .

(AcGB2 on the left and AcF on the right). For checking the goodness of fit of $\{Y_t\}$, in the graph we also draw density curve (red colored) of the true distributions (GB2 (1, 1, 2, 2) on the left and Fréchet (1, 1) on the right). According to the two graphs in Figure 2.2.1, it seems both models work fine: the pink line is close to the red line for both graphs. It's also worth to mention that for the left graph the two lines are even closer, and they almost overlap with each other.

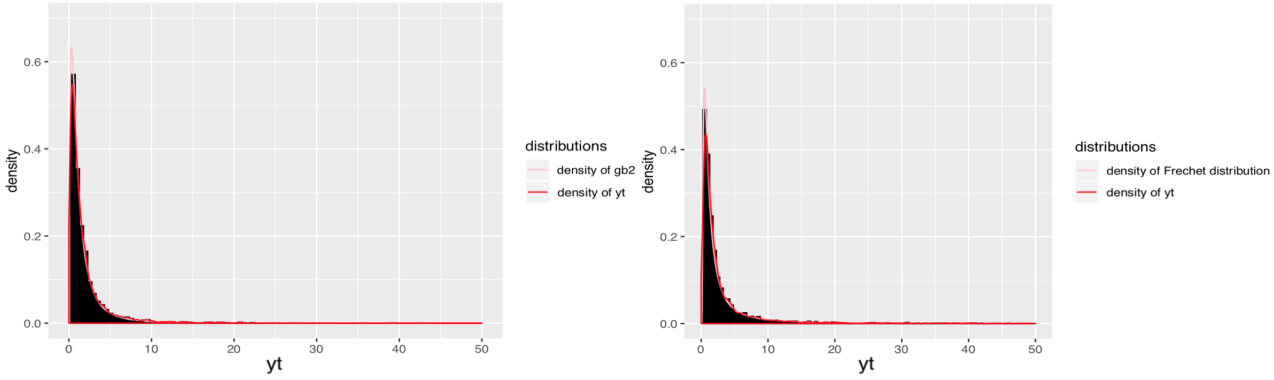

Figure 2.2.1: *Histogram and smooth lines (pink colored) for* $\{Y_t\}$*, and density curves (red colored) of GB2 (1, 1, 2, 2) (on the left) and Fréchet (1, 1) (on the right).*

To formally confirm the idea that the AcGB2 model fits better. We perform K-S tests for $\{Y_t\}$ of two data sets. Test 1 with null hypothesis that $\{Y_t\}$ is GB2 (1, 1, 2, 2) distributed and Test 2 with null hypothesis that $\{Y_t\}$ is Fréchet (1, 1) distributed. The results are shown below:

	$Test\ 1$	$Test\ 2$
$P-value\ of\ K-S\ Test$	0.7518	0.1284

Table 2.2.5: *P-values of K-S tests, Test 1 compares* $\{Y_t\}$ *from AcGB2 model with GB2 (1, 1, 2, 2) distribution; Test 2 compares* $\{Y_t\}$ *from AcF model with Fréchet (1, 1).*

From Table 2.2.5 above, both tests have p-values bigger than 0.05, which means for both tests we do not reject the null hypothesis and both models fit well. But the p-value from Test 1 (0.7518) is

way bigger than that from Test 2 (0.1284). It can be said the AcGB2 model works even better than AcF model, although both models fit well. This example shows that we can use the AcGB2 model to approximate the AcF model, especially when the data length is short or intermediately long.

2.2.3 Deal with p and q

In this subsection, we focus on how to deal with p and q. If we already know the true values of them, we can directly assume they are known and set the values of them to be the true values then estimate other parameters. However, usually we do not have much information about these two parameters. Another way is to pick some constants and assume these parameters are known and estimate (μ, β_0, β_1, β_2, β_3, α_0, α_1, α_2, α_3). If we pick enough large amount of pairs, it's possible to find a pair that fits well. But this method is still not desirable and time consuming. So we consider to treat them as unknown but fix values and estimate them.

To compare these methods, we start with simulation data with the AcGB2 model and set p and q to be (2, 2). Then for the estimation part we use three methods. First, we fit the AcGB2 model with $\{Y_t\}$ following GB2 (a = 1, b = 1, p = 2, q = 2) distribution (which is equivalent to beta prime (2, 2)). Second, we fit the AcGB2 model with $\{Y_t\}$ follow GB2 (a = 1, b = 1, p = 3, q = 3) distribution (which is equivalent to beta prime (3, 3)). Last, the model we choose is that $\{Y_t\}$ follows GB2 (a = 1, b = 1, p, q) where p and q unknown, and we estimate these two parameters. Table 2.2.6 shows the estimated parameters, and we compare them with the true parameters.

One may notice that the estimated parameters are not as close to the true parameters as they are in Section 2.2.1. The reason is that for this section, we only simulate data with length T=1500. In this way, it is clearer about which method will outperform others even with small data size. We also recover $\{Y_t\}$ for three method and compare the smooth lines of them with distribution curve of GB2 (a = 1, b = 1, p = 2, q = 2), GB2 (a = 1, b = 1, p = 3, q = 3) and GB2 (a = 1, b = 1, p = 2.630, q = 3.910). Need to notice that for the last method, we only know that the true distribution of Y_t is GB2 and parameter a = 1, b = 1, but the true values of p and q are unknown. Here we assume the estimated p and q are true values and fulfill the comparison work. We perform K-S tests for all three methods and the results are shown in Table 2.2.7.

	μ	α_0	α_1	α_2	α_3	β_0
$True\ Value$	-0.00714	-0.321	0.904	0.463	7.150	-1.670
$Estimate\ with$ $(p=2, q=2)$	-0.00812	-0.240	0.888	0.400	8.470	-1.330
$Estimate\ with$ $(p=3, q=3)$	-0.00850	-0.258	0.886	0.395	8.750	-1.330
$Estimate\ with$ $p, q\ Unknown$	-0.00646	-0.178	0.902	0.294	12.500	-0.920
	β_1	β_2	β_3	p	q	
$True\ Value$	0.391	-0.673	9.230	2.000	2.000	
$Estimate\ with$ $(p=2, q=2)$	0.597	-0.233	32.200	2.000	2.000	
$Estimate\ with$ $(p=3, q=3)$	0.592	-0.238	30.700	3.000	3.000	
$Estimate\ with$ $p, q\ Unknown$	0.657	-0.366	20.600	2.630	3.910	

Table 2.2.6: *Estimated parameters for different setting of p and q: Method 1 with (p=2, q=2), Method 2 with (p=3, q=3) and Method 3 with p, q unknown).*

	$Method\ 1$	$Method\ 2$	$Method\ 3$
$P-value\ of\ K-S\ Test$	0.7581	0.4307	0.4885

Table 2.2.7: *P-values of K-S tests, Method 1 compares $\{Y_t\}$ with GB2 (1, 1, 2, 2) distribution; Method 2 compares $\{Y_t\}$ with GB2 (1, 1, 3, 3) and Method 3 compares $\{Y_t\}$ with GB2 (1, 1, 2.630, 3.910).*

From the results of K-S tests, we find out that even though Method 1 performs the best as we expected, Method 2 and Method 3 also work well (with p-values way bigger than 0.05). This result shows that setting p and q to be unknown and estimate their values is a good way to deal with them when we do not know much of them. Also, p=3 and q=3 are not a terrible pick for them. But in reality, we may not be as luck as in simulation. It may take times for us to pick a pair that works as good as p=3 and q=3. Taking into account the time consuming and complexity we may choose the Method 3.

To further confirm our viewpoint, we also tried simulated data with the AcGB2 model and set p and q to be (0.368, 0.807). Then for the estimation part we use three methods: fitting the AcGB2 model with $\{Y_t\}$ following GB2 (a = 1, b = 1, p = 0.368, q = 0.807), fitting the AcGB2 model with $\{Y_t\}$ following GB2 (a = 1, b = 1, p = 10, q = 10), and fitting the AcGB2 model with $\{Y_t\}$ following

GB2 (a = 1, b = 1, p, q), where p and q unknown. We also perform K-S tests for these methods and the results are shown below in Table 2.2.8. This time for Method 2 where we pick the pair of p and q to be (10, 10), the estimation does not work well and the p-value of Kolmogorov-Simirnov is smaller than 0.05. But still, Method 3 fits well.

	$Method\,1$	$Method\,2$	$Method\,3$
$P-value\ of\ K-S\ Test$	0.8313	<0.0001	0.6380

Table 2.2.8: *P-values of K-S tests, method 1 compares $\{Y_t\}$ with GB2 (1, 1, 0.368, 0.807) distribution; method 2 compares $\{Y_t\}$ with GB2 (1, 1, 10, 10) and method 3 compares $\{Y_t\}$ with GB2 (1, 1, 0.436, 0.916), where (0.436, 0.916) is the estimation of (p, q).*

One may think can we also let parameters p and q be autoregressive as a and b? To answer this question, first we need to go back to take a look of our model. In the AcGB2 model, $X_t = \mu + b_t Y_t^{\frac{1}{a_t}}$, where $\{Y_t\}$ is a sequence of $i.i.d.$ GB2 (1, 1, p, q). We have one location parameter μ, three shape parameters a, p, q and one scale parameter b. We let one shape parameter (a) which controls both tails and one scale parameter (b) be autoregressive and showed that the model works well. To add more shape parameters (p and q) may not be necessary. To confirm our thought, We tried adding only p to be autoregressive, only q to be autoregressive or both p and q to be autoregressive. In these ways, we would have 15 or 19 parameters to estimate and when we are trying to optimize log-likelihood function to get parameters using real data, we found out that we barely have good results. Alternatively, we can set a_t to be constant, and set q which controls right tail to be q_t. We will discuss more about this setting in next chapter.

2.3 Real Data Applications

2.3.1 DJIA

2.3.1.1 Data Description and Model Fitting

In this subsection, we apply the AcGB2 model to maxima of negative log-returns of stocks of companies involved in Dow Jones Industrial Average (DJIA). The Dow Jones Industrial Average is one of the most commonly followed stock market index that measures the stock performance of 30

large companies listed on stock exchanges in the United States (e.g. Apple and Disney). We collect the stock prices of these 30 companies from November 24, 2014 to November 22, 2019. Then we take maxima of the negative log-returns of the stock prices as our data to analyse. The reason that we choose log-returns instead of single period return is that the log-returns are time additive, which means to get the n-period log return, we can simply add the consecutive single period log returns. Also, logs and exponents are easier to manipulate with calculus. The graph below shows the change of maxima against time. It seems there are some extreme large values that appear at around early 2016 and middle of 2019.

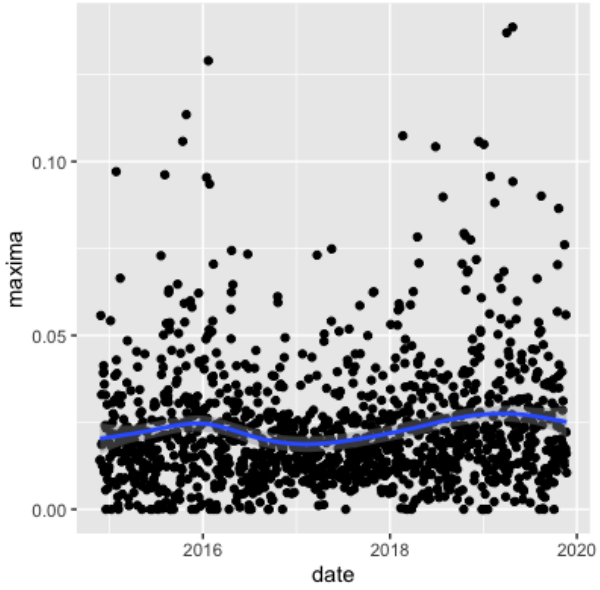

Figure 2.3.1: *Change of maxima over time of DJIA data. The trend is shown by a regression line (blue line) with* 95% *confidence intervals (gray shade).*

To see the displacement of the maxima, we need to count the maxima and draw a histogram with density curve. Also we can get a table to show the descriptive statistics of our data. According to the table below, the differences between the means (0.0230) and the medians (0.0188) imply that the data are skewed to the right. This finding is consistent with the histogram and the density shown in Figure 2.3.2 that most of the data are smaller than 0.05 and there are some extreme large values.

We fit the real data with the AcGB2 model and the AcF model. Table 2.3.2 below shows the estimated parameters. Before comparing the two models, we firstly analyze the recovered dynamic

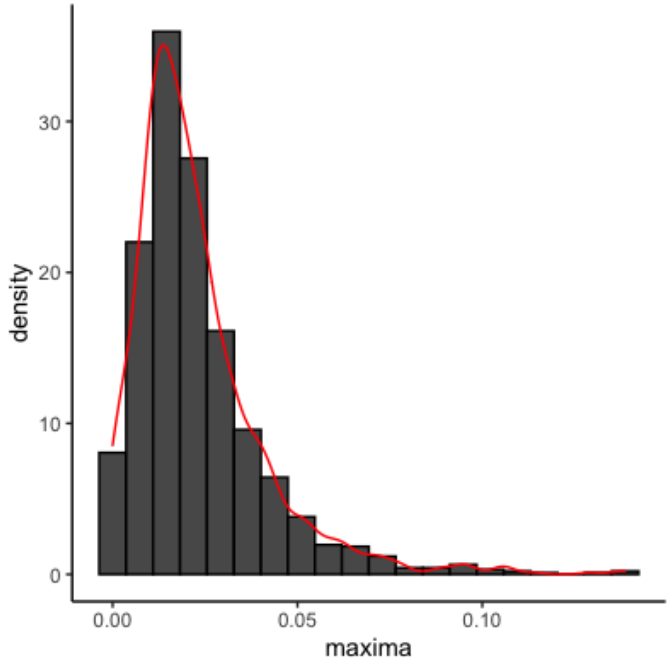

Figure 2.3.2: *Histogram and density curve for maxima of DJIA data.*

	Min	$1st\ Quantile$	$Median$	$Mean$	$3rd\ Quantile$	Max
$Maxima$	0.0000	0.0117	0.0188	0.0230	0.0295	0.138

Table 2.3.1: *Statistical summary for maxima of DJIA data.*

parameters a_t and b_t of our AcGB2 model to see whether it agrees with the trend of maxima. We recover $\{a_t\}$ and $\{b_t\}$ by the autoregressive equations with the estimated parameters:

$$\begin{cases} \log a_t = \widehat{\alpha_0} + \widehat{\alpha_1} \log a_{t-1} + \widehat{\alpha_2} \exp(-\widehat{\alpha_3}(X_{t-1})), \\ \log b_t = \widehat{\beta_0} + \widehat{\beta_1} \log b_{t-1} - \widehat{\beta_2} \exp(-\widehat{\beta_3}(X_{t-1})). \end{cases}$$

The range of recovered a_t and b_t are from 2.463 to 5.081 and from 0.0210 to 0.0590 respectively. Also we can find from the plot above that the two valleys of estimated a_t and two peaks of estimated b_t are located at about same places which are early 2016 and middle of 2019. Lower value of shape parameter a and higher value of scale parameter b imply larger tail of maxima. So this observation agrees with the trend of maxima in Figure 2.3.1, which has more extreme values in early 2016 and middle of 2019.

Fitting Model	μ	α_0	α_1	α_2	α_3	β_0	β_1	β_2	β_3
AcGB2 Mean	-0.000652	-0.162	0.891	0.378	8.820	0.214	0.830	-0.924	5.460
AcGB2 S.D.	5.33E-14	4.19E-04	1.36E-08	4.05E-04	1.88E-01	5.19E-04	5.96E-08	4.60E-04	2.87E-02
	p	q							
AcGB2 Mean	0.369	0.810							
AcGB2 S.D.	3.58E-09	3.81E-08							
	μ	γ_0	γ_1	γ_2	γ_3	β_0	β_1	β_2	β_3
AcF Mean	-0.0645	-0.229	0.863	0.602	8.320	-1.870	0.0377	-0.596	2.000
AcF S.D.	8.46E-06	1.46E-03	8.52E-06	1.46E-03	2.52E-02	4.35E-02	2.38E-03	3.75E-02	1.31E-01

Table 2.3.2: *Estimated parameters with standard deviations of two models .*

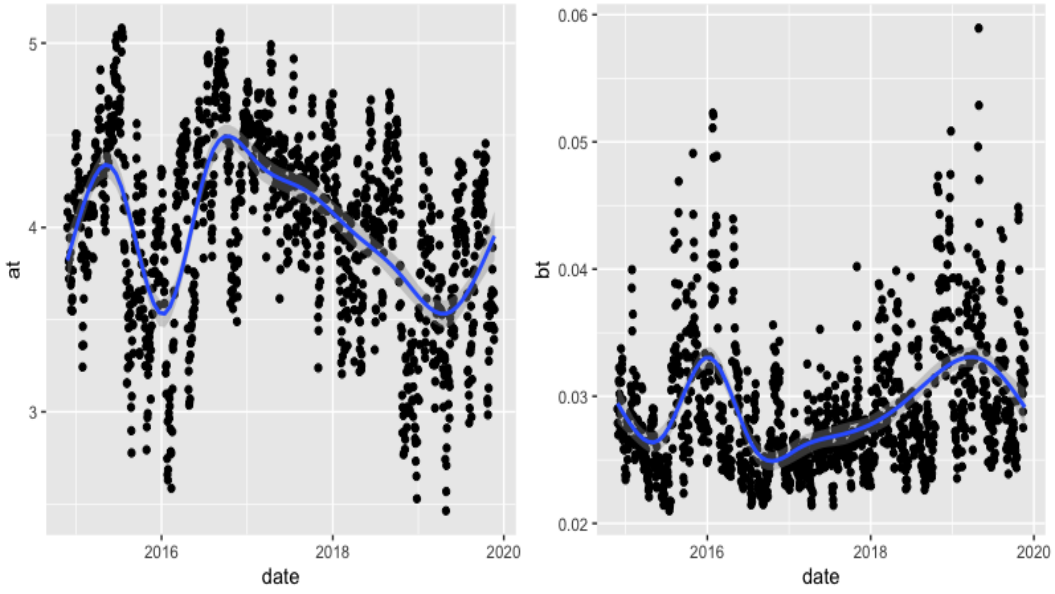

Figure 2.3.3: *Change of estimated a_t and b_t over time, a_t and b_t are dynamic estimated parameters of the AcGB2 model. The trend is shown by a regression line (blue line) with 95% confidence intervals (gray shade).*

2.3.1.2 Comparing with the AcF Model

We use three methods to compare these two models: comparing the recovered $\{X_t\}$ with the real data set; comparing the recovered scale parameter time series with the daily average volatility given by the GARCH models and comparing the results of prediction.

For the first method, we recover a set of $\{X_t\}$ with the same length of the real data set. Then we compare the distributions of this recovered $\{X_t\}$ with the real data set using Q-Q plot. For

recovering $\{X_t\}$, after getting a_t and b_t values based on autoregressive equations with estimated parameters and value of X_{t-1}, we can get value of X_t which equals to $\widehat{\mu}$ plus a variable with GB2 (a_t, b_t, $\widehat{p}$, $\widehat{q}$) distribution and we repeat the steps above to recover a series of $\{X_t\}$. Based on the result in Figure 2.3.4, the left plot that compares the real data with recovered maxima series with the AcGB2 model shows the AcGB2 model fits well with all points around the 45-degree-line. As for the right plot that compares the real data with recovered maxima series with the AcF model, only part of the points are close to the line and we can observe obvious separation starts from the middle part to the top right and this shows the AcF model not performs as good as the AcGB2 model.

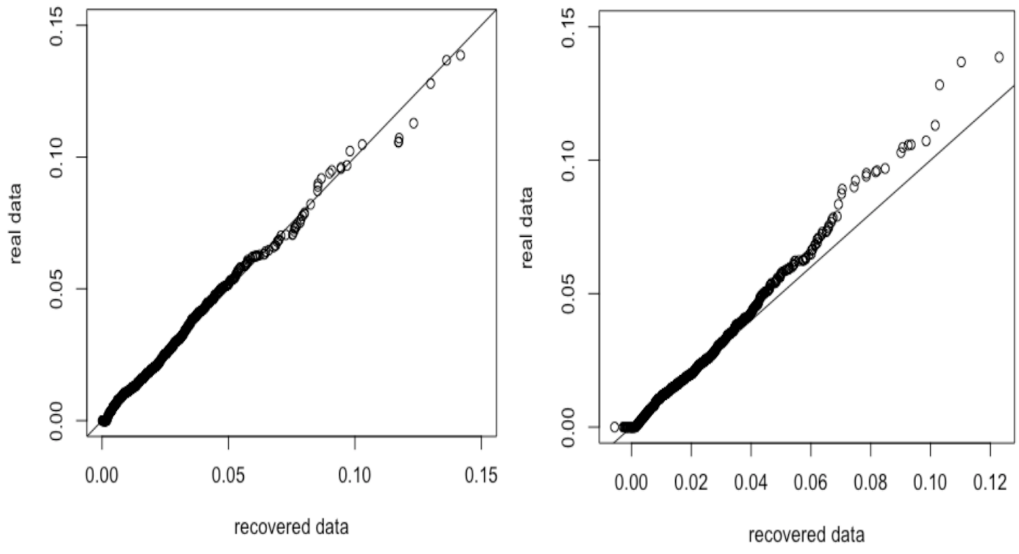

Figure 2.3.4: *Q-Q plots for the AcGB2 model and the AcF model. The left plot compares the real data with recovered maxima series with the AcGB2 model, while the right plot compares the real data with recovered maxima series with the AcF model.*

Secondly, to check whether dynamic scale parameters of the two models can accurately measure market volatility, we compare the recovered scale parameter time series with the daily average volatility given by the GARCH models following the steps below.

First We fit a GARCH(1,1) model for each component stock involved in DJIA from November 24, 2014 to November 22, 2019 and calculate the daily average volatility from the 30 GARCH models. This daily average values generated from GARCH models can be used to represent the

market volatility. Secondly, we also recover the dynamic scale parameter series with estimated parameters and autoregerssive equations for the two models respectively. Then we plot the dynamic scale parameter series together with the daily average volatility given by the GARCH models and calculate the correlations between them for the two models.

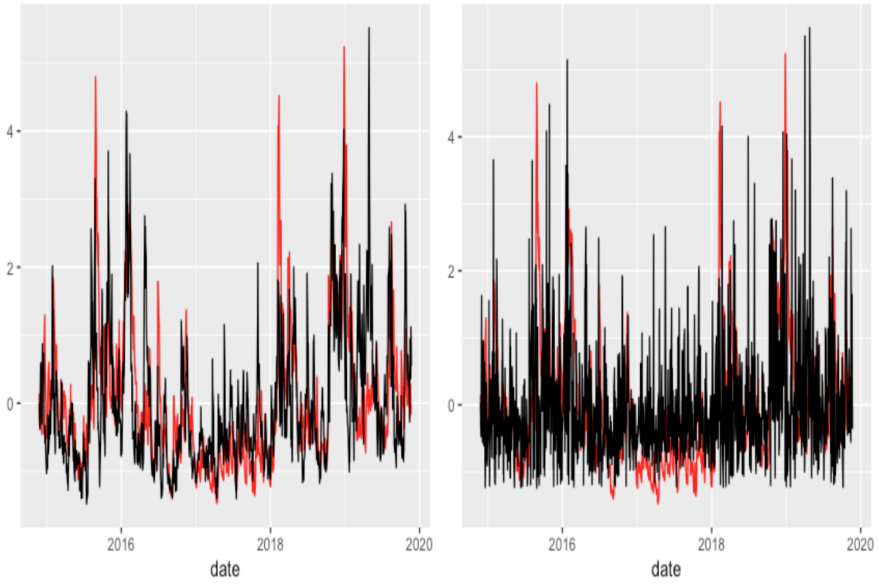

Figure 2.3.5: *Recovered dynamic scale parameter series (black line) with the daily average volatility given by the GARCH models (red line) for two models: the AcGB2 model (left) and the AcF model (right).*

In the plots of Figure 2.3.5, the red line represents the daily average volatility of 30 GARCH models, and the black line in the left plot is the recovered scale parameter series from the AcGB2 model while the black line in the right plot is the recovered scale parameter sereis from the AcF model. All the values of daily average of volatility and the recovered scale parameter are standardized by minus their mean and then divide their standard deviation to better compare them in the same plot. Clearly, we can observe that the left plot has the black line closer to the red line and can better show the trend of the red line. Also, in the right plot we have recovered scale parameter changed rapidly in short terms. All these shows that the AcGB2 model can better catch the volatility of the market. To further support our point view, we also calculate the correlations between recovered scale parameter with daily average volatility for the two model and find out that the correlation

from the AcGB2 model is over 0.67 and it from the AcF model is around 0.18. This finding is consistent with the two plots in Figure 2.3.5.

We also compare the two models through checking their prediction ability. The prediction interval is from November 25, 2019 to November 09, 2020. To fulfill the prediction, taking the AcGB2 model as an example, based on the real value of maxima at $t-1$, and estimated parameters, we get the values of a_t and b_t. Next, based on a_t, b_t, $\widehat{p}$, $\widehat{q}$ and $\widehat{\mu}$ we simulate 500 values of x_t. For these 500 values, we use median as estimator and use the range between between 0.025th percentile and 0.975th percentile as 95% confidence interval. Then we repeat the steps to generate estimated values and confidence intervals for the time interval from November 25, 2019 to November 09, 2020 and compare with the real time series of maxima. The result is shown in Figure 2.3.6.

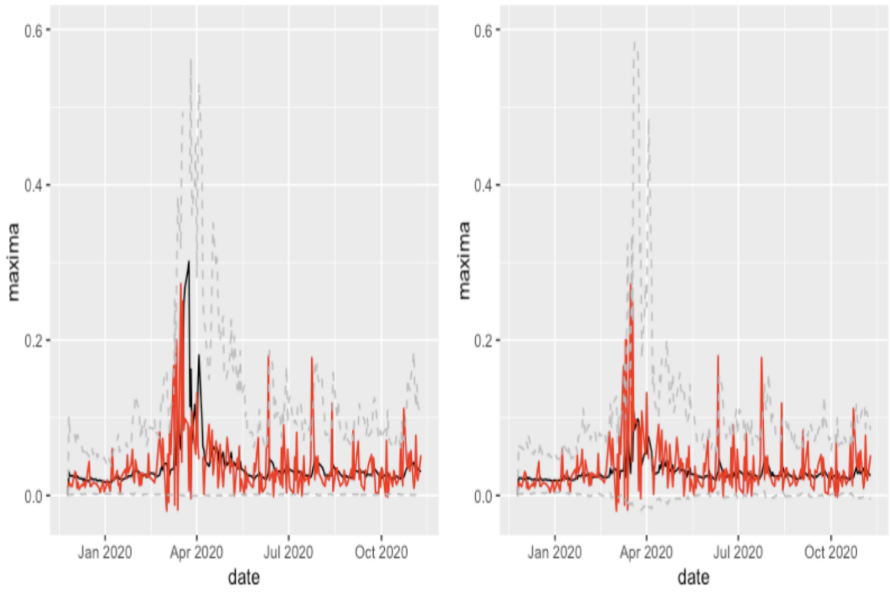

Figure 2.3.6: *Predicted maxima values from November 25, 2019 to November 09, 2020 (black line) with 95% CI (range within gray dash lines) against the real maxima values (red line) for two models: the AcGB2 model (left) and the AcF model (right).*

In two plots of Figure 2.3.6, the red line represents the real values of maxima, the black lines represent the estimated maxima, and the range between two gray dash line represent the 95% confidence intervals (the AcGB2 model on the left and the the AcF model on the right). Comparing the two plots, the black line on the left can better catch the trend of the red line, especially at the time

interval around April 2020. In this time interval, we can observe a peak of maxima time series of real values, which means their are lots of extreme large drops of stock prices, and the reason may be that April 2020 is the about the time that COVID-19 started to brake out and spread. In the left plot, the black line can better catch the behavior of red line around April 2020, while the black line in the right plot is more flat. This demonstrates that the AcGB2 model performs better than the AcF model at prediction. As for the confidence interval, both plots have less than 10% of points of real maxima out of the range, so from this perspective, both models work well.

2.3.2 S&P 100

2.3.2.1 Data Description and Model Fitting

Need to mention that with the increasing number of time series (larger m in $\{X_{it}\}_{i=1}^{m}$), the accuracy of the AcF model will increase, because of the assumption that the maxima X_t ($X_t = \max_{1\leq i\leq m} X_{it}$) fits the AcF model if m goes to infinity. So we also consider another data set with larger m: stocks of companies involved in S&P 100 (a stock market index of United States stocks maintained by Standard & Poor's). Comparing with 30 companies involved in DJIA, the S&P 100 includes 101 leading U.S. stocks with exchange-listed options (one of the companies has two classes of stock). We measure the stock price of these 101 stock prices from May 27, 2015 to May 26, 2020. The trend of the maxima is shown in Figure 2.3.7. From the graph we can see there are two peaks located around early 2016 and early 2019. Also there are some extreme large values in 2020.

We also draw a histogram with density curve and get a table with descriptive statistics of the maxima to show the displacement. The histogram in Figure 2.3.8show the data is skewed to the right and it has a long right tail. We can get the same finding from the Tale 2.3.3 that the median value 0.0438 is way bigger than the mean value 0.347.

	Min	$1st\,Quantile$	$Median$	$Mean$	$3rd\,Quantile$	Max
$Maxima$	0.0000	0.0234	0.0347	0.0438	0.0536	0.346

Table 2.3.3: *Statistical summary for maxima of S&P 100 data.*

We fit the data with the AcGB2 model and the AcF model. The estimated parameters are shown

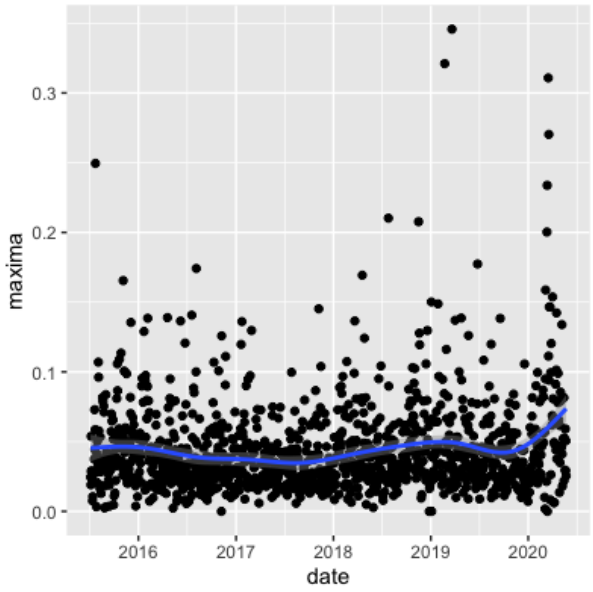

Figure 2.3.7: *Change of maxima over time of S&P 100 data. The trend is shown by a regression line (blue line) with* 95% *confidence intervals (gray shade).*

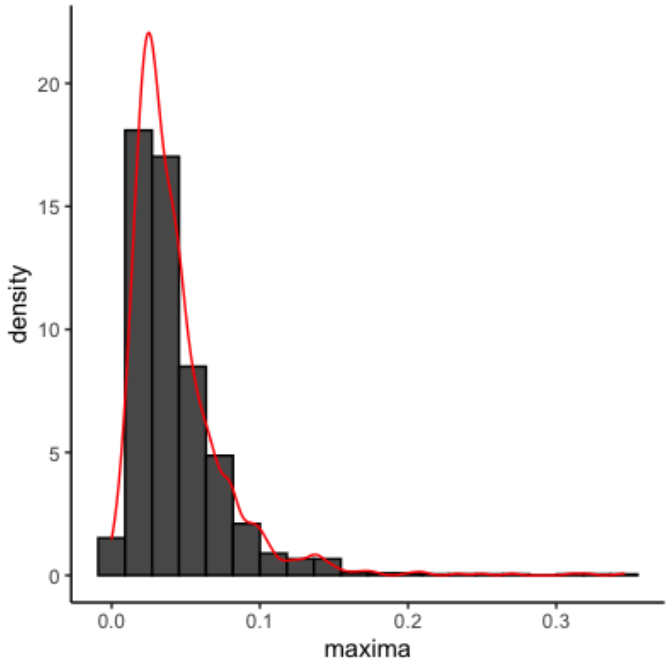

Figure 2.3.8: *Histogram and density curve for maxima of S&P 100 data.*

in Table 2.3.4.

$Fitting\ Model$	μ	α_0	α_1	α_2	α_3	β_0	β_1	β_2	β_3
$AcGB2\ Mean$	-0.00399	-0.0326	0.904	0.182	15.700	-0.343	0.771	-0.523	6.85
$AcGB2\ S.D.$	4.65E-09	4.11E-06	1.33E-08	3.23E-05	2.13E-01	3.37E-03	4.55E-07	2.11E-03	5.33E-02
	p	q							
$AcGB2\ Mean$	2.100	1.973							
$AcGB2\ S.D.$	3.23E-07	2.14E-06							
	μ	γ_0	γ_1	γ_2	γ_3	β_0	β_1	β_2	β_3
$AcF\ Mean$	-0.0582	-0.237	0.841	0.625	5.430	-0.458	0.763	-0.163	7.580
$AcF\ S.D.$	9.08E-07	3.23E-05	3.63E-06	3.67E-05	4.09E-04	1.77E-03	4.67E-06	1.89E-03	1.40E-02

Table 2.3.4: *Estimated parameters with standard deviations of two models* .

2.3.2.2 Comparing with the AcF Model

Here we also compare the fittings of the two models use three methods as we dealing with DJIA data: comparing the recovered $\{X_t\}$ with real maxima time series, comparing the ability of measuring market volatility and comparing the prediction results.

For the first method, taking the AcGB2 model as an example, Using the estimated parameters, and following the model:

$$\begin{cases} X_t = \widehat{\mu} + b_t Y_t^{\frac{1}{a_t}}, \\ \log a_t = \widehat{\alpha_0} + \widehat{\alpha_1} \log a_{t-1} + \widehat{\alpha_2} \exp(-\widehat{\alpha_3}(X_{t-1})), \\ \log b_t = \widehat{\beta_0} + \widehat{\beta_1} \log b_{t-1} - \widehat{\beta_2} \exp(-\widehat{\beta_3}(X_{t-1})), \end{cases}$$

we recover a set of $\{X_t\}$ with the same length of the real data set. Then we compare the distributions of this recovered $\{X_t\}$ with the real data set using Q-Q plot. The two plots (the AcGB2 model on the left and the AcF model on the right) in Figure 2.3.9 both have almost all points around the 45-degree-line and this shows both models fit well this time.

The result of comparing the recovered scale parameter series with the average daily volatility from GARCH models from May 27, 2015 to May 26, 2020 is presented in Figure 2.3.10, where the red line is the daily average volatility given by GARCH models, and black line at the left plot shows the recovered scale parameter series from the AcGB2 model while the black line at the right plot shows the recovered scale parameter series from the AcF model. Also, all values are

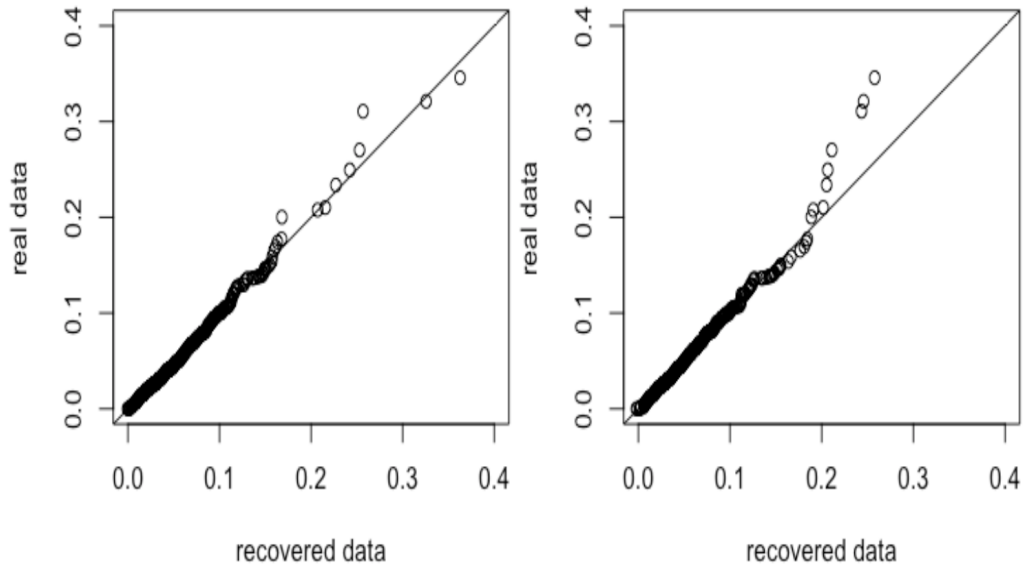

Figure 2.3.9: *Q-Q plots for the AcGB2 model and the AcF model. The left plot compares the real data with recovered maxima series with the AcGB2 model, while the right plot compares the real data with recovered maxima series with the AcF model.*

standardized. For both plots, the black lines are close to the red line and can well catch the trend of it. The correlations between the recovered scale parameter series with daily average volatility from GARCH models are calculated. For the AcGB2 model, the correlation is 0.71 which is slightly bigger than that of the AcF model which is 0.67.

In addition, since CBOE Volatility Index (VIX) that is calculated based on stock prices of components involved in S&P 100 can also be used to measure the volatility of market, we also calculate the correlations between the recovered scale parameter series with VIX from May 27, 2015 to May 26, 2020 for the two models. For the AcGB2 model, the correlation is 0.67 which is also slightly bigger than that of the AcF model which is 0.62.

We also compare the prediction ability of the two models and the predicted maxima time series is compared with the real values of maxima from May 27, 2020 to November 13, 2020. In the Figure 2.3.11, we can see that the two models work similarly, the black lines (the predicted maxima time series, left for the AcGB2 model and right for the AcF model) can well present the trend for the red line (true values of maxima). Since the time interval we applied here is about half length of

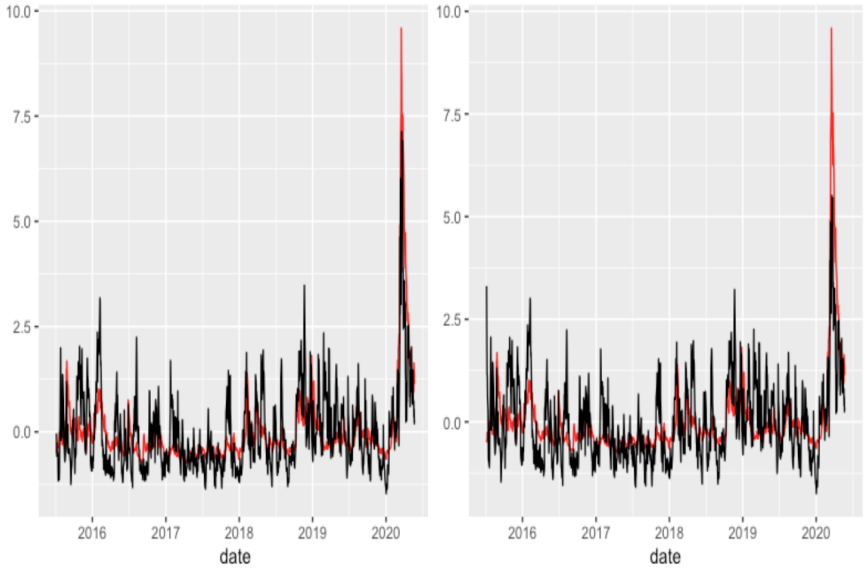

Figure 2.3.10: *Recovered dynamic scale parameter series (black line) with the daily average volatility given by the GARCH models (red line) for two models: the AcGB2 model (left) and the AcF model (right).*

the time interval we used in prediction of DJIA data, and the trend of the maxima in this interval is more stable, two black lines are more flat comparing with them of DJIA data.

From two real data applications, we can see that the AcF model requires the number of time series to be large. When the number of time series is small or intermediate, the AcGB2 model is preferred.

2.3.3 Subgroups of S&P 100

2.3.3.1 Data Description and Model Fitting

To further demonstrate that the AcGB2 model fits better than the AcF model when the number of time series is not large, we separate the 101 time series of S&P 100 into three parts with relatively small number of time series and then fit both the AcGB2 model and the AcF model. We compare results of these two models.

According to the Global Industry Classification Standard (GICS) which takes the market-orientated approach to assigning companies to different industries, the stocks in S&P 100 can be classified into

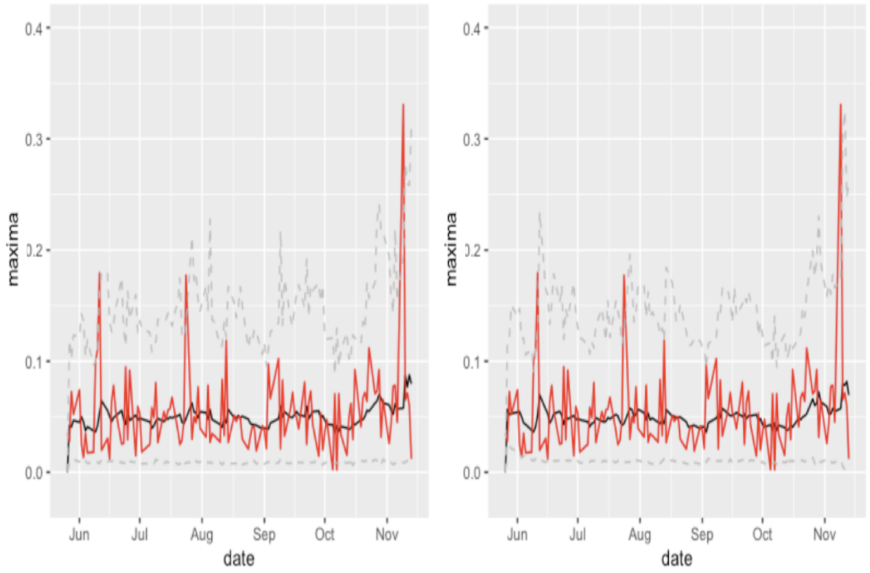

Figure 2.3.11: *Predicted maxima values from May 27, 2020 to November 13, 2020 (black line) with 95% CI (range within gray lines) against the real maxima values (red line) for two models: the AcGB2 model (left) and the AcF model (right).*

11 sectors and they are: consumer discretionary, consumer staples, energy, materials, industrial, healthcare, financial, information technology, real estate, communication services and utilities. We further separate these 11 sectors into three groups. Since the sectors financial (e.g. JPMORGAN CHASE & CO and and CAPITAL ONE FINANCIAL CORP), information technology (e.g. MICROSOFT CORP and APPLE INC) and industrial (e.g. UNITED PARCEL SERVICE INC CLASS B and BOEING) play major roles in affecting the stock market, so these three sectors are assigned to one group. It contains 42 companies. The sectors that have the closed correlations with the daily consuming of people are consumer discretionary, consumer staples and communication services. For example, NIKE and MCDONALD'S are belong to sector consumer discretionary; AMAZON and WALMART are belong to sector consumer staple and AT&T and FACEBOOK are belong to sector communication. So these three sectors are assigned to second group and it contains 30 companies. The rest 29 companies are assigned to third group.

We fit the two models to these three subgroups and comparing results. Here I use the group that contains sectors consumer discretionary, consumer staples and communication services as an

example to show the data description and parameter estimation. This group contains 30 companies. We can draw the change of maxima against time to take a look the trend of maxima for negative log-returns of stock prices of these 30 companies. We can see the regression line of trend in Figure 2.3.12 is more flat comparing with the trend of maxima of all 101 companies of S&P 100 in Figure 2.3.7. But still we can observe that the maxima decreases first and then increases from around early 2017, and then decreases again from early 2019 and at last from 2020 it increases. There are some negative values in this plot.

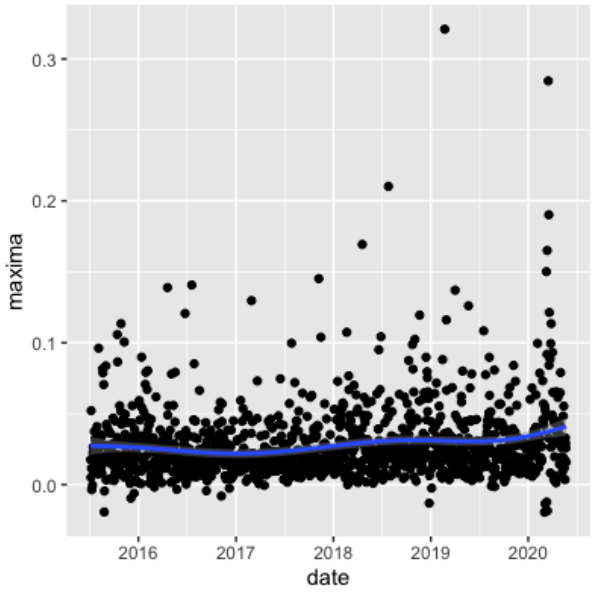

Figure 2.3.12: *Change of maxima over time of subgroup of S&P 100 data contains sectors of consumer discretionary, consumer staples and communication services. The trend is shown by a regression line (blue line) with* 95% *confidence intervals (gray shade).*

We check the displacement of maxima using histogram with density curve and table of statistical summary. Comparing the histogram in Figure 2.3.13 with the histogram of all time sires involved in S&P 100 in Figure 2.3.8, overall values are pulled to the left for this group and majorities are smaller than the maxima of whole stocks of S&P 100. More specifically, in the statistical summary table of this group, all values are smaller then them in Table 2.3.3. For example, the mean in this group is 0.0280 which is smaller than 0.0443 the mean of maxima of 101 time series; the median in this group is 0.0210 which is smaller than 0.347 the median of maxima of 101 time series. Still,

the mean is bigger than the median for this group and this is consistent with the shape of histogram that is skewed to the right.

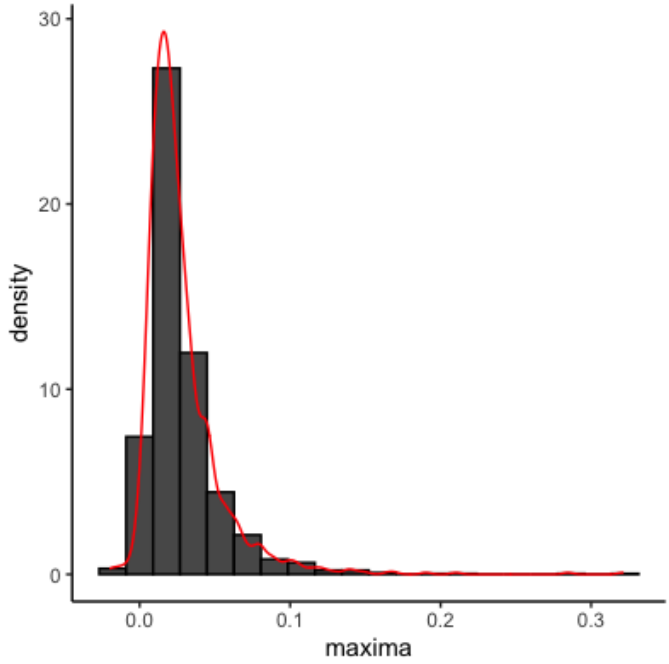

Figure 2.3.13: *Histogram and density curve for maxima of subgroup of S&P 100 data contains sectors consumer discretionary, consumer staples and communication services.*

	Min	$1st\ Quantile$	$Median$	$Mean$	$3rd\ Quantile$	Max
$Maxima$	-0.0195	0.0128	0.0210	0.0280	0.0345	0.320

Table 2.3.5: *Statistical summary for maxima of subgroup of S&P 100 data contains sectors consumer discretionary, consumer staples and communication services.*

After analyzing trend and displacement of maxima for subgroup that contains sectors consumer discretionary, consumer staples and communication services, we start to fit two models to the data: the AcGB2 model and the AcF model.The estimated parameters together with standard deviations are presented in Table 2.3.6. Based on the estimated parameter, we can recover two time-varying parameters a_t and b_t according to autoregressive equations. The results are shown in Figure 2.3.14. There are two valleys for recovered a_t and two peaks for recovered b_t. The positions of two valleys and two peaks are close and they are around the beginning of 2016 and the beginning of 2019. Also, the trend of a_t drops and the trend of b_t increases rapidly around 2020. Although, there are

not obvious peaks in the trend of maxima of this group in Figure 2.3.12, the fluctuation pattern is consistent with trends of recovered a_t and b_t.

Fitting Model	μ	α_0	α_1	α_2	α_3	β_0	β_1	β_2	β_3
AcGB2 Mean	-0.00178	-0.172	0.920	0.265	2.710	-0.348	0.817	-0.432	12.500
AcGB2 S.D.	9.48E-11	2.50E-06	1.26E-08	1.36E-06	2.69E-01	2.69E-03	4.63E-06	2.92E-03	6.50E-01
	p	q							
AcGB2 Mean	0.776	0.970							
AcGB2 S.D.	1.06E-05	2.93E-06							
	μ	γ_0	γ_1	γ_2	γ_3	β_0	β_1	β_2	β_3
AcF Mean	-0.0858	-0.125	0.807	0.614	8.652	-0.443	0.793	-0.0414	23.457
AcF S.D.	5.56E-04	8.84E-03	1.00E-03	1.22E-02	9.42E-02	8.68E-03	4.08E-03	9.23E-03	7.70E-01

Table 2.3.6: *Estimated parameters with standard deviations of two models .*

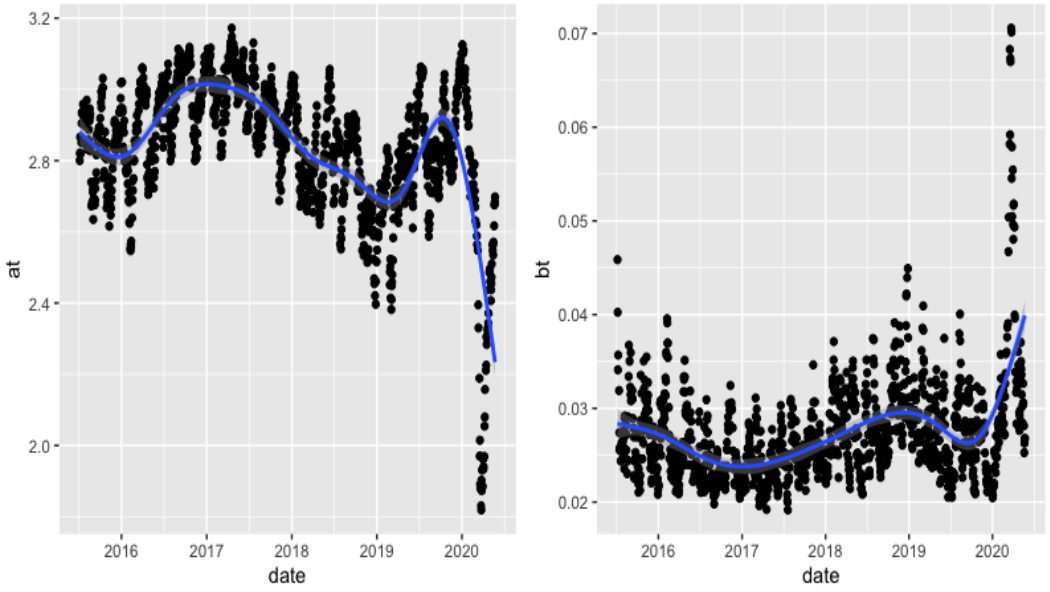

Figure 2.3.14: *Change of estimated a_t and b_t over time for data of subgroup of S&P 100 contains sectors consumer discretionary, consumer staples and communication services. a_t and b_t are dynamic estimated parameters of the AcGB2 model. The trend is shown by a regression line (blue line) with 95% confidence intervals (gray shade).*

2.3.3.2 Comparing with the AcF Model

The two models are compared in three ways. Firstly, the maxima time series for the two models are recovered and compared with real values of maxima time series. Results are shown in Figure 2.3.15.

For the left plot, we compare the recovered maxima time series from the AcGB2 model with the real values of maxima. We can see almost all values are around the 45-degree-line, but at the left bottom there are some points not as close to the line as others. This may means that there are some negative values generated from the AcGB2 model not very match with the real values of maxima. For the right plot of comparing the recovered maxima time series from the AcF model with the real values of maxima, the obvious separation can be observed from near the middle part of the 45-degree-line towards the top right. Since for financial time series, we care more about the right tail, the small unmatched part at the left corner of the left plot can be ignored and we still prefer the AcGB2 model for fitting this data set.

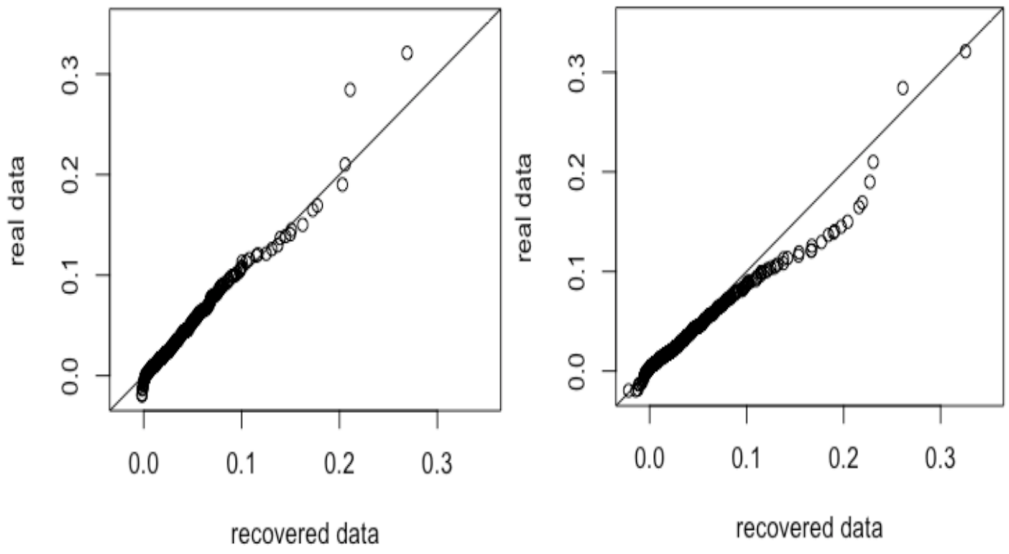

Figure 2.3.15: *Q-Q plots for the AcGB2 model and the AcF model. The left plot compares the real data with recovered maxima series with the AcGB2 model, while the right plot compares the real data with recovered maxima series with the AcF model.*

Next, we compare recovered scale parameters of the two models with the daily average volatility of GARCH models to check which model can better measure the volatility of market. We standardize values of recovered scale parameters and daily average volatility of GARCH models and plot them in the same plot to compare trends. Results are shown in Figure 2.3.16. The red line represents the daily average of volatility and black lines represent recovered scale parameters (the left plot is

from the AcGB2 model and the right plot is from the AcF model). Overall, the left plot has black line more close to the red line then the right plot. Two lines are almost overlap in the left plot. Take the time interval of 2020 as an example, for the right plot we can observe obvious separation from two lines. The black line's peak is lower than the red line's peak. But in the left plot, two peaks are almost overlap. Correlations between recovered scale parameter with the daily average volatility of GARCH models for the two models also confirm our observation, since the correlation for the AcGB2 model is 0.82 and the correlation for the AcF model is 0.46.

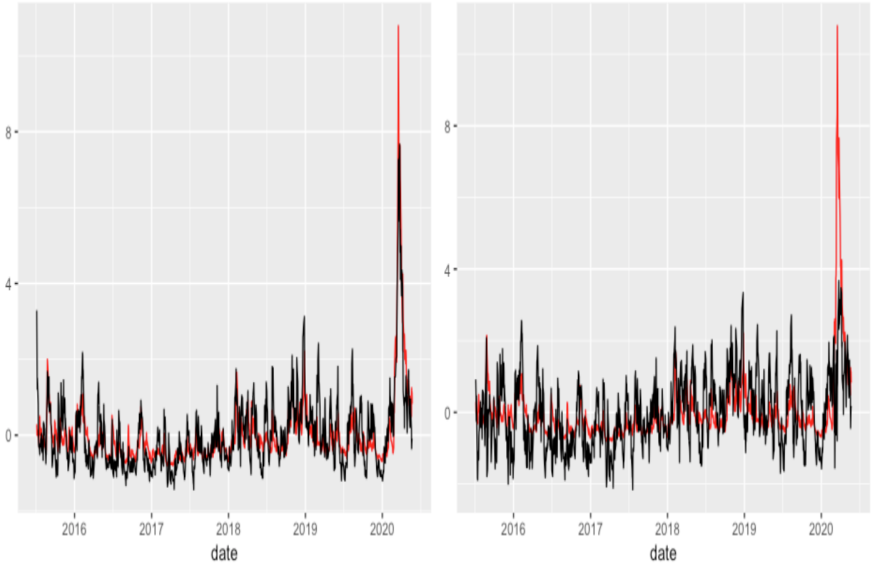

Figure 2.3.16: *Recovered dynamic scale parameter series (black line) with the daily average volatility given by the GARCH models (red line) for two models: the AcGB2 model (left) and the AcF model (right).*

For the third method of comparing the two models, we estimate maxima time series from May 27, 2020 to November 13, 2020 and compare them with real values of maxima at the same time period. In Figure 2.3.17, red lines represents real values of maxima; black lines represent estimated values of maxima and ranges within two gray dash lines represents the 95% confidence intervals, and left plot is from the AcGB2 model while the right plot is from the AcF model. Comparing these two plots, for the red line there are three parts that more extreme large values are gathered: around July, September and November. From the left plot we can find three peaks of black line around

the same location but for right plot there are only two peaks: one around July and the other around September, but the trend of red line at November is not well caught by the AcF model. Base on this observation the AcGB2 model is preferred.

Figure 2.3.17: *Predicted maxima values from May 27, 2020 to November 13, 2020 (black line) with 95% CI (range within gray lines) against the real maxima values (red line) for two models: the AcGB2 model (left) and the AcF model (right).*

After comparing the two models through three methods for fitting the subgroup that contains sectors: consumer discretionary, consumer staples and communication services, we further confirm that when the number of time series is large enough, the two models perform similarly, but with the smaller number of time series, the advantage of choosing the AcGB2 model appeared.

For the other two subgroups: one contains sectors financial, information technology and industrial (42 companies) and the other contains sectors energy, materials, healthcare, real estate and utilities (29 companies), results of fitting are similar to the subgroup contains sectors consumer discretionary, consumer staples and communication services (30 companies) that we analysed above. Similar conclusions can be generated after comparing the AcGB2 model and AcF model through the three methods Take the method of comparing the ability of measuring market volatility as an example, the correlation between recovered scale parameter from the AcGB2 model with the daily

average volatility from GARCH model for subgroup with sectors financial, information technology and industrial is 0.70, while the correlation for the AcF model is 0.45. The correlation between recovered scale parameter from the AcGB2 model with the daily average volatility from GARCH model for subgroup with sectors energy, materials, healthcare, real estate and utilities is 0.71, while the correlation for the AcF model is 0.65. In this case the correlation of the AcF model is relatively large, although it is still smaller than the value for the AcGB2 model. And this implies that the AcF model may fit well in some cases when the number of time series is small, but overall it cannot perform consistently well. Through observing the Figure 2.3.18 and Figure 2.3.16, we can get the same conclusion as comparing the values of correlations.

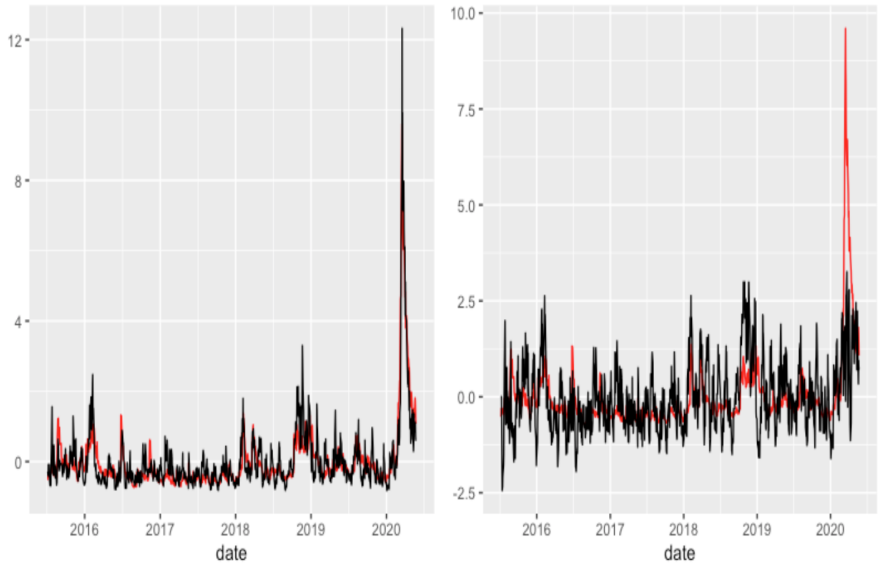

Figure 2.3.18: *Recovered dynamic scale parameter series (black line) with the daily average volatility given by the GARCH models (red line) for two models: the AcGB2 model (left) and the AcF model (right).*

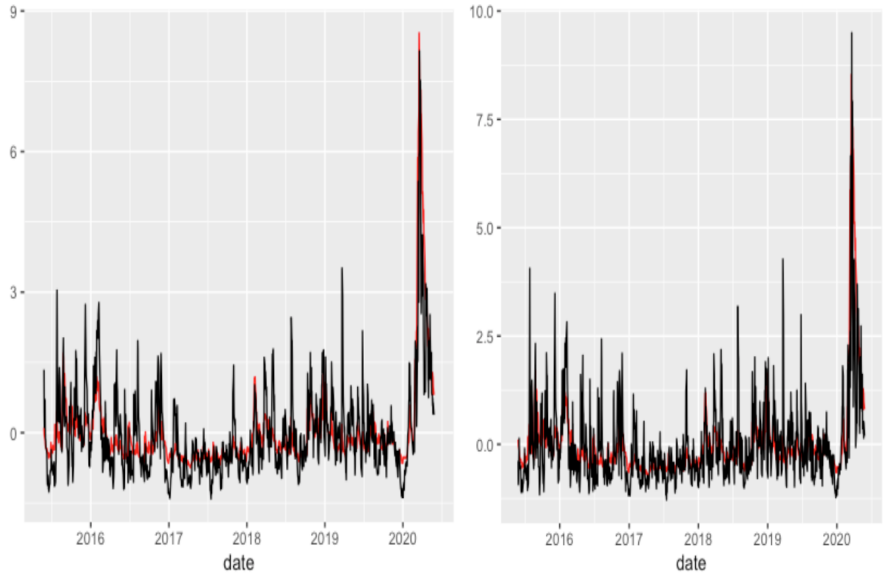

Figure 2.3.19: *Recovered dynamic scale parameter series (black line) with the daily average volatility given by the GARCH models (red line) for two models: the AcGB2 model (left) and the AcF model (right).*

Chapter 3

Dynamics at the Right Tail Parameter and the Scale Parameter

3.1 Model Specification

For the last chapter we have discussed the AcGB2 model with the two tails shape parameter a and the scale parameter b dynamic. Since for the financial market we focus more on the right tail, in this chapter, we try to update the model by letting the right tail shape parameter q instead of a to be dynamic and we keep the scale parameter b dynamic to see the performance of this new model.

We consider model the maxima time series X_t by GB2 with parameters (μ, a, b_t, p, q_t):

$$X_t = \mu + F_{Z_t}^{-1}(Z_t),$$

where μ is location parameter, b_t and q_t are two time-varying parameters; Z_t follows uniform distribution on (0,1); F_{Z_t} is the CDF of GB2 distribution with (a, b_t, p, q_t) and F^{-1} is inverse function of F.

According to the setting above we will have $X_t - \mu$ ($= F_{Z_t}^{-1}(Z_t)$) follow GB2 (a, b_t, p, q_t) and the p.d.f of X_t will be:

$$f_{X_t}(x_t) = \frac{1}{B(p,q_t)} \frac{a(x_t-\mu)^{ap-1}}{b_t^{ap}(1+(\frac{x_t-\mu}{b_t})^a)^{p+q_t}}, X_t > \mu, \tag{3.1.1}$$

where $B(p, q_t)$ is a beta function with parameters (p, q_t).

Then the two time-varying parameters b_t and q_t are assumed to have the dynamic pattern:

$$\log q_t = \gamma_0 + \gamma_1 \log q_{t-1} + \gamma_2 \exp(-\gamma_3 X_{t-1}), \tag{3.1.2}$$

$$\log b_t = \lambda_0 + \lambda_1 \log b_{t-1} - \lambda_2 \exp(-\lambda_3 X_{t-1}), \tag{3.1.3}$$

where γ_0, λ_0, $\mu \in \mathbb{R}$, $0 \leq \gamma_1, \lambda_1 < 1$ $\gamma_1 \neq \lambda_1$ and $\gamma_2, \gamma_3, \lambda_2, \lambda_3 > 0$.

Remark 3.1.1. $\gamma_2 \exp(-\gamma_3 X_{t-1})$ in equation 3.1.2 is a decreasing function of X_{t-1} while $-\lambda_2 \exp(-\lambda_3 X_{t-1})$ in equation 3.1.3 is an increasing function of X_{t-1}. Since large values of maxima tend to happen around the same period, a large X_{t-1} is usually followed by a large X_t. Decreasing of function $\gamma_2 \exp(-\gamma_3 X_{t-1})$ and increasing of function $-\lambda_2 \exp(-\lambda_3 X_{t-1})$ ensure that larger X_{t-1} is fol-

lowed by smaller q_t and larger b_t. Then smaller q_t and larger b_t will lead to a large tail and there is more possible to have a large value of X_t.

By the setting of the model, we obtain the following Theorem 3.1.1. The stationarity and ergodicity of process $\{q_t, b_t\}$ can be demonstrated follow the idea of Chan and Tong (1994). As a result, process $\{X_t\}$ is also stationary and ergodic.

Theorem 3.1.1. *For AcGB2 with q_t and b_t dynamics and γ_0, λ_0, $\mu \in \mathbb{R}$, $0 \leq \gamma_1, \lambda_1 < 1$ $\gamma_1 \neq \lambda_1$ and $\gamma_2, \gamma_3, \lambda_2, \lambda_3, a, p > 0$, the process $\{q_t, b_t\}$ is stationary and geometrically ergodic.*

The proof of Theorem 3.1.1 can be found in 5.2.1

As for parameter estimation, the parameter set we try to estimate for this model is (μ, γ_0, γ_1, γ_2, γ_3, λ_0, λ_1 λ_2, λ_3, a, p) and we defined it as ϕ. We want to find the local maximizer of log-likelihood function with arbitrary initial value q_1 and b_1, and we define this log-likelihood function as:

$$\begin{aligned} L_n(\phi) &= \frac{1}{n}\sum_{t=1}^{n} l_t(\phi) \\ &= \frac{1}{n}\sum_{t=1}^{n}\Bigg[-\log(B(p,q_t)) + \log(a) - ap\log(b_t) + (ap-1)\log(x_t-\mu) \\ &\quad - (p+q_t)\log[1+(\frac{x_t-\mu}{b_t})^a]\Bigg], \end{aligned}$$

where $l_t(\phi) = -\log(B(p,q)) + \log(a) - ap\log(b_t) + (ap-1)\log(x_t-\mu) - (p+q_t)\log[1+(\frac{x_t-\mu}{b_t})^a]$; q_t and b_t are generated with ϕ and arbitrary initial values q_1 and b_1.

Even though with arbitrary initial values the consistency, asymptotic normality and uniqueness can be proved for the local maximizer of the log-likelihood function above in the following theories: the Theorem 3.1.2, Theorem 3.1.3 and Theorem 3.1.4.

Theorem 3.1.2. *(Consistency) There exists a local maximizer $\widehat{\phi_n}$ of $L_n(\phi)$ for $\{X_t\}_{t=1}^n$ from AcGB2 with q_t and b_t dynamics and γ_0, λ_0, $\mu \in \mathbb{R}$, $0 \leq \gamma_1, \lambda_1 < 1$ $\gamma_1 \neq \lambda_1$ and $\gamma_2, \gamma_3, \lambda_2, \lambda_3, a, p > 0$ and true parameter ϕ^0 within a compact space, such that $\widehat{\phi_n} \xrightarrow{p} \phi^0$, $\|\widehat{\phi_n} - \phi^0\| \leq \delta_n$, for $\delta_n \searrow 0$, $n\delta_n \to +\infty$, $\frac{1}{\delta_n}(\min\{X_i\}_{i=1}^t - \mu^0) \to +\infty$, where $L_n(\phi)$ is log-likelihood function with ϕ and arbitrary initial value q_1, b_1.*

The proof of Theorem 3.1.2 can be found in 5.2.2.

Theorem 3.1.3. *(Asymptotic Normality) The local maximizer* $\widehat{\phi_n}$ *of* $L_n(\phi)$ *for* $\{X_t\}_{t=1}^n$ *from AcGB2 with* q_t *and* b_t *dynamics and* γ_0, λ_0, $\mu \in \mathbb{R}$, $0 \leq \gamma_1, \lambda_1 < 1$ $\gamma_1 \neq \lambda_1$ *and* $\gamma_2, \gamma_3, \lambda_2, \lambda_3, a, p > 0$ *and true parameter* ϕ^0 *within a compact space, such that* $\widehat{\phi_n} \to \phi^0$, $\|\widehat{\phi_n} - \phi^0\| \leq \delta_n$, *for* $\delta_n \searrow 0$, $n\delta_n \to +\infty$, $\frac{1}{\delta_n}(\min\{X_i\}_{i=1}^t - \mu^0) \to +\infty$, *is asymptotic normal. Equivalently,* $\sqrt{n}(\widehat{\phi_n} - \phi^0) \xrightarrow{d} N\left(0, \frac{1}{-E_{\phi^0}(\frac{\partial^2 l_t^0(\phi^0)}{\partial\phi\partial\phi^T})}\right)$.

The proof of Theorem 3.1.3 can be found in 5.2.2. It follows the same idea as the proof of Theorem 2.1.3, by applying the method from Smith (1985).

Theorem 3.1.4. *(Uniqueness) Under conditions in Theorem 3.1.1, there is an asymptotic unique MLE over* M_n, *where* $M_n = \{\phi \in \phi | \mu \leq c \min_{1 \leq t \leq n} X_t + (1-c)\mu_0\}$; ϕ *is a compact set of* $\{\phi | \gamma_0, \lambda_0, \mu \in \mathbb{R}, 0 \leq \gamma_1, \lambda_1 < 1, \gamma_1 \neq \lambda_1, \gamma_2, \gamma_3, \lambda_2, \lambda_3, a, p > 0\}$; $0 < c < 1$.

The proof of Theorem 3.1.4 can be found in 5.2.2.

3.2 Simulation Study

In this section, we compare how the sample sizes will affect the accuracy of estimation use our model. And then shows the flexibility and robustness of the model when comparing it with AcF model.

3.2.1 Simulation of Different Data Size

We simulate data with three different lengths: 1000, 3000 and 5000. For each length, we simulate 100 data sets with the parameters (μ, γ_0, γ_1, γ_2, γ_3, λ_0, λ_1 λ_2, λ_3, a, p) to be (-0.000854, -0.309, 0.950, 0.353, 7.180, -2.340, 0.197, -0.675, 13.500, 3.880, 0.383). Then we fit the simulate data with our model to get estimators. We calculate means and standard deviations of estimators for different data length, then compare the means with true values and compare the standard deviation with each other. The result are shown in two tables below.

$Parameter$	μ	γ_0	γ_1	γ_2	γ_3	λ_0
$True\ Value$	-0.000854	-0.309	0.950	0.353	7.180	-2.340
$Mean\ of$ $T = 1000$	-0.000683	-0.231	0.938	0.284	15.596	-2.0341
$Mean\ of$ $T = 3000$	-0.000769	-0.280	0.950	0.324	11.695	-2.216
$Mean\ of$ $T = 5000$	-0.000813	-0.309	0.949	0.348	7.455	-2.275
$Parameter$	λ_1	λ_2	λ_3	a	p	
$True\ Value$	0.197	-0.675	13.500	3.880	0.383	
$Mean\ of$ $T = 1000$	0.304	-0.585	18.150	4.048	0.386	
$Mean\ of$ $T = 3000$	0.210	-0.750	12.174	3.863	0.388	
$Mean\ of$ $T = 5000$	0.206	-0.698	13.223	3.811	0.391	

Table 3.2.1: *Mean of MLE for 100 data sets with length to be 1000, 3000 and 5000.*

$Parameter$	μ	γ_0	γ_1	γ_2	γ_3	λ_0
$S.D.\ of$ $T = 1000$	2.30E-04	2.07E-01	2.34E-02	1.98E-01	8.72E+00	5.82E-01
$S.D.\ of$ $T = 3000$	9.64E-05	1.98E-01	9.89E-03	2.37E-01	6.63E+00	2.66E-01
$S.D.\ of$ $T = 5000$	5.99E-05	1.44E-01	1.24E-02	1.44E-01	3.86E+00	2.11E-01
$Parameter$	λ_1	λ_2	λ_3	a	p	
$S.D.\ of$ $T = 1000$	1.63E-01	1.41E-01	6.77E+00	9.98E-01	1.30E-01	
$S.D.\ of$ $T = 3000$	6.18E-02	1.61E-01	3.78E+00	4.51E-01	5.25E-02	
$S.D.\ of$ $T = 5000$	4.90E-02	1.05E-01	3.19E+00	2.95E-01	3.67E-02	

Table 3.2.2: *Standard deviation of MLE for 100 data sets with length to be 1000, 3000 and 5000.*

We can see from Table 3.2.1, with the increase of data length, the estimated values are closer to the true vales. If we take parameter γ_0 as an example, the true value is -0.309, with data length to be 1000, the estimator is -0.231, with data length to be 3000, the estimator is -0.280 witch is closer to the true value, and with data length to be 5000, the estimator is -0.309 witch is the same as true value. For other parameters, they have the same trend. Base on Table 3.2.2, we find with

the increase of data length, the standard deviation will be smaller. We still take parameter γ_0 as an example, with data length to be 1000, the standard deviation equals to 4.29e-02, then with data length to be 3000, we observed a smaller standard deviation witch is 3.91e-02, and when the data length is 5000, the standard deviation is 2.08e-02, which is smallest. So according to the results from these two tables we have the conclusion that when the data lengths increase, we will have more accurate and more stable estimators.

3.2.2 Compare with the AcF Model

To compare with the AcF model and show the flexibility and robustness, we follow the steps as below. First, we simulate data with AcF model, then fit AcGB2 model with q and b dynamics to get estimates of ϕ. After that, we recover time series of maxima $\{X_t\}$ and compare the distribution of the recovered maxima with the simulated data to check whether model fits well by using Q-Q plot. The way to recover time series of maxima $\{X_t\}$ is shown below: with estimate parameters ($\widehat{\mu}$, $\widehat{\gamma_0}$, $\widehat{\gamma_1}$, $\widehat{\gamma_2}$, $\widehat{\gamma_3}$, $\widehat{\lambda_0}$, $\widehat{\lambda_1}$ $\widehat{\lambda_2}$, $\widehat{\lambda_3}$, $\widehat{a}$, $\widehat{p}$)), we have X_t equals to $\widehat{\mu}$ plus a GB2 distributed distributed variable with parameters $\widehat{a}$, b_t, $\widehat{p}$ and q_t, where b_t and q_t can be generated from equations:

$$\begin{cases} \log q_t = \widehat{\gamma_0} + \widehat{\gamma_1} \log q_{t-1} + \widehat{\gamma_2} \exp(-\widehat{\gamma_3} X_{t-1}), \\ \log b_t = \widehat{\lambda_0} + \widehat{\lambda_1} \log b_{t-1} - \widehat{\lambda_2} \exp(-\widehat{\lambda_3} X_{t-1}). \end{cases}$$

Second, we simulate data with AcGB2 model with q and b dynamics, then fit AcF model to get estimate parameters. After that, we recover $\{X_t\}$ with AcF model and estimated parameters. Then to check the goodness of fit, we compare the recovered data with simulated data through Q-Q plot.

The results of two Q-Q plots are shown. The plot on the left which is generated following the step that simulating with the AcF model and fitting with the AcGB2 model has points all close to 45-degree-line. As for the plot on the right which is generated following the step that simulating with the AcGB2 model and fitting with the AcF model has points around the 45-degree-line at beginning but the separation can be observed at the top right. This result implies that the recovered data on the left plot is closer to the simulated data comparing with the right plot and it demonstrates that

considering about the flexibility and robustness, the AcGB2 model with parameters q and b dynamic is a better choice comparing with AcF model.

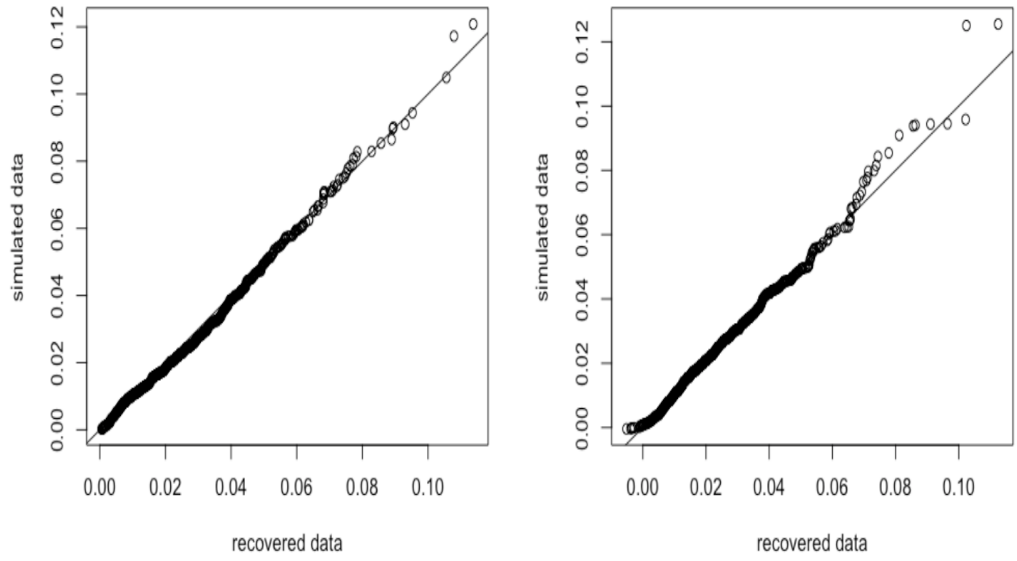

Figure 3.2.1: *Q-Q plots for the AcGB2 model with q and b dynamics and the AcF model. The left plot is generated with the method of simulating with the AcF model and fitting with the AcGB2 model with q and b dynamics, while the right plot is generated with the method of simulating with the AcGB2 model with q and b dynamics and fitting with the AcF model.*

3.3 Real Data Application

3.3.1 DJIA

In this section, the data we choose to fit our model is negative log-returns of companies involved in DJIA witch is the data we applied in Chapter 2. The estimated parameters with standard deviations are shown in Table 3.3.1. Comparing with the Table 2.3.2 of the AcGB2 model with a and b dynamic, we can notice that the standard deviations in Table 2.3.2 are almost all smaller than them of Table 3.3.1, and this indicates that the estimations for the AcGB2 model with q and b dynamic are less stable. For example, the standard deviation for estimation of μ in Table 2.3.2 is 5.33E-14 which is smaller than 3.43E-06 in Table 3.3.1. And the standard deviation for estimation of γ_0 in Table 2.3.2 is 4.19E-04 which is smaller than 2.33E-03 in Table 3.3.1. Also, we find that the

estimation for the AcGB2 model with a and b dynamic are more time consuming than them for for the AcGB2 model with q and b dynamic.

$Fitting\ Model$	μ	γ_0	γ_1	γ_2	γ_3	λ_0	λ_1	λ_2	λ_3
$AcGB2\ Mean$	-0.000854	-0.309	0.950	0.353	7.180	-2.340	0.197	-0.675	13.500
$AcGB2\ S.D.$	3.43E-06	2.33E-03	6.78E-05	1.03E-03	3.22E-01	8.25E-04	4.56E-06	2.03E-02	8.74E-02
	a	p							
$AcGB2\ Mean$	3.880	0.383							
$AcGB2\ S.D.$	3.12E-5	7.41E-05							

Table 3.3.1: *Estimated parameters with standard deviations of the AcGB2 model with q and b dynamics for data of DJIA .*

With the estimated parameters, original data $\{X_t\}$ and autoregressive equations:

$$\begin{cases} \log q_t = \widehat{\gamma_0} + \widehat{\gamma_1} \log q_{t-1} + \widehat{\gamma_2} \exp(-\widehat{\gamma_3} X_{t-1}), \\ \log b_t = \widehat{\lambda_0} + \widehat{\lambda_1} \log b_{t-1} - \widehat{\lambda_2} \exp(-\widehat{\lambda_3} X_{t-1}), \end{cases}$$

we can recover $\{q_t\}$ and $\{b_t\}$. Figure 3.3.1 shows the recovered $\{q_t\}$ has range within 0.482 and 1.226, and there are two valleys located at around early 2016 and middle 2019. Also, the $\{b_t\}$ has range within 0.0237 and 0.0438, and there are two peaks located at around the same place as where the valleys of $\{q_t\}$ located. Since lower values of q_t and higher values of b_t imply large tail of maxima which means there are more possible to get large maxima, the trends of $\{q_t\}$ and $\{b_t\}$ are consistent with that of $\{X_t\}$ showed in Figure 2.3.1. This shows that $\{q_t\}$ and $\{b_t\}$ recovered well with estimated parameters of autoregressive equations.

To further check the goodness of fit, we recover $\{X_t\}$ based on estimate parameters and use Q-Q plot to compare it with original maxima data and the result is shown in Figure 3.3.2. According to the Q-Q plot basically all of the points are around 45-degree-line and this demonstrate that our model fit well.

Next, we compare the standardized recovered scale parameter series $\{b_t\}$ at interval from November 24, 2014 to November 22, 2019 with the standardized daily average volatility from GARCH models to analyse whether the AcGB2 model with q and b dynamics can well measure the volatility of market. And we compare results with the AcGB2 model with a and b dynamics and the AcF

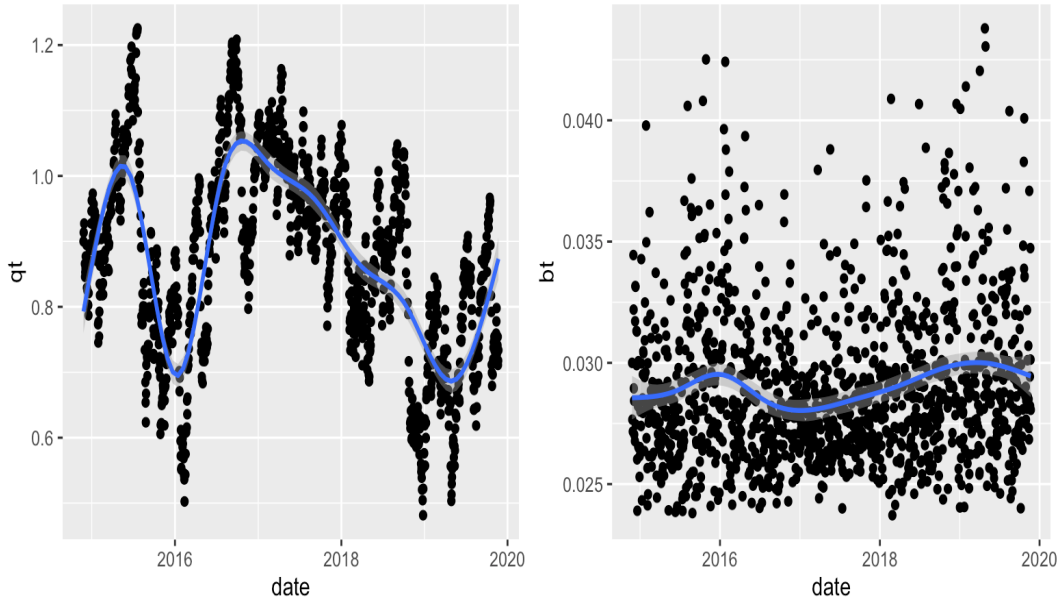

Figure 3.3.1: *Change of estimated q_t and b_t over time for data of DJIA, q_t and b_t are dynamic estimated parameters of the AcGB2 model. The trend is shown by a regression line (blue line) with* 95% *confidence intervals (gray shade).*

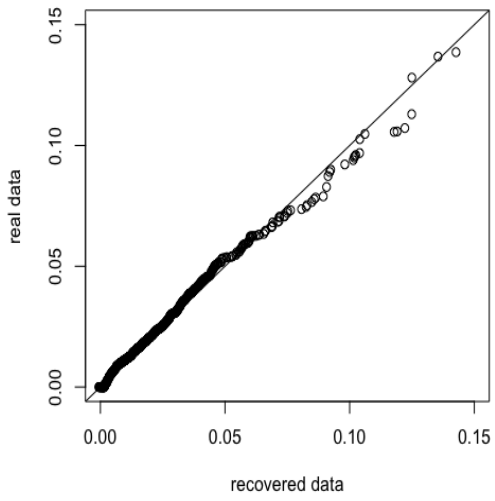

Figure 3.3.2: *Q-Q plots for the AcGB2 model with q and b dynamics for data of DJIA.*

model. Through comparing the plot in Figure 3.3.3 with plots in Figure 2.3.5, the black line for the AcGB2 model with q and b dynamics performs similar with the AcF model, the recovered scale parameter have large standard deviation and can not match with the red line as well as the the AcGB2

model with a and b dynamics. The correlation is also calculated and the value is 0.22, which is close the value for the AcF model and smaller then the value of the AcGB2 model with q and b dynamics.

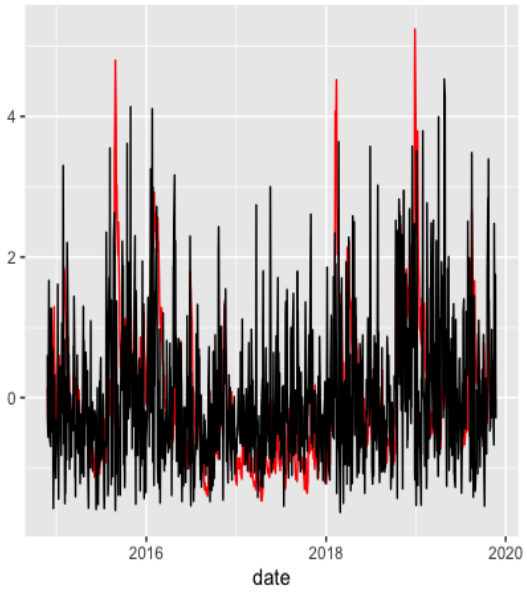

Figure 3.3.3: *Recovered dynamic scale parameter series (black line) with the daily average volatility given by the GARCH models (red line) for the AcGB2 model with q and b dynamics.*

The prediction ability of the AcGB2 model with q and b dynamics is also checked. We predict maxima time series from November 25, 2019 to November 09, 2020 and compare them with the real maxima values in Figure 3.3.4. Comparing the plot with prediction results in Figure 2.3.6, we can find the prediction of the AcGB2 model with q and b dynamics still performs similar with the AcF model and cannot catch the trend of real maxima values as well as the the AcGB2 model with q and b dynamics, especially at around the April 2020.

After analyzing and researching above, while being applied to DJIA data, the AcGB2 model with q and b dynamics performs similarly as the AcF model and not outperforms the AcGB2 model with a and b dynamics. To further confirm our findings, we also compare these models use other data sets.

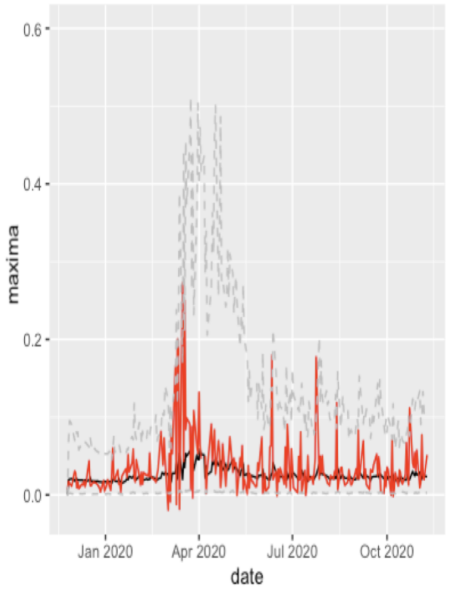

Figure 3.3.4: *Predicted maxima values from November 25, 2019 to November 09, 2020 (black line) with 95% CI (range within gray lines) against the real maxima values (red line) for the AcGB2 model with q and b dynamics.*

3.3.2 S&P 100

We also fit our model to the data of S&P 100, the estimated parameters with standard deviations are shown in Table 3.3.2. Comparing with Table 2.3.4, we have the same finding as estimating parameters for DJIA data that the standard deviations of estimation in the AcGB2 model with q and b dynamics are relatively larger than them of the AcGB2 model with a and b dynamics. Still using parameters μ and γ_0 as example, the two values in Table 3.3.2 are 8.13E-06 and 4.11E-04 and they are larger than 9.08E-07 and 3.23E-05 in Table 3.3.2 respectively.

$Fitting\ Model$	μ	γ_0	γ_1	γ_2	γ_3	λ_0	λ_1	λ_2	λ_3
$AcGB2\ Mean$	-0.00632	-0.356	0.833	0.545	7.670	-1.870	0.397	-0.175	9.530
$AcGB2\ S.D.$	8.13E-06	4.11E-04	1.23E-05	8.30E-03	3.11E-01	2.25E-03	2.68E-06	8.01E-03	8.21E-02
	a	p							
$AcGB2\ Mean$	2.640	1.660							
$AcGB2\ S.D.$	2.09E-06	3.55E-06							

Table 3.3.2: *Estimated parameters with standard deviation of the AcGB2 model with q and b dynamics for data of S&P 100 .*

We recover $\{q_t\}$ and $\{b_t\}$ and record trends in Figure 3.3.5. The recovered trends are consistent with that of maxima showed in Figure 2.3.7 and this demonstrates that the autoregressive equations works well: the smooth line of $\{q_t\}$ has two large valleys at around the early 2016 and early 2019, also it decreases at 2020 while the smooth line of $\{b_t\}$ has two large peaks at around the early 2016 and early 2019,and it increases at 2020.

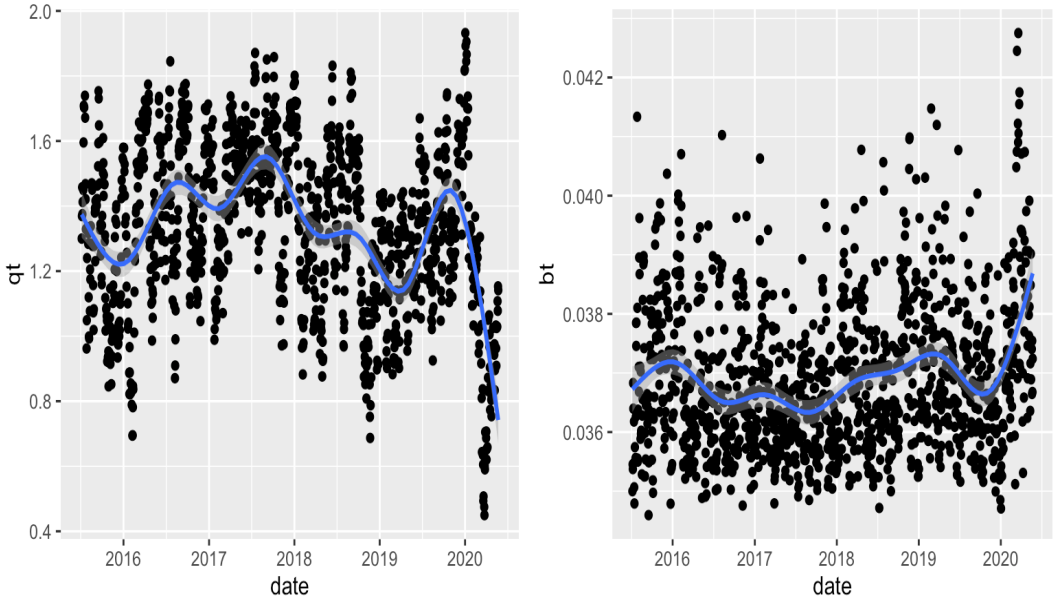

Figure 3.3.5: *Change of estimated q_t and b_t over time for data of S&P 100, q_t and b_t are dynamic estimated parameters of the AcGB2 model. The trend is shown by a regression line (blue line) with* 95% *confidence intervals (gray shade).*

We also recover $\{X_t\}$ with estimated parameters and compare it with original maxima data using Q-Q plot to check the goodness of fit for the model. Based on Figure 3.3.6, we observe that basically all of the points are located around 45-degree-line and this confirms the idea that our model fit well. Comparing with Q-Q plots in Figure 2.3.9, all of these three three models perform equally well and we can not tell the difference based on recovering maxima time series $\{X_t\}$.

Another method to compare the AcGB2 model with q and b dynamics is to check the performance of measuring market volatility. The recovered scale parameter series from May 27, 2015 to May 26, 2020 are compared with daily average of volatility from GARCH models after standardizing. From Figure 3.3.7, basically the black line of the recovered $\{b_t\}$ can well represent the trend

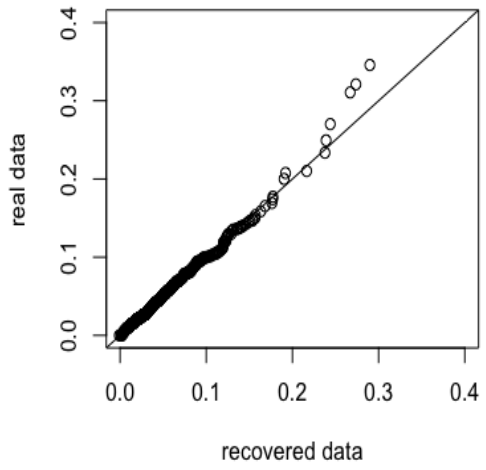

Figure 3.3.6: *Q-Q plots for the AcGB2 model with q and b dynamics for data of S&P 100.*

of the red line of daily average volatility from GARCH models, but while comparing with two plots in Figure 2.3.10, the black line is not as close as it in Figure 2.3.10. The calculation of correlation also confirm this point view: the correlation is 0.45 for model AcGB2 with q and b dynamics which is smaller than 0.71 and 0.67 for the other two models.

The prediction of maxima time series from May 27, 2020 to November 13, 2020 together with 95% confidence intervals are shown in Figure 3.3.8. The black line of predicted maxima can well catch the trend of real values of maxima (red line), and the probability of real values of maxima within the confidence intervals is over 95%. While comparing with two plots of prediction in Figure 2.3.11, the AcGB2 model with q and b dynamics is as good as the other two models.

The application of S&P data shows the AcGB2 model with q and b dynamics not outperforms the AcF model. The AcGB2 model with a and b model performs similar with the AcF model and measures the market volatility slightly better than the AcF model.

3.3.3 Subgroups of S&P 100

To see the performances of the AcGB2 model with q and b dynamics while applying to more data sets, we also fit the model to the three subgroups of S&P 100. The subgroup with sectors consumer

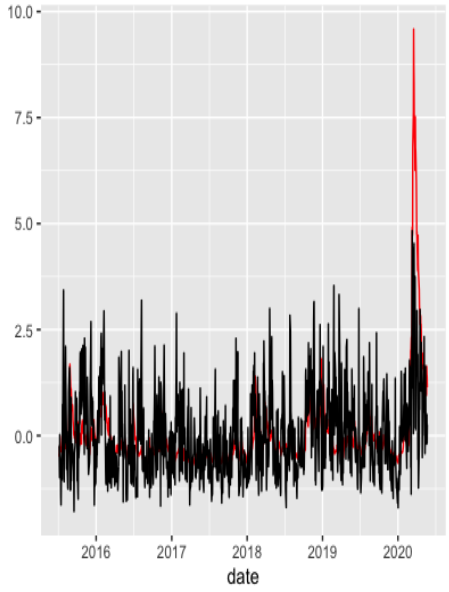

Figure 3.3.7: *Recovered dynamic scale parameter series (black line) with the daily average volatility given by the GARCH models (red line) for the AcGB2 model with q and b dynamics.*

discretionary, consumer staples and communication services that contains 30 companies is picked as an example to show the parameter estimation and compare the goodness of fit with the AcF model and the AcGB2 model with a and b dynamics. In the Table 3.3.3, the estimated parameters and their standard deviations can be found. The estimations of this model overall have large standard deviations comparing with results of estimations for the AcGB2 model with a and b dynamics in Table 2.3.6.

Then Based on the estimated parameter, we can recover two time-varying parameters q_t and b_t according to autoregressive equations. The results are shown in Figure 3.3.9. The recovered shape parameter $\{q_t\}$ has two valleys: one at around beginning of 2016 and the other at around beginning of 2019, while the recovered scale parameter $\{b_t\}$ has two peaks at around beginning of 2016 and beginning of 2019. These trends are close to the recovered parameters series showed in Figure 2.3.14 for the AcGB2 model with parameter a and b dynamics, but we can observe that the points at Figure 3.3.9 are spread widely.

To check the goodness of fit for the model, first we recover maxima time series using the estimated parameters and compare the maxima with the real data through Q-Q plot to see whether

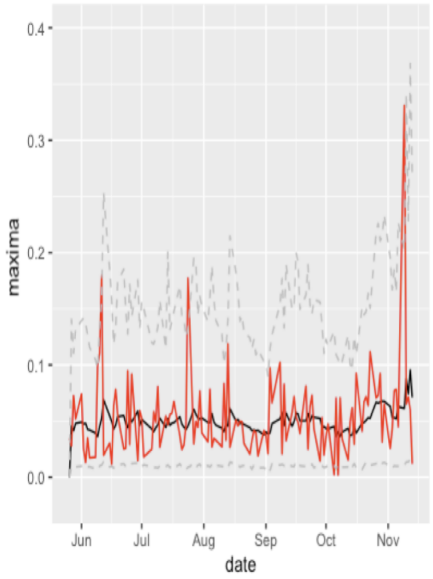

Figure 3.3.8: *Predicted maxima values from May 27, 2020 to November 13, 2020 (black line) with 95% CI (range within gray lines) against the real maxima values (red line) for the AcGB2 model with q and b dynamics.*

Fitting Model	μ	γ_0	γ_1	γ_2	γ_3	λ_0	λ_1	λ_2	λ_3
AcGB2 Mean	-0.0828	-0.398	0.914	0.320	9.804	-1.640	0.268	-0.098	5.99
AcGB2 S.D.	2.14E-06	1.11E-05	4.12E-04	1.35E-03	5.19E-01	6.85E-02	7.68E-06	6.99E-03	1.25E-01
	a	p							
AcGB2 Mean	33.342	0.533							
AcGB2 S.D.	6.50E-04	5.66E-06							

Table 3.3.3: *Estimated parameters with standard deviation of the AcGB2 model with q and b dynamics for data of subgroup of S&P 100 contains sectors consumer discretionary, consumer staples and communication services .*

the plot has points close to 45-degree-line. As we can see from the results in Figure 3.3.10, most of the points are stick to the 45-degree-line, although there are two points not close to the line, but since the amount is very small, they do not need to be worried about. According to recovered $\{X_t\}$, the AcGB2 model with q and b dynamics performs closely with the AcGB2 model with a and b dynamics and out performs the AcF model. Results of recovered $\{X_t\}$ for the two models can be found in Table 2.3.15.

Then we also check the whether the AcGB2 model with q and b dynamics can well measure

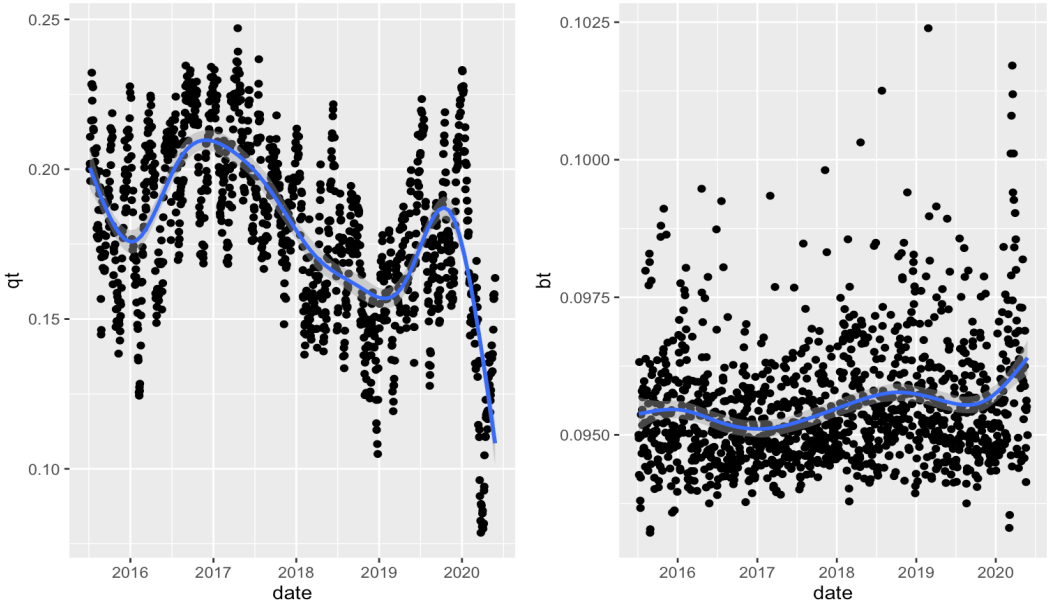

Figure 3.3.9: *Change of estimated q_t and b_t over time for data of subgroup of S&P 100 contains sectors consumer discretionary, consumer staples and communication services. q_t and b_t are dynamic estimated parameters of the AcGB2 model with q and b dynamics. The trend is shown by a regression line (blue line) with 95% confidence intervals (gray shade).*

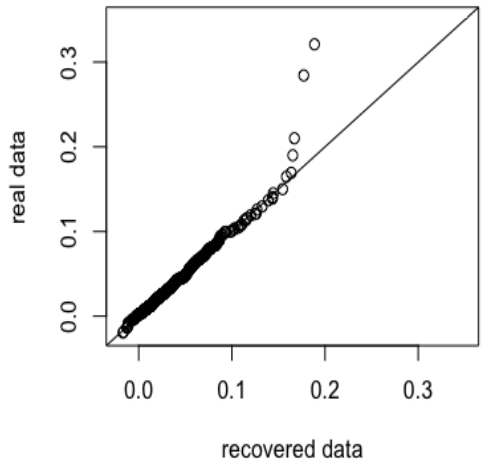

Figure 3.3.10: *Q-Q plots for the AcGB2 model with q and b dynamics for data of subgroup of S&P 100 contains sectors consumer discretionary, consumer staples and communication services.*

the volatility of market through comparing the recovered scale parameter with the average daily volatility from GARCH models (after standardizing). Comparing the plot in the Figure 3.3.11 with plots in the Figure 2.3.16, we can find the AcGB2 model with a and b dynamics performs best, the AcGB2 model with q and b dynamics has large standard deviations of recovered scale parameter, but it performs better than the AcF model at period of 2020. The correlation for the AcGB2 model with q and b dynamics is 0.50 which is slightly bigger than 0.46 of the AcF model but smaller than 0.82 of the the AcGB2 model with a and b dynamics.

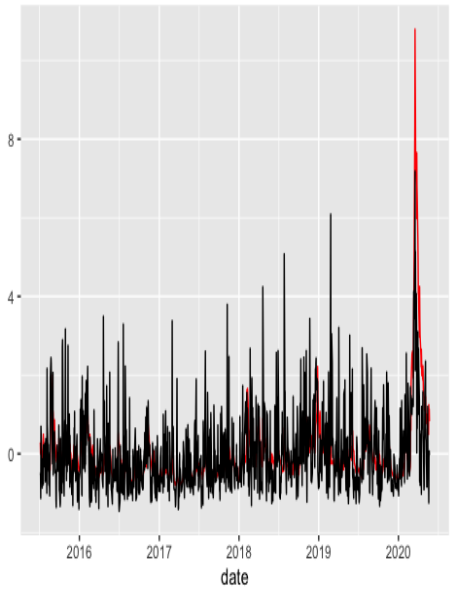

Figure 3.3.11: *Recovered dynamic scale parameter series (black line) with the daily average volatility given by the GARCH models (red line) for the AcGB2 model with q and b dynamics.*

The result of prediction for this data set is shown in Figure 3.3.12. Comparing with predictions of the other two models, the estimated value in this plot is flat and can not accurately present the trend of maxima from May 27, 2020 to November 13, 2020.

In conclusion, after fitting the three models with different data, the AcGB2 model with a and b dynamics performs best and consistently no matter the number of time series is large or small. The AcF model fits well when the number of time series is large, but not as good as the AcGB2 model with a and b dynamics while the number of time seires is small. The AcGB2 model with q and b dynamics performs not consistently, sometimes it outperforms the AcF model and sometimes not.

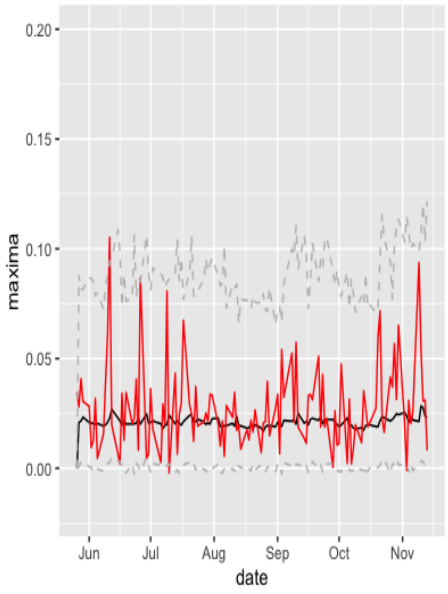

Figure 3.3.12: *Predicted maxima values from May 27, 2020 to November 13, 2020 (black line) with 95% CI (range within gray lines) against the real maxima values (red line) for the AcGB2 model with q and b dynamics.*

Chapter 4

Discussion

In order to model highly skewed and long tailed data in financial market. Currently, there are two major approaches widely analyzed and applied: the BM method and the POT method. For BM method the extreme value distributions like Fréchet, Gumble and Weibull are applied to fit maximum observations of periods with equal size, while for POT method, the observations are selected by the rule that they exceed a threshold and the distribution used is generalized Pareto distribution. However there are two challenges for current methods: these distributions not perform well enough when the financial data is intermediate extreme; also the behavior and structure of the maxima time series are changing over time but these two methods usually use independent and identically distributed series. In order to overcome these two shortcomings of current models, we introduce a novel dynamic GB2 framework to model the time-varying behavior of maxima in financial time series.

In our model the GB2 distribution is chosen to fit maxima time series. Majorly there are three advantages for this four parameter distribution: it can provide an excellent description for highly skewed and long-tailed data and it is more flexible to fit the financial data at less extreme levels and many distributions including extreme value distributions can be approximated by GB2 with different parameter values. In the GB2 family tree that describe in McDonald and Xu (1995), many distributions are included like generalized gamma, Weibull and chi-square. We can use one single path as an example, when shape parameter a of GB2 equals to 1, it becomes Beta distribution of the second kind; then when another shape parameter q converges to infinity with $b = \beta * q^{\frac{1}{a}}$, it becomes Gamma distribution; further when β equals to 2, it becomes Chi-square distribution. The solution to deal with the time-varying behavior of maxima in financial time series is that since volatility dynamics can be directly related to changing parameters of fitting distribution, we make parameters of GB2 vary according to time. For the dynamic equations of time-varying parameters, we let the parameters depending not only on their own history but also on the history of maxima.

The next question is which parameters of GB2 distribution we should choose to be dynamic and change over time. For GB2 distribution, it has four parameters and three of them are shape parameters including a, p and q and one scale parameter b. In Chapter 2, shape parameter a which controls both left tail and right tail and scale parameter b are chosen to be dynamic. In the dynamic

equations, for a_t we include a decreasing function of maxima X_{t-1} while for b_t we include a increasing function of maxima X_{t-1}. Since large values of maxima tend to happen around the same period, a large X_{t-1} is more often followed by a large X_t. An increasing function of maxima in dynamic equation of shape parameter a and a decreasing funciton of maxima in dynamic equation of scale parameter b ensure that larger X_{t-1} is followed by smaller a_t and larger b_t. Then smaller a_t and larger b_t will have large tail and there is more possible to have large value of X_t. In Chapter 3, we tried another combination of dynamic parameters shape parameter q and scale parameter b. The reason we choose q is that this shape parameter controls right tail and in financial market literature, we care more about right tail and usually the data is skewed to the right. Smaller value of q also tend to have larger tail, so in the setting of dynamic equation for q_t, we also let the function of maxima to be decreasing for the same reason as in Chapter 2. Through the setting of time series pair $\{a_t\}$ and $\{b_t\}$ and pair $\{q_t\}$ and $\{b_t\}$ are proved to have the properties of stationarity and ergodicity. The parameters are estimated using maximum likelihood estimation method and consistency, asymptotic normality and uniqueness are proved for estimators.

Simulation and real data examples shows the performance of the AcGB2 model in a variety of situations. After using maximum likelihood estimation method to get the estimate parameters, we use them to recover time series pairs $\{a_t\}$ and $\{b_t\}$ or $\{q_t\}$ and $\{b_t\}$ and check whether the trends consistent with each other within each pair and whether the trends are consistent with the trend of original maxima time series. Also, the maxima time series is recovered base on the estimated parameters and the model, and Q-Q plot is applied to compare recovered maxima with real data or simulated data to check the goodness of fit for our model. The abilities of measuring market volatility and prediction are also compared for different models. The two real data we applied are stock prices of companies involved in DJIA and stock prices of companies involved in S&P 100. We take the maxima of negative log returns of stock prices as our data to analyze. Also, we further separate the companies involved in S&P 100 into three groups according to sectors to show that when the number of time sires is small our models still perform well especially the AcGB2 model with a and b dynamic. During the simulation study and real data analysis, the flexibility, robustness and goodness of fit are demonstrated.

Still, there should be possible improvement of our model. For the four parameters of GB2 distribution, in Chapter 2 and Chapter 3 we choose two different pairs of parameters to be dynamic. We also considered using three or four of them to be time-varying, but by doing so, we will have fifteen or nineteen parameters to estimate and it will take much longer time to get estimation work done and it is hard to get applicable estimators. How to overcome the efficiency issue and let more parameters to be dynamic may need more considerations and can be left as further work.

Chapter 5

Appendix: Lemmas and Proofs

5.1 Proofs for Model with Parameters a and b Dynamic

5.1.1 Proof of Stationarity and Ergodicity

Proof of Theorem 2.1.1:

Proof. Our proof follows the results of Chan and Tong (1994). They showed under appropriate conditions, an embeded deterministic dynamic system which admits a compact attractor will be stationary and ergodic. The main theory of Chan and Tong (1994) is shown below:

Let $(\mathbf{X}_n)$ be the data follows the stochastic difference equation: $\mathbf{X}_{n+1} = T(\mathbf{X}_n) + S(\mathbf{X}_n, e_{n+1})$, where $\mathbf{X}_n$ and e_n are vector in $\mathbf{R}^m$, $\{e_n\}$ is a sequence of independent and identically distributed random variables, $T(\mathbf{X}_n)$ is a function of $\mathbf{X}_n$ and $S(\mathbf{X}_n, e_{n+1})$ ia a function of $\mathbf{X}_n$ and e_{n+1}. Let Λ be a compact attractor, and G be an absorbing open subset of the domain of attraction for Λ. Then if the following five conditions hold:

(a) for any positive integer k, Λ has a dense orbit with respect to T^k where T^k is the k-fold composition of T with itself;

(b) Λ is exponentially attracting, i.e., $\exists$ K, $\lambda > 0$ such that $\forall x \in G$, dist$(T^n x, \lambda) \leq K \exp(-\lambda n)$, where dist$(.,.)$ denotes the Euclidean distance from a point to a set;

(c) T is Lipschitz continuous over G;

(d) for all $x \in G$, 0 is in the support of $\mid S(\mathbf{X}, e) \mid$, Where $\mid . \mid$ is the norm of the vector; $\exists$ integer $k \geq 1$, $\forall x \in G$, there is a continuous and positive function $r(x)$ such that the kth step transition probability for $(\mathbf{X}_n)$, $P^k(x, dy)$, has an absolutely continuous component whose probability density function is positive over $\mathbf{B}(T^k(x), r(x))$, where $\mathbf{B}(x, \delta)$ denotes the open ball in G with the centre at x and the radius equal to δ;

(e) $E\{\mid S(\mathbf{X}_n, e_{n+1}) \mid\mid \mathbf{X}_n = x\}$ is uniformly bounded above for $x \in G$;

then $(\mathbf{X}_n)$ as a Markov chain on G is stationary and ergodic. More details can be found in Chan and Tong (1994).

Next, we verify all five conditions (a)-(e) with our model in three steps.

First step: Define T and S such that $\mathbf{X}_{n+1} = T(\mathbf{X}_n) + S(\mathbf{X}_n, e_{n+1})$.

Without loss of generality, we assume $\mu = 0$, then we have:

$$\begin{cases} \log a_t = \alpha_0 + \alpha_1 \log a_{t-1} + \alpha_2 \exp(-\alpha_3 b_{t-1} Y_{t-1}^{\frac{1}{a_{t-1}}}), \\ \log b_t = \beta_0 + \beta_1 \log b_{t-1} - \beta_2 \exp(-\beta_3 b_{t-1} Y_{t-1}^{\frac{1}{a_{t-1}}}). \end{cases}$$

We can write it as

$$\begin{bmatrix} \log a_t \\ \log b_t \end{bmatrix} = \begin{bmatrix} \alpha_0 + \alpha_1 \log a_{t-1} \\ \beta_0 + \beta_1 \log b_{t-1} \end{bmatrix} + \begin{bmatrix} \alpha_2 \exp(-\alpha_3 b_{t-1} Y_{t-1}^{\frac{1}{a_{t-1}}}) \\ -\beta_2 \exp(-\beta_3 b_{t-1} Y_{t-1}^{\frac{1}{a_{t-1}}}) \end{bmatrix}, \tag{5.1.1}$$

and let $\mathbf{X}_t = \begin{bmatrix} \log a_t \\ \log b_t \end{bmatrix}$, $T(\mathbf{X}_{t-1}) = \begin{bmatrix} \alpha_0 + \alpha_1 \log a_{t-1} \\ \beta_0 + \beta_1 \log b_{t-1} \end{bmatrix}$,

and $S(\mathbf{X}_{t-1}, Y_{t-1}) = \begin{bmatrix} \alpha_2 \exp(-\alpha_3 b_{t-1} Y_{t-1}^{\frac{1}{a_{t-1}}}) \\ -\beta_2 \exp(-\beta_3 b_{t-1} Y_{t-1}^{\frac{1}{a_{t-1}}}) \end{bmatrix}$, i.e., $e_t = Y_{t-1}$.

The condition (d) in Chan and Tong (1994) requires that 0 is the support of $\mid S(\mathbf{X}_{t-1}, Y_{t-1}) \mid$. But Since $\alpha_2 \exp(-\alpha_3 b_{t-1} Y_{t-1}^{\frac{1}{a_{t-1}}}) = 0$, $-\beta_2 \exp(-\beta_3 b_{t-1} Y_{t-1}^{\frac{1}{a_{t-1}}}) = 0$ and $\alpha_2, \beta_2 > 0$ can not hold at the same time, condition (d) does not hold.

So we update the equation (5.1.1) as

$$\begin{bmatrix} \log a_t \\ \log b_t \end{bmatrix} = \begin{bmatrix} \alpha_0 + \alpha_1 \log a_{t-1} + m_1 \\ \beta_0 + \beta_1 \log b_{t-1} - m_2 \end{bmatrix} + \begin{bmatrix} -m_1 + \alpha_2 \exp(-\alpha_3 b_{t-1} Y_{t-1}^{\frac{1}{a_{t-1}}}) \\ m_2 - \beta_2 \exp(-\beta_3 b_{t-1} Y_{t-1}^{\frac{1}{a_{t-1}}}) \end{bmatrix}. \tag{5.1.2}$$

Then $-m_1 + \alpha_2 \exp(-\alpha_3 b_{t-1} Y_{t-1}^{\frac{1}{a_{t-1}}}) = 0$, $m_2 - \beta_2 \exp(-\beta_3 b_{t-1} Y_{t-1}^{\frac{1}{a_{t-1}}}) = 0$, and $b_{t-1} Y_{t-1}^{\frac{1}{a_{t-1}}} > 0$ implies: $\begin{cases} 0 < m_2 < \beta_2, \\ m_1 = \alpha_2 \exp(-\alpha_3 b_{t-1} Y_{t-1}^{\frac{1}{a_{t-1}}}) = \alpha_2 \exp(-\alpha_3(-\frac{1}{\beta_3} \log(\frac{m_2}{\beta_2}))) = \alpha_2 \exp(\frac{\alpha_3}{\beta_3} \log(\frac{m_2}{\beta_2})). \end{cases}$

Step 2: Define Λ and G, and to have T satisfy $T^n(x) \to \Lambda$ as $n \to +\infty$.

Let's define $T^n(x) = \begin{bmatrix} T^n(x)[1] \\ T^n(x)[2] \end{bmatrix}$ and focus on $T^n(x)[2]$ first:

$T^1(\mathbf{X}_{t-1})[2] = \beta_0 + \beta_1 \log(b_{t-1}) - m_2 = (\beta_0 - m_2)(1) + \beta_1 \log(b_{t-1})$,

$T^2(\mathbf{X}_{t-1})[2] = \beta_0 + \beta_1 T^1(\mathbf{X}_{t-1}) - m_2 = (\beta_0 - m_2)(1 + \beta_1) + \beta_1^2 \log(b_{t-1})$,

$T^3(\mathbf{X}_{t-1})[2] = \beta_0 + \beta_1 T^2(\mathbf{X}_{t-1}) - m_2 = (\beta_0 - m_2)(1 + \beta_1 + \beta_1^2) + \beta_1^3 \log(b_{t-1})$,

$\vdots$

$T^n(\mathbf{X}_{t-1})[2] = \beta_0 + \beta_1 T^{n-1}(\mathbf{X}_{t-1}) - m_2 = (\beta_0 - m_2)(1 + \beta_1 + \beta_1^2 + \cdots + \beta_1^{n-1}) + \beta_1^n \log(b_{t-1})$,

$\to (\beta_0 - m_2)(\frac{1}{1-\beta_1})$.

Similarly, $T^n(\mathbf{X}_{t-1})[1] \to (\alpha_0 + m_1)(\frac{1}{1-\alpha_1})$. So we get $\Lambda = \begin{bmatrix} \frac{\alpha_0 + m_1}{1-\alpha_1} \\ \frac{\beta_0 - m_2}{1-\beta_1} \end{bmatrix}$.

We want to define G to be an absorbing set for $\mathbf{X}_t$.

Since $\Lambda = \begin{bmatrix} \frac{\alpha_0 + m_1}{1-\alpha_1} \\ \frac{\beta_0 - m_2}{1-\beta_1} \end{bmatrix}$, $0 < m_1 < \alpha_2$, $0 < m_2 < \beta_2$, and G is an open subset of the domain of attraction for Λ, we can set $G = \left(\frac{\alpha_0}{1-\alpha_1}, \frac{\beta_0 - \beta_2}{1-\beta_1} \right) \times \left(\frac{\alpha_0 + \alpha_2}{1-\alpha_1}, \frac{\beta_0}{1-\beta_1} \right)$.

If $\log b_t > \frac{\beta_0 - \beta_2}{1-\beta_1}$, then

$\log b_{t+1} = \beta_0 + \beta_1 \log b_t - \beta_2 \exp(-\beta_3 b_t Y_t^{\frac{1}{a_t}}) > \beta_0 + \beta_1 \frac{\beta_0 - \beta_2}{1-\beta_1} - \beta_2 \exp(-\beta_3 b_t Y_t^{\frac{1}{a_t}})$

$= \frac{\beta_0 - \beta_2 [\beta_1 + \exp(-\beta_3 b_t Y_t^{\frac{1}{a_t}})(1-\beta_1)]}{1-\beta_1}$.

Since $\beta_1 + \exp(-\beta_3 b_t Y_t^{\frac{1}{a_t}})(1-\beta_1) - 1 = (\beta_1 - 1) - \exp(-\beta_3 b_t Y_t^{\frac{1}{a_t}})(\beta_1 - 1) = (\beta_1 - 1)(1 - \exp(-\beta_3 b_t Y_t^{\frac{1}{a_t}}))$ and $\begin{cases} \beta_1 - 1 < 0, \\ 1 - \exp(-\beta_3 b_t Y_t^{\frac{1}{a_t}}) > 0, \end{cases} \Rightarrow \beta_1 + \exp(-\beta_3 b_t Y_t^{\frac{1}{a_t}})(1-\beta_1) < 1$

$\Rightarrow \log b_{t+1} > \frac{\beta_0 - \beta_2}{1-\beta_1}$.

If $\log b_t < \frac{\beta_0}{1-\beta_1}$, then $\log b_{t+1} = \beta_0 + \beta_1 \log b_t - \beta_2 \exp(-\beta_3 b_t Y_t^{\frac{1}{a_t}}) < \beta_0 + \beta_1 \frac{\beta_0}{1-\beta_1} - \beta_2 \exp(-\beta_3 b_t Y_t^{\frac{1}{a_t}})$

$< \beta_0 + \beta_1 \frac{\beta_0}{1-\beta_1} = \frac{\beta_0}{1-\beta_1}$. We have $\log b_t$ is absorbing. $\log a_t$ absorbing can be similarly verified. Hence G is absorbing.

Step 3: Prove all five conditions.

For (a):

It is equivalent to prove Λ being topologically transitive with respect to T^k, for all $k = 1, 2, 3, \cdots$.

It is also equivalent to: for any two open sets U, V of Λ and for all k, $\exists\, n$ such that $T^{nk}(U) \cap V \neq \varnothing$, which means Λ cannot be decomposed into disjoint attractors under T^k for all k.

Given Λ= $\begin{bmatrix} \frac{\alpha_0+m_1}{1-\alpha_1} \\ \frac{\beta_0-m_2}{1-\beta_1} \end{bmatrix}$ is a singleton, (a) holds.

For (b):

We want to apply Proposition 3.8 in Freĭdlin and Wentzell (1988), which indicates that if T and G defined above satisfy certain restrictions, then the hyperbolic attractor Λ will have the following property: $\exists$ c,r $>$0 such that $\forall x \in G$, we can find $y \in \Lambda$ for all n $\geq$ 0, dist$(T^n(x), T^n(y)) \leq ce^{-rn}$dist$(x, y)$.

We check the conditions first: T restricted to G is twice continuously differentiable; Λ is a hyperbolic attractor; and G satisfies I.4.2 in Freĭdlin and Wentzell (1988), which means: G is an open neighborhood of Λ; the closure $\bar{G}$ is compact; and $\bar{G}$ is disjoint from other basic equivalence classes expect for Λ.

Then by Proposition 3.8 in Freĭdlin and Wentzell (1988): $\exists$ c,r $>$0 such that $\forall x \in G$, we can find $y \in \Lambda$ for all n $\geq$ 0, dist$(T^n(x), T^n(y)) \leq ce^{-rn}$dist$(x, y)$.
The result implies $\exists\, K$ such that dist$(T^n(x), T^n(y)) \leq Ke^{-rn}$.
Since $T^n(y) = y$ for all n, we set dist$(T^n(x), y) \leq Ke^{-rn}$, and then (b) holds.

For (c):

We want to prove T is Lipschitz continuous over G. $\Leftrightarrow$ $\exists$ constant k, such that $\frac{\|T(x)-T(y)\|}{\|x-y\|} \leq k$ for $x, y \in G$.

Let x= $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$,y= $\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$. Then,

$$\frac{\|T(x)-T(y)\|}{\|x-y\|} = \frac{\sqrt{[(\beta_0+\beta_1 x_2-m_2)-(\beta_0+\beta_1 y_2-m_2)]^2+[(\alpha_0+\alpha_1 x_1+m_1)-(\alpha_0+\alpha_1 y_1+m_1)]^2}}{\sqrt{(x_1-y_1)^2+(x_2-y_2)^2}}$$
$$= \frac{\sqrt{[\beta_1(x_2-y_2)]^2+[\alpha_1(x_1-y_1)]^2}}{\sqrt{(x_1-y_1)^2+(x_2-y_2)^2}}.$$

Since $0 \leq \alpha_1, \beta_1 < 1, \Rightarrow \frac{\|T(x)-T(y)\|}{\|x-y\|} \leq 1$.

For (e):

For $\mathbf{X}_t \in G$,

$$E\left[\mid S(\mathbf{X}_t, Y_t) \mid \Big| \mathbf{X}_t = x \right]$$

$$= E\left[\begin{bmatrix} \mid -m_1 + \alpha_2 \exp[-\alpha_3 (b_t Y_t^{\frac{1}{a_t}})] \mid \\ \mid m_2 - \beta_2 \exp[-\beta_3 (b_t Y_t^{\frac{1}{a_t}})] \mid \end{bmatrix} \mid \begin{bmatrix} a_t = a \\ b_t = b \end{bmatrix} \right]$$

$$= E\left[\begin{bmatrix} \mid -m_1 + \alpha_2 \exp[-\alpha_3 (b Y_t^{\frac{1}{a}})] \mid \\ \mid m_2 - \beta_2 \exp[-\beta_3 (b Y_t^{\frac{1}{a}})] \mid \end{bmatrix} \right].$$

Since $\exp[-\beta_3 (b Y_t^{\frac{1}{a}})] < 1$ and $\exp[-\alpha_3 (b Y_t^{\frac{1}{a}})] < 1, \Rightarrow \mid -m_1 + \alpha_2 \exp[-\alpha_3 (b Y_t^{\frac{1}{a}})] \mid$ is bounded above and $\mid m_2 - \beta_2 \exp[-\beta_3 (b Y_t^{\frac{1}{a}})] \mid$ is bounded above. Therefore, (e) holds.

For (d):

$$S(\mathbf{X}_t, Y_t) = \begin{bmatrix} -m_1 + \alpha_2 \exp(-\alpha_3 b_t Y_t^{\frac{1}{a_t}}) \\ m_2 - \beta_2 \exp(-\beta_3 b_t Y_t^{\frac{1}{a_t}}) \end{bmatrix}, \text{ where } \begin{cases} 0 < m_2 < \beta_2, \\ m_1 = \alpha_2 \exp(\frac{\alpha_3}{\beta_3} \log(\frac{m_2}{\beta_2})). \end{cases}$$

$S(\mathbf{X}_t, Y_t) = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Leftrightarrow Y_t = [-\frac{1}{\beta_3 b_t} \log(\frac{m_2}{\beta_2})]^{a_t}$ (define as Y_t'). So $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ is support of $S(\mathbf{X}_t, Y_t)$ for all $\mathbf{X}_t \in G$.

Then: $\begin{cases} \log a_t = \alpha_0 + \alpha_1 \log a_{t-1} + m_1 + (-m_1) + \alpha_2 \exp(-\alpha_3 b_{t-1} Y_{t-1}^{\frac{1}{a_{t-1}}}), \\ \log b_t = \beta_0 + \beta_1 \log b_{t-1} - m_2 + m_2 - \beta_2 \exp(-\beta_3 b_{t-1} Y_{t-1}^{\frac{1}{a_{t-1}}}), \end{cases}$

$$\Rightarrow \begin{cases} \log a_t = \alpha_0 + \alpha_1 [\alpha_0 + \alpha_1 \log a_{t-2} + m_1 + (-m_1) + \alpha_2 \exp(-\alpha_3 X_{t-2})] \\ + m_1 + (-m_1) + \alpha_2 \exp(-\alpha_3 b_{t-1} Y_{t-1}^{\frac{1}{a_{t-1}}}), \\ \log b_t = \beta_0 + \beta_1 [\beta_0 + \beta_1 \log b_{t-2} - m_2 + m_2 - \beta_2 \exp(-\beta_3 X_{t-2})] \\ - m_2 + m_2 - \beta_2 \exp(-\beta_3 b_{t-1} Y_{t-1}^{\frac{1}{a_{t-1}}}), \end{cases}$$

$$\Rightarrow \begin{cases} \log a_t = \alpha_0 + \alpha_1[\alpha_0 + \alpha_1 \log a_{t-2} + m_1] \\ -m_1 + \alpha_1[(-m_1) + \alpha_2 \exp(-\alpha_3 X_{t-2})] + (-m_1) + \alpha_2 \exp(-\alpha_3 b_{t-1} Y_{t-1}^{\frac{1}{a_{t-1}}}), \\ \log b_t = \beta_0 + \beta_1[\beta_0 + \beta_1 \log b_{t-2} - m_2] \\ +m_2 + \beta_1[a_1 - \beta_2 \exp(-\beta_3 X_{t-2})] + m_2 - \beta_2 \exp(-\beta_3 b_{t-1} Y_{t-1}^{\frac{1}{a_{t-1}}}), \end{cases}$$

$$\Rightarrow \begin{cases} \log a_t = T^2(\begin{bmatrix} \log(a_{t-2}) \\ \log(b_{t-2}) \end{bmatrix})[1] \\ +\alpha_1[(-m_1) + \alpha_2 \exp(-\alpha_3 b_{t-2} Y_{t-2}^{\frac{1}{a_{t-2}}})] + (-m_1) + \alpha_2 \exp(-\alpha_3 b_{t-1} Y_{t-1}^{\frac{1}{a_{t-1}}}), \\ \log b_t = T^2(\begin{bmatrix} \log(a_{t-2}) \\ \log(b_{t-2}) \end{bmatrix})[2] \\ +\beta_1[m_2 - \beta_2 \exp(-\beta_3 b_{t-2} Y_{t-2}^{\frac{1}{a_{t-2}}})] + m_2 - \beta_2 \exp(-\beta_3 b_{t-1} Y_{t-1}^{\frac{1}{a_{t-1}}}). \end{cases}$$

Let $\begin{bmatrix} \log(a_t) \\ \log(b_t) \end{bmatrix} = \begin{bmatrix} f_1(Y_{t-1}, Y_{t-2}) \\ f_2(Y_{t-1}, Y_{t-2}) \end{bmatrix} = f(Y_{t-1}, Y_{t-2})$. Then the Jacobian matrix of $f(Y_{t-1}, Y_{t-2})$ at (Y'_{t-1}, Y'_{t-2}) is:

$$\begin{bmatrix} \frac{\partial f_1(Y_{t-1},Y_{t-2})}{\partial Y_{t-1}} & \frac{\partial f_1(Y_{t-1},Y_{t-2})}{\partial Y_{t-2}} \\ \frac{\partial f_2(Y_{t-1},Y_{t-2})}{\partial Y_{t-1}} & \frac{\partial f_2(Y_{t-1},Y_{t-2})}{\partial Y_{t-2}} \end{bmatrix} \Bigg| \begin{bmatrix} Y_{t-1} = Y'_{t-1} \\ Y_{t-2} = Y'_{t-2} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{-\alpha_2\alpha_3 b_{t-1}}{a_{t-1}} Y_{t-1}'^{\frac{1}{a_{t-1}}-1} \exp(-\alpha_3 b_{t-1} Y_{t-1}'^{\frac{1}{a_{t-1}}}) & \frac{-\alpha_1\alpha_2\alpha_3 b_{t-2}}{a_{t-2}} Y_{t-2}'^{\frac{1}{a_{t-2}}-1} \exp(-\alpha_3 b_{t-2} Y_{t-2}'^{\frac{1}{a_{t-2}}}) \\ \frac{\beta_2\beta_3 b_{t-1}}{a_{t-1}} Y_{t-1}'^{\frac{1}{a_{t-1}}-1} \exp(-\beta_3 b_{t-1} Y_{t-1}'^{\frac{1}{a_{t-1}}}) & \frac{\beta_1\beta_2\beta_3 b_{t-2}}{a_{t-2}} Y_{t-2}'^{\frac{1}{a_{t-2}}-1} \exp(-\beta_3 b_{t-2} Y_{t-2}'^{\frac{1}{a_{t-2}}}) \end{bmatrix}.$$

The determinant of the Jacobian matrix is:

$\frac{\beta_2\beta_3\alpha_2\alpha_3 b_{t-1} b_{t-2}}{a_{t-1} a_{t-2}} (\alpha_1 - \beta_1) Y_{t-1}'^{\frac{1}{a_{t-1}}-1} Y_{t-2}'^{\frac{1}{a_{t-2}}-1} \exp(-\beta_3 b_{t-1} Y_{t-1}'^{\frac{1}{a_{t-1}}} - \alpha_3 b_{t-2} Y_{t-2}'^{\frac{1}{a_{t-2}}})$. Since

$$\begin{cases} \alpha_2, \alpha_3, \beta_2, \beta_3 > 0, \\ a_{t-1}, a_{t-2}, b_{t-1}, b_{t-2} > 0, \\ \alpha_1 \neq \beta_1, \\ Y'_{t-1}, Y'_{t-2} > 0, \end{cases}$$

$\Rightarrow$ the determinant$\neq$0.

$\Rightarrow \exists$ a neighborhood of point (Y'_{t-1}, Y'_{t-2}), such that $f(.)$ invertible.

$\Rightarrow f^{-1}(.)$ exists at open neighborhood of $f((Y'_{t-1}, Y'_{t-2}) = T^2(\begin{bmatrix}\log(a_{t-2})\\ \log(b_{t-2})\end{bmatrix})$, which can be defined as $B\left(T^2(\begin{bmatrix}\log(a_{t-2})\\ \log(b_{t-2})\end{bmatrix}), c\right)$ where c is a constant.

$\Rightarrow \exists\, K = 2$ such that the second step transition probability for $\begin{bmatrix}\log(a_{t-2})\\ \log(b_{t-2})\end{bmatrix}$ has an absolutely continuous component whose probability density function is positive over $B\left(T^2(\begin{bmatrix}\log(a_{t-2})\\ \log(b_{t-2})\end{bmatrix}), c\right)$.

□

5.1.2 Proof of Consistency and Asymptotic Normality

Need to mention that during the proof of consistency, asymptotic normality and uniqueness, for convenience we let θ to be (μ, α_0, α_1, α_2, α_3, β_0, β_1, β_2, β_3) and it can be extended to be (μ, α_0, α_1, α_2, α_3, β_0, β_1, β_2, β_3, p, q). Similarly, in Section 5.2, we let ϕ to be (μ, γ_0, γ_1, γ_2, γ_3, λ_0, λ_1, λ_2, λ_3), and it can be extended to be (μ, γ_0, γ_1, γ_2, γ_3, λ_0, λ_1, λ_2, λ_3, a, p).

Proof of Theorem 2.1.2:

Proof. Want to prove $\exists$ a local maximizer $\widehat{\theta_n}$ of $L_n(\theta)$ for $\{X_t\}_{t=1}^n$ from the model with θ^0 such that

$$\begin{cases} \widehat{\theta_n} \xrightarrow{p} \theta^0, \\ \|\widehat{\theta_n} - \theta^0\| \leq \delta_n, \end{cases}$$

for

$$\begin{cases} \delta_n \searrow 0, \\ n\delta_n \to +\infty, \\ \frac{1}{\delta_n}(\min\{X_i\}_{i=1}^t - \mu^0) \to +\infty, \end{cases}$$

where $L_n(.)$ is likelihood with arbitrary initial value a_1, b_1.

To prove $L_n(\theta)$ has a local maximizer that converges to θ^0, we can try to prove $h_n(y,z) = \frac{1}{\delta_n^2} L_n(\mu^0 + \delta_n y, \psi^0 + \delta_n z)$, where $z \in \mathbb{R}$, $y \in \mathbb{R}^8$ and ψ^0 = (α_0^0, α_1^0, α_2^0, α_3^0, β_0^0, β_1^0, β_2^0, β_3^0), has a local maximizer $\widehat{y}$, $\widehat{z}$ over open set $\|(y,z)\| < 1$.

Since

$$\begin{cases} \mu^0 + \delta_n \widehat{y} \to \mu^0, \\ \psi^0 + \delta_n \widehat{z} \to \psi^0, \\ \|(\mu_0 + \delta_n \widehat{y} - \mu_0, \psi_0 + \delta \widehat{z} - \psi_0 + \delta \widehat{z})\| \leq \delta_n, \end{cases}$$

by using Lemma 5 in Smith (1985) , we need to show

$$x \frac{\partial h_n(x)}{\partial x} = y \frac{\partial h_n(y,z)}{\partial y} + \sum_{i=1}^{8} z_i \frac{\partial h_n(y,z)}{\partial z_i}$$

is negative when $\|x\| = 1$ ($\|(y,z)\| = 1$).

We start from $\frac{\partial h_n(y,z)}{\partial y}$.

Let $\mu = \mu^0 + \delta_n y$, $\psi = \psi^0 + \delta_n z$, we have $\frac{\partial h_n(y,z)}{\partial y} = \frac{1}{\delta_n} \frac{\partial L_n(\mu^0 + \delta_n y, \psi^0 + \delta_n z)}{\partial \mu}$. By Taylor expansion of $\frac{1}{\delta_n} \frac{\partial L_n(\mu^0 + \delta_n y, \psi^0 + \delta_n z)}{\partial \mu}$ at $(\mu, \psi) = (\mu^0, \psi^0)$, we have:

$$\begin{aligned}
\frac{1}{\delta_n}\frac{\partial L_n(\mu^0+\delta_n y,\psi^0+\delta_n z)}{\partial \mu} &= \frac{1}{\delta_n}\frac{\partial L_n(\mu^0,\psi^0)}{\partial \mu} + \frac{1}{\delta_n}\frac{\partial^2 L_n(\mu^c,\psi^c)}{\partial \mu^2}(\mu-\mu_0) \\
&\quad + \frac{1}{\delta_n}\sum_{i=1}^{8}\frac{\partial^2 L_n(\mu^c,\psi^c)}{\partial \mu \partial \psi_i}(\psi_i-\psi_i^0) \\
&= \frac{1}{\delta_n}\frac{\partial L_n(\mu^0,\psi^0)}{\partial \mu} + \frac{\partial^2 L_n(\mu^c,\psi^c)}{\partial \mu^2}y + \sum_{i=1}^{8}\frac{\partial^2 L_n(\mu^c,\psi^c)}{\partial \mu \partial \psi_i}(\psi_i-\psi_i^0)z_i \\
&= \overbrace{\frac{1}{\delta_n}\frac{\partial L_n(\mu^0,\psi^0)}{\partial \mu} - \frac{1}{\delta_n}\frac{\partial L_n^0(\mu^0,\psi^0)}{\partial \mu}}^{\text{(Part 1)}} \\
&\quad + \overbrace{\frac{1}{\delta_n}\frac{\partial L_n^0(\mu^0,\psi^0)}{\partial \mu}}^{\text{(Part 2)}} \\
&\quad + \overbrace{\frac{\partial^2 L_n(\mu^c,\psi^c)}{\partial \mu^2}y + \sum_{i=1}^{8}\frac{\partial^2 L_n(\mu^c,\psi^c)}{\partial \mu \partial \psi_i}(\psi_i-\psi_i^0)z_i}^{\text{(Part 3)}},
\end{aligned}$$

where $\mid \mu^c-\mu_0 \mid < \delta_n y$, $\mid \psi_i^c-\psi_i^0 \mid < \delta_n z_i$ (i from 1 to 8) and $L_n^0(.)$ is the likelihood with true initial value a_1', b_1'.

Then we analyze the limiting behavior of Part 1, Part 2, and Part 3.

For Part 3, we want to show $\frac{\partial^2 L_n(\mu^c,\psi^c)}{\partial \mu^2} \xrightarrow{p} -m_{\mu\mu}(\theta^0)$, where $m_{\theta_i\theta_j} = -E_{\theta^0}(\frac{\partial^2 l_t^0(\theta^0)}{\partial \theta_i \partial \theta_j})$. Since by the law of large number we have $\frac{\partial^2 L_n^0(\theta^0)}{\partial \mu^2} \xrightarrow{p} -m_{\mu\mu}(\theta^0)$ ($E_{\theta^0}(\frac{\partial l_t^0(\theta^0)}{\partial \theta}) = 0$, $Var_{\theta^0}(\frac{\partial l_t^0(\theta^0)}{\partial \theta}) = -E_{\theta^0}(\frac{\partial^2 l_t^0(\theta^0)}{\partial \theta \partial \theta^T})$), we want to prove:

$$\begin{cases}
(*)_1 : \frac{\partial^2 L_n(\theta)}{\partial \mu^2} \xrightarrow{p} \frac{\partial^2 L_n^0(\theta)}{\partial \mu^2}, \\
(*)_2 : \frac{\partial^2 L_n^0(\theta)}{\partial \mu^2} \xrightarrow{p} \frac{\partial^2 L_n^0(\theta^0)}{\partial \mu^2},
\end{cases}$$

for $\|\theta-\theta^0\| \leq \delta_n$.

For $(*)_2$, we have:

$\frac{\partial L_n(\theta)}{\partial \mu} = \frac{1}{n}\sum_{t=1}^{n}[(-1)(a_t p-1)\frac{1}{X_t-\mu} - (p+q)\frac{1}{1+(\frac{X_t-\mu}{b_t})^{a_t}}\frac{a_t}{-b_t}(\frac{X_t-\mu}{b_t})^{a_t-1}]$,

$\frac{\partial^2 L_n(\theta)}{\partial \mu^2} = \frac{1}{n}\sum_{t=1}^{n}\big[(a_t p-1)(X_t-\mu)^{-2} + (p+q)\frac{a_t(1-a_t)}{b_t^2}(\frac{X_t-\mu}{b_t})^{a_t-2}[1+(\frac{X_t-\mu}{b_t})^{a_t}]^{-1} + (p+q)\frac{a_t^2}{b_t^2}(\frac{X_t-\mu}{b_t})^{2a_t-2}[1+(\frac{X_t-\mu}{b_t})^{a_t}]^{-2}\big]$.

Then:

$$\frac{\partial^2 L_n^0(\theta)}{\partial \mu^2} - \frac{\partial^2 L_n^0(\theta^0)}{\partial \mu^2} = \overbrace{\frac{1}{n}\sum_{t=1}^{n}[(a_t' p-1)(X_t-\mu)^{-2} - (a_t'^0 p-1)(X_t-\mu^0)^{-2}]}^{\text{Part 3(1)}}$$

$$+(p+q)\overbrace{\frac{1}{n}\sum_{t=1}^{n}\big[\frac{a_t'(1-a_t')}{b_t'^2}(\frac{X_t-\mu}{b_t'})^{a_t'-2}[1+(\frac{X_t-\mu}{b_t'})^{a_t'}]^{-1} - \frac{a_t'^0(1-a_t'^0)}{b_t'^{02}}(\frac{X_t-\mu^0}{b_t'^0})^{a_t'^0-2}[1+(\frac{X_t-\mu^0}{b_t'^0})^{a_t'^0}]^{-1}\big]}^{\text{Part 3(2)}}$$

$$+(p+q)\overbrace{\frac{1}{n}\sum_{t=1}^{n}\big[\frac{a_t'^2}{b_t'^2}(\frac{X_t-\mu}{b_t'})^{2a_t'-2}[1+(\frac{X_t-\mu}{b_t'})^{a_t'}]^{-2} - \frac{a_t'^{02}}{b_t'^{02}}(\frac{X_t-\mu^0}{b_t'^0})^{2a_t'^0-2}[1+(\frac{X_t-\mu^0}{b_t'^0})^{a_t'^0}]^{-2}\big]}^{\text{Part 3(3)}}$$

where a_t, b_t is generated with θ and arbitrary a_1, b_1; a_t', b_t' is generated with θ and true a_1', b_1'; a_t^0, b_t^0 is generated with θ^0 and arbitrary a_1, b_1 while $a_t'^0, b_t'^0$ is generated with θ^0 and true a_1', b_1'.

Then we analyze the limiting behavior of Part 3(1), Part 3(2), and Part 3(3) respectively:

For Part 3(1):

$$|\text{ Part 3(1) }| \le \overbrace{\frac{1}{n}\sum_{t=1}^{n}|\ (a_t' p-1)(X_t-\mu)^{-2} - (a_t' p-1)(X_t-\mu^0)^{-2}\ |}^{\text{Part 3(1)(1)}}$$
$$+\overbrace{\frac{1}{n}\sum_{t=1}^{n}|\ (a_t' p-1)(X_t-\mu^0)^{-2} - (a_t'^0 p-1)(X_t-\mu^0)^{-2}\ |}^{\text{Part 3(1)(2)}}.$$

We have Part 3(1)(1) $= \frac{1}{n}\sum_{t=1}^{n}|\ (a_t' p-1)\ ||\ (X_t-\mu)^{-2} - (X_t-\mu^0)^{-2}\ |$. By Lemma 2, Part 3(1)(1)→0. Part 3(1)(2) $= \sum_{t=1}^{n} p\ |\ (a_t'-a_t'^0)\ ||\ (X_t-\mu^0)^{-2}\ |$. By Lemma 3 and Lemma 1, Part 3(1)(2)→0. So, Part 3(1)→0.

For Part 3(2):

| Part 3(2) |≤

$$\overbrace{\frac{1}{n}\sum_{t=1}^{n}\mid\frac{a_t'(1-a_t')}{b_t'^2}\mid\mid\big[(\frac{X_t-\mu}{b_t'})^{a_t'-2}[1+(\frac{X_t-\mu}{b_t'})^{a_t'})]^{-1}-(\frac{X_t-\mu^0}{b_t'})^{a_t'-2}[1+(\frac{X_t-\mu^0}{b_t'})^{a_t'}]^{-1}\big]\mid}^{\text{Part 3(2)(1)}}$$

$$+\overbrace{\frac{1}{n}\sum_{t=1}^{n}\mid\frac{a_t'(1-a_t')}{b_t'^2}\mid\mid\big[(\frac{X_t-\mu^0}{b_t'})^{a_t'-2}[1+(\frac{X_t-\mu^0}{b_t'})^{a_t'})]^{-1}-(\frac{X_t-\mu^0}{b_t'^0})^{a_t'^0-2}[1+(\frac{X_t-\mu^0}{b_t'^0})^{a_t'^0}]^{-1}\big]\mid}^{\text{Part 3(2)(2)}}$$

$$+\overbrace{\frac{1}{n}\sum_{t=1}^{n}\mid\frac{a_t'(1-a_t')}{b_t'^2}-\frac{a_t'^0(1-a_t'^0)}{b_t'^0{}^2}\mid(\frac{X_t-\mu^0}{b_t'^0})^{a_t'^0-2}[1+(\frac{X_t-\mu^0}{b_t'^0})^{a_t'^0}]^{-1}}^{\text{Part 3(2)(3)}}.$$

We can analyze the three parts: Part 3(2)(1), Part 3(2)(2) and Part 3(2)(3) separately.

$$\text{Part 3(2)(1)}\le\overbrace{\frac{1}{n}\sum_{t=1}^{n}\mid\frac{a_t'(1-a_t')}{b_t'^2}\mid\mid\frac{(\frac{X_t-\mu}{b_t'})^{a_t'-2}}{1+(\frac{X_t-\mu}{b_t'})^{a_t'}}-\frac{(\frac{X_t-\mu}{b_t'})^{a_t'-2}}{1+(\frac{X_t-\mu^0}{b_t'})^{a_t'}}\mid}^{\text{Part 3(2)(1)(1)}}$$

$$+\overbrace{\frac{1}{n}\sum_{t=1}^{n}\mid\frac{a_t'(1-a_t')}{b_t'^2}\mid\mid\frac{(\frac{X_t-\mu}{b_t'})^{a_t'-2}-(\frac{X_t-\mu^0}{b_t'})^{a_t'-2}}{1+(\frac{X_t-\mu^0}{b_t'})^{a_t'}}\mid}^{\text{Part 3(2)(1)(2)}}.$$

$$\begin{aligned}\text{Part 3(2)(1)(1)}&\le\frac{1}{n}\sum_{t=1}^{n}\mid\frac{a_t'(1-a_t')}{b_t'^2}\mid\mid\frac{(\frac{X_t-\mu}{b_t'})^{a_t'-2}}{(\frac{X_t-\mu}{b_t'})^{a_t'}}-\frac{(\frac{X_t-\mu}{b_t'})^{a_t'-2}}{(\frac{X_t-\mu^0}{b_t'})^{a_t'}}\mid\\&=\frac{1}{n}\sum_{t=1}^{n}\mid(1-a_t')a_t'\mid(X_t-\mu)^{a_t'-2}\mid\frac{(X_t-\mu^0)^{a_t'}-(X_t-\mu)^{a_t'}}{(X_t-\mu)^{a_t'}(X_t-\mu^0)^{a_t'}}\mid\\&=\frac{1}{n}\sum_{t=1}^{n}\mid(1-a_t')a_t'\mid(X_t-\mu)^{-2}(X_t-\mu^0)^{-a_t'}\mid(X_t-\mu^0)^{a_t'}-(X_t-\mu)^{a_t'}\mid.\end{aligned}$$

By Lemma 2 and Lemma 1, Part 3(2)(1)(1)→0.

Since Part 3(2)(1)(2)= $\frac{1}{n}\sum_{t=1}^{n}\mid a_t'(1-a_t')b_t'^{-a_t'}\mid\frac{1}{1+(\frac{X_t-\mu^0}{b_t'})^{a_t'}}\mid(X_t-\mu)^{a_t'-2}-(X_t-\mu^0)^{a_t'-2}\mid$, by Lemma 2 and Lemma 1, Part 3(2)(1)(2)→0.

So Part 3(2)(1)→0.

$$\text{Part 3(2)(2)} \leq \overbrace{\frac{1}{n}\sum_{t=1}^{n} \mid \frac{a'_t(1-a'_t)}{b'^2_t} \mid\mid \frac{(\frac{X_t-\mu^0}{b'_t})^{a'_t-2}}{1+(\frac{X_t-\mu^0}{b'_t})^{a'_t}} - \frac{(\frac{X_t-\mu^0}{b'_t})^{a'_t-2}}{1+(\frac{X_t-\mu^0}{b'^0_t})^{a'^0_t}} \mid}^{\text{Part 3(2)(2)(1)}}$$
$$+ \overbrace{\frac{1}{n}\sum_{t=1}^{n} \mid \frac{a'_t(1-a'_t)}{b'^2_t} \mid\mid \frac{(\frac{X_t-\mu^0}{b'_t})^{a'_t-2} - (\frac{X_t-\mu^0}{b'^0_t})^{a'^0_t-2}}{1+(\frac{X_t-\mu^0}{b'^0_t})^{a'^0_t}} \mid}^{\text{Part 3(2)(2)(2)}}$$

$$\begin{aligned}
\text{Part 3(2)(2)(1)} &\leq \frac{1}{n}\sum_{t=1}^{n} \mid \frac{a'_t(1-a'_t)}{b'^2_t} \mid\mid \frac{(\frac{X_t-\mu^0}{b'_t})^{a'_t-2}}{(\frac{X_t-\mu^0}{b'_t})^{a'_t}} - \frac{(\frac{X_t-\mu^0}{b'_t})^{a'_t-2}}{(\frac{X_t-\mu^0}{b'^0_t})^{a'^0_t}} \mid \\
&= \frac{1}{n}\sum_{t=1}^{n} \mid a'_t(1-a'_t)b'^{-a'_t}_t \mid (X_t-\mu^0)^{a'_t-2} \mid \frac{(\frac{X_t-\mu^0}{b'^0_t})^{a'^0_t} - (\frac{X_t-\mu^0}{b'_t})^{a'_t}}{(\frac{X_t-\mu^0}{b'_t})^{a'_t}(\frac{X_t-\mu^0}{b'^0_t})^{a'^0_t}} \mid \\
&= \frac{1}{n}\sum_{t=1}^{n} \mid a'_t(1-a'_t)(b'^0_t)^{a'^0_t} \mid (X_t-\mu^0)^{-2-a'^0_t} \mid (b'^0_t)^{-a'^0_t}(X_t-\mu^0)^{a'^0_t} - b'^{-a'_t}_t(X_t-\mu^0)^{a'_t} \mid \\
&= \frac{1}{n}\sum_{t=1}^{n} \mid a'_t(1-a'_t)(b'^0_t)^{a'^0_t} \mid (X_t-\mu^0)^{-2-a'^0_t} \mid (b'^0_t)^{-a'^0_t} \mid\mid (X_t-\mu^0)^{a'^0_t} - (X_t-\mu^0)^{a'_t} \mid \\
&\quad + \frac{1}{n}\sum_{t=1}^{n} \mid a'_t(1-a'_t)(b'^0_t)^{a'^0_t} \mid (X_t-\mu^0)^{-2-a'^0_t} \mid (b'^0_t)^{-a'^0_t} - (b'_t)^{-a'_t} \mid\mid (X_t-\mu^0)^{a'_t} \mid \\
&= \frac{1}{n}\sum_{t=1}^{n} \mid a'_t(1-a'_t) \mid (X_t-\mu^0)^{-2-a'^0_t} \mid (X_t-\mu^0)^{a'^0_t} - (X_t-\mu^0)^{a'_t} \mid \\
&\quad + \frac{1}{n}\sum_{t=1}^{n} \mid a'_t(1-a'_t)(b'^0_t)^{a'^0_t} \mid (X_t-\mu^0)^{a'_t-2-a'^0_t} \mid (b'^0_t)^{-a'^0_t} - (b'_t)^{-a'_t} \mid .
\end{aligned}$$

By Lemma 5, Lemma 3 and Lemma 1, Part 3(2)(2)(1)→0.

$$\begin{aligned}
\text{Part 3(2)(2)(2)} &= \frac{1}{n}\sum_{t=1}^{n} \mid \frac{a'_t(1-a'_t)}{b'^2_t} \mid \frac{1}{1+(\frac{X_t-\mu^0}{b'^0_t})^{a'^0_t}} b'^{2-a'_t}_t \mid (X_t-\mu^0)^{a'_t-2} - (X_t-\mu^0)^{a'^0_t-2} \mid \\
&\quad + \frac{1}{n}\sum_{t=1}^{n} \mid \frac{a'_t(1-a'_t)}{b'^2_t} \mid \frac{1}{1+(\frac{X_t-\mu^0}{b'^0_t})^{a'^0_t}} (X_t-\mu^0)^{a'^0_t-2} \mid b'^{2-a'_t}_t - (b'^0_t)^{2-a'^0_t} \mid .
\end{aligned}$$

Also, by Lemma 5, Lemma 3 and Lemma 1, Part 3(2)(2)(2)→0.

So Part 3(2)(2)→0.

Part 3(2)(3) $\leq \frac{1}{n}\sum_{t=1}^{n} \mid \frac{a_t'(1-a_t')}{b_t'^2} - \frac{a_t'^0(1-a_t'^0)}{b_t'^{0\,2}} \mid (\frac{X_t-\mu^0}{b_t'^0})^{-2}$. By Lemma 3 and Lemma 1, Part 3(2)(3)→0.

Putting all things together, we have Part 3(2)→0.

Similar to Part 3(2), we can prove Part 3(3)→0. So we have $(*)_2$ proved ($\frac{\partial^2 L_n^0(\theta)}{\partial \mu^2} \xrightarrow{p} \frac{\partial^2 L_n^0(\theta^0)}{\partial \mu^2}$). Similar to $(*)_2$, by Lemma 6 and Lemma 4 (instead of Lemma 5 and Lemma 3) we can prove $(*)_1$. So we have $\frac{\partial^2 L_n(\mu^c,\psi^c)}{\partial \mu^2} \xrightarrow{p} -m_{\mu\mu}(\theta^0)$, where $m_{\theta_i\theta_j} = -E_{\theta^0}(\frac{\partial^2 L_n^0(\theta^0)}{\partial\theta_i \partial\theta_j})$. We can also prove $\frac{\partial^2 L_n(\mu^c,\psi^c)}{\partial \mu \partial\psi_i} \xrightarrow{p} -m_{\mu\psi_i}(\theta^0)$, by similar arguments as above. Now we have Part 3 $\xrightarrow{p} -m_{\mu\mu}(\theta^0)y - \sum_{i=1}^{8} m_{\mu\psi_i}(\theta^0)z_i$, over $\|(y,z)\| \leq 1$.

Next we want to prove Part 1 $\to 0$.

$$\text{Part 1} = \overbrace{\frac{1}{\delta_n}\frac{1}{n}\sum_{t=1}^{n}[(1-a_t^0)\frac{1}{X_t-\mu^0} - (1-a_t'^0)\frac{1}{X_t-\mu^0}]}^{\text{Part 1(1)}}$$
$$+\overbrace{\frac{1}{\delta_n n}\sum_{t=1}^{n}(p+q)[\frac{a_t'^0}{b_t'^0}(\frac{X_t-\mu^0}{b_t'^0})^{a_t'^0-1}(1+(\frac{X_t-\mu^0}{b_t'^0})^{a_t'^0})^{-1} - \frac{a_t^0}{b_t^0}(\frac{X_t-\mu^0}{b_t^0})^{a_t^0-1}(1+(\frac{X_t-\mu^0}{b_t^0})^{a_t^0})^{-1}]}^{\text{Part 1(2)}}.$$

For Part 1(1):

$\mid$ Part 1(1) $\mid = \frac{1}{\delta_n}\frac{1}{n}\sum_{t=1}^{n} \mid a_t^0 - a_t'^0 \mid \frac{1}{X_t-\mu^0}$. By Lemma 4 $\mid$ Part 1(1) $\mid \leq \frac{C}{\delta_n n}\sum_{t=1}^{n} U_{r,b}^{t-1}\frac{1}{X_t-\mu^0}$. This converges to zero, since $\sum_{t=1}^{n} U_{r,b}^{t-1}\frac{1}{X_t-\mu^0}$ is finite.

For Part 1(2):

$$\mid \text{Part 1(2)} \mid \leq \overbrace{\frac{p+q}{n\delta_n}\sum_{t=1}^{n}\frac{a_t'^0}{b_t'^0} \mid \frac{(\frac{X_t-\mu^0}{b_t'^0})^{a_t'^0-1}}{1+(\frac{X_t-\mu^0}{b_t'^0})^{a_t'^0}} - \frac{(\frac{X_t-\mu^0}{b_t^0})^{a_t^0-1}}{1+(\frac{X_t-\mu^0}{b_t^0})^{a_t^0}} \mid}^{\text{Part 1(2)(1)}}$$
$$+\overbrace{\frac{p+q}{n\delta_n}\sum_{t=1}^{n} \mid \frac{a_t'^0}{b_t'^0} - \frac{a_t^0}{b_t^0} \mid \frac{(\frac{X_t-\mu^0}{b_t^0})^{a_t^0-1}}{1+(\frac{X_t-\mu^0}{b_t^0})^{a_t^0}}}^{\text{Part 1(2)(2)}}.$$

We analyze Part 1(2)(1) and Part 1(2)(2) separately.

According to Part3 (2)(2), we have

$$\text{Part 1(2)(1)} \leq \overbrace{\frac{p+q}{n\delta_n}\sum_{t=1}^{n} \mid a_t'^0 b_t'^0 \mid (X_t-\mu^0)^{-1-a_t^0} \mid (X_t-\mu^0)^{a_t^0} - (X_t-\mu^0)^{a_t'^0} \mid}^{\text{Part 1(2)(1)(1)}}$$
$$+ \overbrace{\frac{p+q}{n\delta_n}\sum_{t=1}^{n} \mid a_t'^0 b_t'^0 (b_t^0)^{a_t^0} \mid (X_t-\mu^0)^{a_t'^0-1-a_t^0} \mid (b_t^0)^{-a_t^0} - (b_t'^0)^{-a_t'^0} \mid}^{\text{Part 1(2)(1)(2)}}$$
$$+ \overbrace{\frac{p+q}{n\delta_n}\sum_{t=1}^{n} \mid a_t'^0 b_t'^0 \mid \frac{1}{1+(\frac{X_t-\mu^0}{b_t^0})^{a_t^0}} (b_t'^0)^{1-a_t'^0} \mid (X_t-\mu^0)^{a_t'^0-1} - (X_t-\mu^0)^{a_t^0-1} \mid}^{\text{Part 1(2)(1)(3)}}$$
$$+ \overbrace{\frac{p+q}{n\delta_n}\sum_{t=1}^{n} \mid a_t'^0 b_t'^0 \mid \frac{1}{1+(\frac{X_t-\mu^0}{b_t^0})^{a_t^0}} (X_t-\mu^0)^{a_t^0-1} \mid (b_t'^0)^{1-a_t'^0} - (b_t^0)^{1-a_t^0} \mid}^{\text{Part 1(2)(1)(4)}}.$$

Then we analyze the four parts: Part 1(2)(1)(1), Part 1(2)(1)(2), Part 1(2)(1)(3) and Part 1(2)(1)(4).

For Part 1(2)(1)(1), by the mean value theorem:

$$\text{Part 1(2)(1)(1)} \leq \tfrac{p+q}{n\delta_n}\sum_{t=1}^{n} \mid a_t'^0 b_t'^0 \mid (X_t-\mu^0)^{-1-a_t^0}(X_t-\mu^0)^{a_t^*} \mid \log(X_t-\mu^0) \mid\mid a_t^0 - a_t'^0 \mid,$$

where a_t^* between a_t^0 and $a_t'^0$. By Lemma 4,

$$\text{Part 1(2)(1)(1)} \leq \tfrac{C(p+q)}{n\delta_n}\sum_{t=1}^{n} \mid a_t'^0 b_t'^0 \mid (X_t-\mu^0)^{-1-a_t^0}(X_t-\mu^0)^{a_t^*} \mid \log(X_t-\mu^0) \mid U_{r,b}^{t-1}.$$

$\sum_{t=1}^{n} \mid a_t'^0 b_t'^0 \mid (X_t-\mu^0)^{-1-a_t^0}(X_t-\mu^0)^{a_t^*} \mid \log(X_t-\mu^0) \mid U_{r,b}^{t-1}$ finite, and $n\delta_n \to \infty$, we have Part 1(2)(1)(1) converge to zero.

For Part 1(2)(1)(2):

$$\text{Part 1(2)(1)(2)} \leq \tfrac{C(p+q)}{n\delta_n}\sum_{t=1}^{n} \mid a_t'^0 b_t'^0 (b_t^0)^{a_t^0} \mid (X_t-\mu^0)^{a_t'^0-1-a_t^0} U_{r,b}^{t-1}.$$

$\sum_{t=1}^{n} \mid a_t'^0 b_t'^0 (b_t^0)^{a_t^0} \mid (X_t-\mu^0)^{a_t'^0-1-a_t^0} U_{r,b}^{t-1}$ finite, we have Part 1(2)(1)(2) converge to zero.

Part 1(2)(1)(3) and Part 1(2)(1)(4) are similar to Part1(2)(1)(1) and Part 1(2)(1)(2). We have Part 1(2)(1)(3) converge to zero and Part 1(2)(1)(4) converge to zero. So Part 1(2)(1) converge to zero.

As for Part 1(2)(2):

Part 1(2)(2) $\leq \frac{p+q}{n\delta_n} \sum_{t=1}^{n} \mid \frac{a_t'^0}{b_t'^0} - \frac{a_t^0}{b_t^0} \mid (\frac{X_t-\mu^0}{b_t^0})^{-1} \leq \frac{C(p+q)}{n\delta_n} \sum_{t=1}^{n} (\frac{X_t-\mu^0}{b_t^0})^{-1} U_{r,b}^{t-1}$.

Since $\sum_{t=1}^{n} (\frac{X_t-\mu^0}{b_t^0})^{-1} U_{r,b}^{t-1}$ finite, we have Part 1(2)(2) converge to zero.

Putting all proofs together, Part 1 converges to zero.

For Part2:

Part 2 = $\frac{1}{\delta_n} \frac{\partial L_n^0(\mu^0,\psi^0)}{\partial \mu} = \frac{1}{\delta_n n} \sum_{t=1}^{n} \frac{\partial l_t^0(\mu^0,\psi^0)}{\partial \mu}$. $E_{\theta_0}(\frac{\partial l_t^0(\theta^0)}{\partial \theta}) = 0$ and $\delta_n n \to \infty$ imply Part 2 converge to zero.

According to results of Part 1, Part 2 and Part 3, we have $\frac{\partial h_n(y,z)}{\partial y} \xrightarrow{p} -m_{\mu\mu}(\theta^0)y - \sum_{i=1}^{8} m_{\mu\psi_i}(\theta^0)z_i$, over $\|(y,z)\| \leq 1$.

Similarly, we can show $\frac{\partial h_n(y,z)}{\partial z_k} \xrightarrow{p} -m_{\mu\psi_k}(\theta^0)y - \sum_{i=1}^{8} m_{\psi_k\psi_i}(\theta^0)z_i$, over $\|(y,z)\| \leq 1$. So, $x\frac{\partial h_n(x)}{\partial x} = y\frac{\partial h_n(y,z)}{\partial y} + \sum_{i=1}^{8} z_i \frac{\partial h_n(y,z)}{\partial z_i} \xrightarrow{p} -y^2 m_{\mu\mu}(\theta^0) - 2y \sum_{i=1}^{8} m_{\mu\psi_i}(\theta^0)z_i - \sum_{i=1}^{8}\sum_{j=1}^{8} m_{\psi_j\psi_i}(\theta^0)z_i z_j$, over $\|(y,z)\| \leq 1$.

By defining $\mathbf{W} = \begin{bmatrix} y, & z_1, & z_2, & z_3, & z_4, & z_5, & z_6, & z_7, & z_8 \end{bmatrix}$, we have:

$$
\begin{aligned}
& - y^2 m_{\mu\mu}(\theta^0) - 2y \sum_{i=1}^{8} m_{\mu\psi_i}(\theta^0)z_i - \sum_{i=1}^{8}\sum_{j=1}^{8} m_{\psi_j\psi_i}(\theta^0)z_i z_j \\
& = (-1)\mathbf{W} \begin{bmatrix} m_{\mu\mu}(\theta^0) & m_{\mu\psi_1}(\theta^0) & \cdots & m_{\mu\psi_8}(\theta^0) \\ m_{\psi_1\mu}(\theta^0) & m_{\psi_1\psi_1}(\theta^0) & \cdots & m_{\psi_1\psi_8}(\theta^0) \\ \vdots & & & \\ & & & m_{\psi_8\psi_8}(\theta^0) \end{bmatrix} \mathbf{W}^{\mathbf{T}} \\
& = (-1)\mathbf{W}(-E_{\theta^0}(\frac{\partial^2 l_t^0(\theta^0)}{\partial\theta\partial\theta^T}))\mathbf{W}^{\mathbf{T}}.
\end{aligned}
$$

Since $-E_{\theta^0}(\frac{\partial^2 l_t^0(\theta^0)}{\partial\theta\partial\theta^T}) = Var_{\theta^0}(\frac{\partial l_t^0(\theta^0)}{\theta})$, which is positive definite, we have: at $\|(y,z)\| = 1$, $x\frac{\partial h_n(x)}{\partial x}$ negative.

The proof is complete.

□

Proof of Theorem 2.1.3:

Proof. We want to show $\sqrt{n}(\widehat{\theta_n} - \theta^0) \xrightarrow{d} N\left(0, \frac{1}{-E_{\theta^0}(\frac{\partial^2 l_t^0(\theta^0)}{\partial\theta\partial\theta^T})}\right)$.

We start from $\widehat{\mu_n}$. By the mean value theorem, we have: $\frac{\frac{\partial L_n(\widehat{\theta_n}')}{\partial\mu} - \frac{\partial L_n(\theta^0)}{\partial\mu}}{\widehat{\mu_n} - \mu^0} = \frac{\partial^2 L_n(\theta^*)}{\partial\mu^2}$, where θ^* is between $\widehat{\theta_n}$ and θ^0, $\widehat{\theta_n}' = (\widehat{\mu_n}, \beta_0^0, \beta_1^0, \beta_2^0, \beta_3^0, \alpha_0^0, \alpha_1^0, \alpha_2^0, \alpha_3^0)$, which means $\widehat{\theta_n}'$ is θ^0 replacing μ_0 by $\widehat{\mu_n}$.

Since $\widehat{\mu_n}$ is the local maximizer of $L_n(.)$ with $(\beta_0^0, \beta_1^0, \beta_2^0, \beta_3^0, \alpha_0^0, \alpha_1^0, \alpha_2^0, \alpha_3^0)$ given, $\frac{\partial L_n(\widehat{\theta_n}')}{\partial\mu} = 0$.

$\Rightarrow \sqrt{n}(\widehat{\mu_n} - \mu^0) = -\frac{\sqrt{n}\frac{\partial L_n(\theta^0)}{\partial\mu}}{\frac{\partial^2 L_n(\theta^*)}{\partial\mu^2}}$.

The denominator $\frac{\partial^2 L_n(\theta^*)}{\partial\mu^2} \to E_{\theta^0}(\frac{\partial^2 l_t^0(\theta^0)}{\partial\mu^2})$, by Theorem 2. The numerator $\sqrt{n}\frac{\partial L_n(\theta^0)}{\partial\mu} \to \sqrt{n}[\frac{1}{n}\sum_{t=1}^{n}\frac{\partial l_t^0(\theta^0)}{\partial\mu}] = \sqrt{n}[\frac{1}{n}\sum_{t=1}^{n}\frac{\partial l_t^0(\theta^0)}{\partial\mu} - E(\frac{l_t^0(\theta^0)}{\partial\mu})] \xrightarrow{d} N(0, -E_{\theta^0}(\frac{\partial^2 l_t^0(\theta^0)}{\partial\mu^2}))$.

So, $-\frac{\sqrt{n}\frac{\partial L_n(\theta^0)}{\partial\mu}}{\frac{\partial^2 L_n(\theta^*)}{\partial\mu^2}} \xrightarrow{d} N\left(0, (-E_{\theta^0}(\frac{\partial^2 l_t^0(\theta^0)}{\partial\mu^2})(-E_{\theta^0}(\frac{\partial^2 l_t^0(\theta^0)}{\partial\mu^2}))^{-2}\right) = N\left(0, \frac{1}{-E_{\theta^0}(\frac{\partial^2 l_t^0(\theta^0)}{\partial\mu^2})}\right)$.

Then we can extend the result to a vector and get the conclusion:

$\sqrt{n}(\widehat{\theta_n} - \theta^0) \xrightarrow{d} N\left(0, \frac{1}{-E_{\theta^0}(\frac{\partial^2 l_t^0(\theta^0)}{\partial\theta\partial\theta^T})}\right)$. □

5.1.3 Proof of Uniqueness

Proof of Theorem 2.1.4:

Proof. There is an asymptotic unique MLE over M_n, where $M_n = \{\theta \in \Theta | \mu \leq c\min_{1\leq t\leq n} X_t + (1-c)\mu_0\}$; Θ is a compact set of $\{\theta | \alpha_0, \beta_0, \mu \in \mathbb{R}, 0 \leq \alpha_1, \beta_1 < 1, \alpha_1 \neq \beta_1, \alpha_2, \alpha_3, \beta_2, \beta_3 > 0\}$; $0 < c < 1$.

To prove uniqueness, we apply Theorem 2.6 in Makelainen et al. (1981). The theorem is shown below: Let $l(\theta)$ be the logarithm of a twice continuously differentiable likelihood function, with θ varying in a connected open subset $\Theta \subset \mathbb{R}^p$. Suppose that (i) the gradient vector ∇l vanishes in at least one point $\theta \in \Theta$ and that (ii) the Hessian matrix of $l(\theta)$ is negative definite at every point $\theta \in \Theta$. Then (a) $l(\theta)$ is a strictly concave function of θ. (b) there is a unique maximum likelihood estimate $\theta \in \Theta$, and (c) $l(\theta)$ has no other maxima or mimima or other stationary points in Θ.

So we want to prove conditions (i) and (ii). First, define $\Theta^a = \{\theta \in M_n | \|\theta - \theta_0\| \geq a\}$, $\Theta^{a^c} = \{\theta \in M_n | \|\theta - \theta_0\| < a\}$, $\Theta^{a\mu_0} = \{\theta \in M_n | \|\theta - \theta_0\| \geq a, \mu > \mu_0\}$, $\Theta^{a\mu_0^c} = \{\theta \in$

$M_n|\|\theta-\theta_0\| \geq a, \mu \leq \mu_0\}$, $\Theta^{a^c\mu_0} = \{\theta \in M_n|\|\theta-\theta_0\| < a, \mu > \mu_0\}$ and $\Theta^{a^c\mu_0^c} = \{\theta \in M_n|\|\theta-\theta_0\| < a, \mu \leq \mu_0\}$. We will have $\Theta^a = \Theta^{a\mu_0} \cup \Theta^{a\mu_0^c}$ and $\Theta^{a^c} = \Theta^{a^c\mu_0} \cup \Theta^{a^c\mu_0^c}$.

Since $\sup_{\Theta^a} \mid L_n(\theta) - L_n^0(\theta) \mid \xrightarrow{p} 0$, $\sup_{\Theta^a} L_n(\theta) = \sup_{\Theta^a} L_n^0(\theta) + Op(1)$
$= \max(\sup_{\Theta^{a\mu_0}} L_n^0(\theta), \sup_{\Theta^{a\mu_0^c}} L_n^0(\theta))$. Since $\sup_{\Theta^{a\mu_0}} \mid L_n^0(\theta) - L_n^0(\mu_0,\phi) \mid \xrightarrow{p} 0$, $\sup_{\Theta^{a\mu_0}} L_n(\theta) = \sup_{\Theta^{a\mu_0}} L_n^0(\mu_0,\phi) + Op(1)$. So $\max(\sup_{\Theta^{a\mu_0}} L_n^0(\theta), \sup_{\Theta^{a\mu_0^c}} L_n^0(\theta)) + Op(1) = \max(\sup_{\Theta^{a\mu_0}} L_n^0(\mu_0,\phi), \sup_{\Theta^{a\mu_0^c}} L_n^0(\theta)) + Op(1)$.

Since $\min_{1<t<n}(X_t) \searrow \mu_0$, $\{\phi|\phi \in \Theta^{a\mu_0}\} \subseteq \{\phi|\phi \in \Theta^{a/2\mu_0^c}\}$. This implies $\sup_{\Theta^{a\mu_0}} L_n^0(\mu_0,\phi) \leq \sup_{\Theta^{a/2\mu_0^c}} L_n^0(\theta)$. Also we have $\sup_{\Theta^{a\mu_0^c}} L_n^0(\theta)) \leq \sup_{\Theta^{a/2\mu_0^c}} L_n^0(\theta)$, so $\max(\sup_{\Theta^{a\mu_0}} L_n^0(\mu_0,\phi), \sup_{\Theta^{a\mu_0^c}} L_n^0(\theta)) \leq \sup_{\Theta^{a/2\mu_0^c}} L_n^0(\theta)$. So we have $\sup_{\Theta^a} L_n(\theta) \leq \sup_{\Theta^{a/2\mu_0^c}} L_n^0(\theta) + Op(1)$. Then we have $\sup_{\Theta^a} L_n(\theta) \leq L_n^0(\theta_0) + Op(1)$ and base on $L_n^0(\theta_0) = L_n(\theta_0) + Op(1)$, we can conclude $\sup_{\Theta^a} L_n(\theta) \leq L_n(\theta_0) + Op(1)$, which is equivalent to that the global maximizer of $L_n(\theta)$ ($\widehat{\theta}$ satisfies $\frac{\partial L_n(\widehat{\theta})}{\partial\theta} = 0$) over M_n located within Θ^{a^c}. So condition (i) is proved.

For condition (ii), we have $\sup_{\Theta^{a^c}} \mid \frac{\partial^2}{\partial\theta_i\partial\theta_j} L_n(\theta) - \frac{\partial^2}{\partial\theta_i\partial\theta_j} L_n^0(\theta) \mid \xrightarrow{p} 0$. For $\Theta^{a^c\mu_0}$, we have $\sup_{\Theta^{a^c\mu_0}} \mid \frac{\partial^2}{\partial\theta_i\partial\theta_j} L_n^0(\theta) - \frac{\partial^2}{\partial\theta_i\partial\theta_j} L_n^0(\mu_0,\phi) \mid \xrightarrow{p} 0$. For $\Theta^{a^c\mu_0^c}$, we have $\sup_{\Theta^{a^c\mu_0^c}} \mid \frac{\partial^2}{\partial\theta_i\partial\theta_j} L_n^0(\theta) - E_{\theta_0}(\frac{\partial^2}{\partial\theta_i\partial\theta_j} l_1(\theta)) \mid \xrightarrow{p} 0$ by ergodicity and Uniform Law of Large numbers. Since $E_{\theta_0}(\frac{\partial^2}{\partial\theta_i\partial\theta_j} l_1(\theta_0))$ is negative definite, we can find an a' such that $E_{\theta_0}(\frac{\partial^2}{\partial\theta_i\partial\theta_j} l_1(\theta))$ is negative definite for $\theta \in \Theta^{a'^c\mu_0^c}$. So we can conclude that $\frac{\partial^2}{\partial\theta_i\partial\theta_j} L_n(\theta)$ is negative definite over $\Theta^{a'^c}$ and (ii) is proved.

Under the condition of (i) and (ii), we have uniqueness proved.

□

5.1.4 Proof of Lemmas

Lemma 5.1.1. *1)* : $\frac{1}{n}\sum_{t=1}^{n}(X_t-\mu^0)^a$ *is finite as* $n \to \infty$*, where* $'a'$ *is a constant. 2)* : $\frac{1}{n}\sum_{t=1}^{n}(X_t-\mu)^a$ *finite as* $n \to \infty$*, where* $'a'$ *is a constant,* $\mid \mu-\mu^0 \mid < \delta_n$*,* $\delta_n \searrow 0$*,* $n\delta_n \to +\infty$*, and* $\frac{1}{\delta_n}(\min\{X_i\}_{i=1}^t - \mu^0) \to +\infty$.

For Lemma 1.1:

Proof. $b_U max(Y_t^{\frac{1}{a_L}}, Y_t^{\frac{1}{a_U}}) > X_t - \mu^0 > b_L \min(Y_t^{\frac{1}{a_L}}, Y_t^{\frac{1}{a_U}})$, where b_U, b_L is the upper bound and lower bound of b_t, a_U, a_L is the upper bound and lower bound of a_t.

By the fact that $0 < Y_t < \infty$, for support $\in (0, \infty)$, $(X_t - \mu^0)^a$ finite for all t.

Then $\frac{1}{n}\sum_{t=1}^{n}(X_t - \mu^0)^a \xrightarrow{p} E_{\theta^0}[(X_1 - \mu^0)^a]$ is finite, by law of large number, ergodicity and stationary. □

For Lemma 1.2:

Proof. We have four situations.

Situation 1 ($a < 0$, $\mu > \mu^0$):

$\frac{1}{n}\sum_{t=1}^{n}(X_t-\mu)^a = \frac{1}{n}\sum_{t=1}^{n}(X_t-\mu^0+\mu^0-\mu)^a < \frac{1}{n}\sum_{t=1}^{n}(X_t-\mu^0-\delta_n)^a$. According to $\frac{1}{\delta_n}(\min\{X_i\}_{i=1}^{t}-\mu^0) \to +\infty$,

$\Rightarrow \exists 0 < c < 1$ such that $P(c(\min\{X_i\}_{i=1}^{t} - \mu^0) > \delta_n) \to 1$

$\Rightarrow \exists 0 < c < 1$ such that $P(c(X_t - \mu^0) > \delta_n) \to 1$.

$\Rightarrow \frac{1}{n}\sum_{t=1}^{n}(X_t - \mu)^a < \frac{1}{n}\sum_{t=1}^{n}[X_t - \mu^0 - c(X_t - \mu^0)]^a = \frac{1}{n}\sum_{t=1}^{n}[(1-c)(X_t - \mu^0)]^a$, which is finite as $n \to \infty$ according to Lemma 1.1.

Situation 2 ($a < 0$, $\mu < \mu^0$):

$\frac{1}{n}\sum_{t=1}^{n}(X_t-\mu)^a = \frac{1}{n}\sum_{t=1}^{n}(X_t-\mu^0+\mu^0-\mu)^a < \frac{1}{n}\sum_{t=1}^{n}(X_t-\mu^0)^a$,which is finite as $n \to \infty$ according to Lemma 1.1.

Situation 3 ($a > 0$, $\mu > \mu^0$):

$\frac{1}{n}\sum_{t=1}^{n}(X_t - \mu)^a = \frac{1}{n}\sum_{t=1}^{n}(X_t - \mu^0 + \mu^0 - \mu)^a < \frac{1}{n}\sum_{t=1}^{n}(X_t - \mu^0)^a$, which is finite as $n \to \infty$ according to Lemma 1.1.

Situation 4 ($a > 0$, $\mu < \mu^0$):

$\frac{1}{n}\sum_{t=1}^{n}(X_t-\mu)^a = \frac{1}{n}\sum_{t=1}^{n}(X_t-\mu^0+\mu^0-\mu)^a < \frac{1}{n}\sum_{t=1}^{n}(X_t-\mu^0+\delta_n)^a$. According to $\frac{1}{\delta_n}(\min\{X_i\}_{i=1}^{t}-\mu^0) \to +\infty$,

$\Rightarrow \exists 0 < c < 1$ such that $P(c(\min\{X_i\}_{i=1}^{t} - \mu^0) > \delta_n) \to 1$

$\Rightarrow \exists 0 < c < 1$ such that $P(c(X_t - \mu^0) > \delta_n) \to 1$.

$\Rightarrow \frac{1}{n}\sum_{t=1}^{n}(X_t-\mu)^a < \frac{1}{n}\sum_{t=1}^{n}[X_t-\mu^0+c(X_t-\mu^0)]^a = \frac{1}{n}\sum_{t=1}^{n}[(1+c)(X_t-\mu^0)]^a$, which is finite as $n\to\infty$ according to Lemma 1.1. □

Lemma 5.1.2. *$\mid \frac{1}{n}\sum_{t=1}^{n}(X_t-\mu)^a - \frac{1}{n}\sum_{t=1}^{n}(X_t-\mu^0)^a \mid \leq O_p(\delta_n)$, with $\mid \mu-\mu^0 \mid < \delta_n$, $\delta_n \searrow 0$, $n\delta_n \to +\infty$, and $\frac{1}{\delta_n}(\min\{X_i\}_{i=1}^{t}-\mu^0)\to+\infty$.*

Proof. We have four situations.

Situation 1 ($a<0$, $\mu>\mu^0$):

$\mid \frac{1}{n}\sum_{t=1}^{n}(X_t-\mu)^a - \frac{1}{n}\sum_{t=1}^{n}(X_t-\mu^0)^a \mid$

$\leq \frac{1}{n}\sum_{t=1}^{n} \mid (X_t-\mu)^a-(X_t-\mu^0)^a \mid$

$= \frac{1}{n}\sum_{t=1}^{n}(X_t-\mu)^{a-1}(X_t-\mu^0) \mid (X_t-\mu)(X_t-\mu^0)^{-1}-[(X_t-\mu)(X_t-\mu^0)^{-1}]^{1-a} \mid$

$\leq \frac{1}{n}\sum_{t=1}^{n}(X_t-\mu)^{a-1}(X_t-\mu^0)(1-a) \mid 1-(X_t-\mu)(X_t-\mu^0)^{-1} \mid$ (by the fact that $x-x^{m+1}\leq (m+1)(1-x)$ with $m>0, 0<x<1$)

$= \frac{1}{n}\sum_{t=1}^{n}(1-a)(\mu-\mu^0)(X_t-\mu)^{a-1}.$

According to Lemma 1, $\mid \frac{1}{n}\sum_{t=1}^{n}(X_t-\mu)^a - \frac{1}{n}\sum_{t=1}^{n}(X_t-\mu^0)^a \mid \leq O_p(\delta_n)$.

Situation 2 ($a<0$, $\mu<\mu^0$):

$\mid \frac{1}{n}\sum_{t=1}^{n}(X_t-\mu)^a - \frac{1}{n}\sum_{t=1}^{n}(X_t-\mu^0)^a \mid$

$\leq \frac{1}{n}\sum_{t=1}^{n} \mid (X_t-\mu)^a-(X_t-\mu^0)^a \mid$

$= \frac{1}{n}\sum_{t=1}^{n}(X_t-\mu)(X_t-\mu^0)^{a-1} \mid (X_t-\mu)^{-1}(X_t-\mu^0)-[(X_t-\mu)^{-1}(X_t-\mu^0)]^{1-a} \mid$

$\leq \frac{1}{n}\sum_{t=1}^{n}(X_t-\mu)(X_t-\mu^0)^{a-1}(1-a) \mid 1-(X_t-\mu)^{-1}(X_t-\mu^0) \mid$ (by the fact that $x-x^{m+1}\leq (m+1)(1-x)$ with $m>0, 0<x<1$)

$= \frac{1}{n}\sum_{t=1}^{n}(1-a)(\mu^0-\mu)(X_t-\mu^0)^{a-1}.$

According to Lemma 1, $\mid \frac{1}{n}\sum_{t=1}^{n}(X_t-\mu)^a - \frac{1}{n}\sum_{t=1}^{n}(X_t-\mu^0)^a \mid \leq O_p(\delta_n)$.

Situation 3 ($a>0$, $\mu>\mu^0$):

$| \frac{1}{n}\sum_{t=1}^{n}(X_t-\mu)^a - \frac{1}{n}\sum_{t=1}^{n}(X_t-\mu^0)^a |$
$\leq \frac{1}{n}\sum_{t=1}^{n} | (X_t-\mu)^a - (X_t-\mu^0)^a |$
$= \frac{1}{n}\sum_{t=1}^{n}(X_t-\mu)^{-1}(X_t-\mu^0)^{1+a} | (X_t-\mu)(X_t-\mu^0)^{-1} - [(X_t-\mu)(X_t-\mu^0)^{-1}]^{1+a} |$
$\leq \frac{1}{n}\sum_{t=1}^{n}(X_t-\mu)^{-1}(X_t-\mu^0)^{1+a}(1+a) | 1-(X_t-\mu)(X_t-\mu^0)^{-1} |$ (by the fact that $x-x^{m+1} \leq (m+1)(1-x)$ with $m>0, 0<x<1$)
$= \frac{1}{n}\sum_{t=1}^{n}(1+a)(\mu-\mu^0)(X_t-\mu)^{-1}(X_t-\mu^0)^a.$

According to Lemma 1, $| \frac{1}{n}\sum_{t=1}^{n}(X_t-\mu)^a - \frac{1}{n}\sum_{t=1}^{n}(X_t-\mu^0)^a | \leq O_p(\delta_n)$.

Situation 4 ($a>0$, $\mu<\mu^0$):

$| \frac{1}{n}\sum_{t=1}^{n}(X_t-\mu)^a - \frac{1}{n}\sum_{t=1}^{n}(X_t-\mu^0)^a |$
$\leq \frac{1}{n}\sum_{t=1}^{n} | (X_t-\mu)^a - (X_t-\mu^0)^a |$
$= \frac{1}{n}\sum_{t=1}^{n}(X_t-\mu)^{1+a}(X_t-\mu^0)^{-1} | (X_t-\mu)^{-1}(X_t-\mu^0) - [(X_t-\mu)^{-1}(X_t-\mu^0)]^{1+a} |$
$\leq \frac{1}{n}\sum_{t=1}^{n}(X_t-\mu)^{1+a}(X_t-\mu^0)^{-1}(1+a) | 1-(X_t-\mu)^{-1}(X_t-\mu^0) |$ (by the fact that $x-x^{m+1} \leq (m+1)(1-x)$ with $m>0, 0<x<1$)
$= \frac{1}{n}\sum_{t=1}^{n}(1+a)(\mu^0-\mu)(X_t-\mu)^a(X_t-\mu^0)^{-1}.$

According to Lemma 1, $| \frac{1}{n}\sum_{t=1}^{n}(X_t-\mu)^a - \frac{1}{n}\sum_{t=1}^{n}(X_t-\mu^0)^a | \leq O_p(\delta_n)$.

So Lemma 2 is proved.

□

Lemma 5.1.3. *$\|(\beta_0,\beta_1,\beta_2,\beta_3)-(\beta_0^0,\beta_1^0,\beta_2^0,\beta_3^0)\| < \delta_n$ where $\delta_n \searrow 0$, then $| b_t' - b_t'^0 | \leq O(\delta_n)$; $\|(\alpha_0,\alpha_1,\alpha_2,\alpha_3)-(\alpha_0^0,\alpha_1^0,\alpha_2^0,\alpha_3^0)\| < \delta_n$ where $\delta_n \searrow 0$, then $| a_t' - a_t'^0 | \leq O(\delta_n)$.*

Proof. We only prove the situation for a_t, it is similar for b_t.

By auto-regressive, we have:

$$\begin{aligned}\log(a_t) &= \alpha_0 + \alpha_1\log(a_{t-1}) + \alpha_2\exp(-\alpha_3 X_{t-1}) \\ &= \alpha_0\sum_{i=1}^{t-1}\alpha_1^{i-1} + \alpha_1^{t-1}\log(a_1) + \alpha_2\sum_{i=1}^{t-1}\alpha_1^{i-1}\exp(-\alpha_3 X_{t-i}).\end{aligned}$$

So:

$$| \log(a'_t) - \log(a'^0_t) | \leq \overbrace{| \alpha_0 \sum_{i=1}^{t-1} \alpha_1^{i-1} - \alpha_0^0 \sum_{i=1}^{t-1} (\alpha_1^0)^{i-1} |}^{\text{Lemma 3(1)}}$$
$$+ \overbrace{| \alpha_1^{t-1} \log(a'_1) - (\alpha_1^0)^{t-1} \log(a'_1) |}^{\text{Lemma 3(2)}}$$
$$+ \overbrace{| \alpha_2 \sum_{i=1}^{t-1} \alpha_1^{i-1} \exp(-\alpha_3 X_{t-i}) - \alpha_2^0 \sum_{i=1}^{t-1} (\alpha_1^0)^{i-1} \exp(-\alpha_3^0 X_{t-i}) |}^{\text{Lemma 3(3)}}.$$

Then we analyze Lemma 3(1), Lemma 3(2) and Lemma 3(3) one by one.

Lemma 3(1):

$$\text{Lemma 3(1)} = | \alpha_0 \sum_{i=1}^{t-1} \alpha_1^{i-1} - \alpha_0 \sum_{i=1}^{t-1} (\alpha_1^0)^{i-1} + \alpha_0 \sum_{i=1}^{t-1} (\alpha_1^0)^{i-1} - \alpha_0^0 \sum_{i=1}^{t-1} (\alpha_1^0)^{i-1} |$$
$$= \overbrace{| \alpha_0 || \sum_{i=1}^{t-1} \alpha_1^{i-1} - \sum_{i=1}^{t-1} (\alpha_1^0)^{i-1} |}^{\text{Lemma 3(1)(1)}} + \overbrace{| \alpha_0 - \alpha_0^0 || \sum_{i=1}^{t-1} (\alpha_1^0)^{i-1} |}^{\text{Lemma 3(1)(2)}}.$$

$$\begin{aligned}
\text{Lemma 3(1)(1)} &\leq | \alpha_0 || \sum_{i=1}^{\infty} \alpha_1^{i-1} - \sum_{i=1}^{\infty} (\alpha_1^0)^{i-1} | \\
&= | \alpha_0 || \frac{1}{1-\alpha_1} - \frac{1}{1-\alpha_1^0} | \\
&= | \alpha_0 || \alpha_1 - \alpha_1^0 || \frac{1}{(1-\alpha_1)(1-\alpha_1^0)} | \\
&\leq \delta_n | \frac{\alpha_0}{(1-\alpha_1)(1-\alpha_1^0)} | \\
&= O(\delta_n).
\end{aligned}$$

$$\text{Lemma 3(1)(2)} < \delta_n | \sum_{i=1}^{t-1} (\alpha_1^0)^{i-1} | = O(\delta_n).$$

$\Rightarrow$ Lemma 3(1) $\leq O(\delta_n)$.

Lemma 3(2):

Lemma 3(2) $=\mid \alpha_1^{t-1} - (\alpha_1^0)^{t-1} \mid\mid \log(a_1') \mid \leq \mid \alpha_1 - \alpha_1^0 \mid\mid \log(a_1') \mid = O(\delta_n)$.

Lemma 3(3):

$$\text{Lemma 3(3)} \leq \overbrace{\mid \alpha_2 \sum_{i=1}^{t-1} \alpha_1^{i-1} \exp(-\alpha_3 X_{t-i}) - \alpha_2^0 \sum_{i=1}^{t-1} (\alpha_1)^{i-1} \exp(-\alpha_3 X_{t-i}) \mid}^{\text{Lemma 3(3)(1)}}$$
$$+ \overbrace{\mid \alpha_2^0 \sum_{i=1}^{t-1} \alpha_1^{i-1} \exp(-\alpha_3 X_{t-i}) - \alpha_2^0 \sum_{i=1}^{t-1} (\alpha_1^0)^{i-1} \exp(-\alpha_3 X_{t-i}) \mid}^{\text{Lemma 3(3)(2)}}$$
$$+ \overbrace{\mid \alpha_2^0 \sum_{i=1}^{t-1} (\alpha_1^0)^{i-1} \exp(-\alpha_3 X_{t-i}) - \alpha_2^0 \sum_{i=1}^{t-1} (\alpha_1^0)^{i-1} \exp(-\alpha_3^0 X_{t-i}) \mid}^{\text{Lemma 3(3)(3)}}.$$

Then we analyze the three parts: Lemma 3(3)(1), Lemma 3(3)(2) and Lemma 3(3)(3).

Lemma 3(3)(1) $=\mid \alpha_2 - \alpha_2^0 \mid\mid \sum_{i=1}^{t-1} \alpha_1^{i-1} \exp(-\alpha_3 X_{t-i}) \mid = O(\delta_n)$.

Lemma 3(3)(2) $\leq \alpha_2^0 \sum_{i=1}^{t-1} \mid \alpha_1^{i-1} - (\alpha_1^0)^{i-1} \mid \exp(-\alpha_3 X_{t-i}) \mid = O(\delta_n)$.

Lemma 3(3)(3) $\leq \alpha_2^0 \sum_{i=1}^{t-1} (\alpha_1^0)^{i-1} \mid \exp(-\alpha_3 X_{t-i}) - \exp(-\alpha_3^0 X_{t-i}) \mid$. By the mean value theorem:
$\mid \exp(-\alpha_3 X_{t-i}) - \exp(-\alpha_3^0 X_{t-i}) \mid = \mid (-X_{t-i}) \exp(-\alpha_3^* X_{t-i})(\alpha_3 - \alpha_3^0) \mid = X_{t-i} \exp(-\alpha_3^* X_{t-i}) \mid \alpha_3 - \alpha_3^0 \mid$, where α_3^* is between α_3 and α_3^0.

$\Rightarrow$ Lemma 3(3)(3) $\leq \alpha_2^0 \sum_{i=1}^{t-1} (\alpha_1^0)^{i-1} X_{t-i} \exp(-\alpha_3^* X_{t-i}) \mid \alpha_3 - \alpha_3^0 \mid = O(\delta_n)$.

$\Rightarrow \mid \log(a_t') - \log(a_t'^0) \mid \leq O(\delta_n)$.

$\Rightarrow \mid a_t' - a_t'^0 \mid \leq O(\delta_n)$. □

Lemma 5.1.4. *$\exists$ a constant C such that $\mid a_t - a_t' \mid \leq C U_{r,b}^{t-1}$; $\mid b_t - b_t' \mid \leq C U_{r,b}^{t-1}$, where $U_{r,b}$ is the upper bound of β_1, α_1.*

Proof. We only prove the situation for a_t, it is similar for b_t.

We start from $\log a_2$, $\log a_2'$:

$$\begin{cases} \log a_2 = \alpha_0 + \alpha_1 \log a_1 + \alpha_2 \exp(-\alpha_3 x_1), \\ \log a_2' = \alpha_0 + \alpha_1 \log a_1' + \alpha_2 \exp(-\alpha_3 x_1). \end{cases}$$

We have $\mid \log a_2 - \log a_2' \mid = \alpha_1 \mid \log a_1 - \log a_1' \mid$. Then $\mid \log a_3 - \log a_3' \mid = \alpha_1 \mid \log a_2 - \log a_2' \mid = \alpha_1^2(\log a_1 - \log a_1')$. Next, keep auto-regressive process we will have $\mid \log a_t - \log a_t' \mid$
$= \alpha_1^{t-1} \mid \log a_1 - \log a_1' \mid \leq U_{r,b}C_1$ for some constant C_1.

So we can conclude $\mid a_t - a_t' \mid \leq CU_{r,b}^{t-1}$ for some C. □

Lemma 5.1.5. *$\mid a_t' - a_t'^0 \mid \leq O(\delta_n)$, then $\frac{1}{n}\sum_{t=1}^{n} \mid (X_t-\mu)^{-a_t'} - (X_t-\mu)^{-a_t'^0} \mid = O_p(\delta_n)$.*

Proof. By the mean value theorem:
$\frac{1}{n}\sum_{t=1}^{n} \mid (X_t-\mu)^{-a_t'} - (X_t-\mu)^{-a_t'^0} \mid = \frac{1}{n}\sum_{t=1}^{n} \mid a_t' - a_t'^0 \mid (X_t-\mu)^{-a_t^*} \mid \log(X_t-\mu) \mid = O_p(\delta_n)$
by Lemma 1, where a_t^* is between a_t' and $a_t'^0$. □

Lemma 5.1.6. $\frac{1}{n}\sum_{t=1}^{n} \mid (X_t-\mu)^{-a_t} - (X_t-\mu)^{-a_t'} \mid \to 0$.

Proof. By the mean value theorem:
$\frac{1}{n}\sum_{t=1}^{n} \mid (X_t-\mu)^{-a_t} - (X_t-\mu)^{-a_t'} \mid$
$= \frac{1}{n}\sum_{t=1}^{n} \mid a_t - a_t' \mid (X_t-\mu)^{-a_t^*} \mid \log(X_t-\mu) \mid$
$\leq \frac{C}{n}\sum_{t=1}^{n}(X_t-\mu)^{-a_t^*} \mid \log(X_t-\mu) \mid U_{r,b}$ (according to Lemma 4).

Since $\sum_{t=1}^{n}(X_t-\mu)^{-a_t^*} \mid \log(X_t-\mu) \mid U_{r,b}$ is finite, $\frac{C}{n}\sum_{t=1}^{n}(X_t-\mu)^{-a_t^*} \mid \log(X_t-\mu) \mid U_{r,b} \to 0$.
□

5.1.5 Transformation Between Distributions

In this section, we show two transformations between distributions.

1: GB2 (a=1, b=1, p, q) $\Leftrightarrow$ Prime Beta (p, q).

Proof. If X $\sim$ GB2 (a, b, p, q), then the p.d.f of X is: $f(x) = \frac{ax^{ap-1}}{b^{ap}B(p,q)(1+(\frac{x}{b})^a)^{p+q}}$. Let $a = 1, b = 1$, we will have: $f(x) = \frac{x^{p-1}}{B(p,q)(1+x)^{p+q}}$, which is the p.d.f of Prime beta (p, q). □

2:If $Y_t \sim$ Prime Beta (p, q) then, $b_t Y_t^{\frac{1}{a_t}} \sim$ GB2 (a_t, b_t, p, q).

Proof. Let $Y_t \sim$ Prime beta (p, q),

$$
\begin{aligned}
[][P(b_t Y_t^{\frac{1}{a_t}} < y)]' &= [P(b_t Y_t^{\frac{1}{a_t}} < y)]' \\
&= [P(Y_t < (\frac{y}{b_t})^{a_t}]' = [I_{\frac{(\frac{y}{b_t})^{a_t}}{1+(\frac{y}{b_t})^{a_t}}}(p,q)]' \\
&= [\frac{1}{B(p,q)} \int_0^{\frac{(\frac{y}{b_t})^{a_t}}{1+(\frac{y}{b_t})^{a_t}}} z^{p-1}(1-z)^{q-1} dz]' \\
&= \frac{1}{B(p,q)} [\frac{(\frac{y}{b_t})^{a_t}}{1+(\frac{y}{b_t})^{a_t}}]' [\frac{(\frac{y}{b_t})^{a_t}}{1+(\frac{y}{b_t})^{a_t}}]^{p-1} [1 - \frac{(\frac{y}{b_t})^{a_t}}{1+(\frac{y}{b_t})^{a_t}}]^{q-1} \\
&= \frac{1}{B(p,q)} \frac{\frac{a_t}{b_t}(\frac{y}{b_t})^{a_t-1}(1+(\frac{y}{b_t})^{a_t}) - \frac{a_t}{b_t}(\frac{y}{b_t})^{a_t-1}(\frac{y}{b_t})^{a_t}}{(1+(\frac{y}{b_t})^{a_t})^2} \frac{[(\frac{y}{b_t})^{a_t}]^{p-1}}{[1+(\frac{y}{b_t})^{a_t}]^{p+q-2}} \\
&= \frac{1}{B(p,q)} \frac{a_t y^{a_t p-1}}{b_t^{a_t p}(1+(\frac{y}{b_t})^{a_t})^{p+q}},
\end{aligned}
$$

which is the p.d.f of GB2 (a_t, b_t, p, q).
$\Rightarrow b_t Y_t^{\frac{1}{a_t}} = (X_t - \mu) \sim$ GB2 (a_t, b_t, p, q). □

5.2 Proofs for Model with Parameters q and b Dynamic

5.2.1 Proof of Stationarity and Ergodicity

Proof. If we let the model to be:

$$
\begin{cases}
X_t = \mu + F_{Z_t}^{-1}(Z_t), \\
\log q_t = \gamma_0 + \gamma_1 \log a_{t-1} + \gamma_2 \exp(-\gamma_3 X_{t-1}), \\
\log b_t = \lambda_0 + \lambda_1 \log b_{t-1} - \lambda_2 \exp(-\lambda_3 X_{t-1}),
\end{cases}
$$

Where Z_t follows uniform distribution with (0,1); F_{Z_t} is CDF of GB2 distribution with (p, q_t, a, b_t) and F^{-1} is inverse function of F.

Then we will have $X_t - \mu = F_{Z_t}^{-1}(Z_t)$ follows GB2 distribution with (p, q_t, a, b_t).

To write as form of $\mathbf{X}_{n+1} = T(\mathbf{X}_n) + S(\mathbf{X}_n, e_{n+1})$, we let Z_t to be the sequence of independent

and identically distributed random variables e_n. We will have

$$\begin{bmatrix} \log q_t \\ \log b_t \end{bmatrix} = \begin{bmatrix} \gamma_0 + \gamma_1 \log q_{t-1} + m_3 \\ \lambda_0 + \lambda_1 \log b_{t-1} - m_4 \end{bmatrix} + \begin{bmatrix} -m_3 + \gamma_2 \exp(-\gamma_3 X_{t-1}) \\ m_4 - \lambda_2 \exp(-\lambda_3 X_{t-1}) \end{bmatrix}. \tag{5.2.1}$$

For condition (d), we only need to show the determinant of the Jacobian matrix of $f(Z_{t-1}, Z_{t-2})$ at (Z'_{t-1}, Z'_{t-2}) not equal to 0, where (Z'_{t-1}, Z'_{t-2}) satisfy that $-m_3 + \gamma_2 \exp(-\gamma_3 F^{-1}_{Z_{t-1}}(Z'_{t-1})) = 0$ and $m_4 - \lambda_2 \exp(-\lambda_3 F^{-1}_{Z_{t-2}}(Z'_{t-2})) = 0$.

For the Jacobian matrix $\begin{bmatrix} \frac{\partial f_1(Z_{t-1},Z_{t-2})}{\partial Z_{t-1}} & \frac{\partial f_1(Z_{t-1},Z_{t-2})}{\partial Z_{t-2}} \\ \frac{\partial f_2(Z_{t-1},Z_{t-2})}{\partial Z_{t-1}} & \frac{\partial f_2(Z_{t-1},Z_{t-2})}{\partial Z_{t-2}} \end{bmatrix} \Big| \begin{bmatrix} Z_{t-1} = Z'_{t-1} \\ Z_{t-2} = Z'_{t-2} \end{bmatrix}$, we have

$\frac{\partial f_1(Z_{t-1},Z_{t-2})}{\partial Z_{t-1}}$

$= \frac{\partial(\gamma_2 exp(-\gamma_3 F^{-1}_{Z_{t-1}}(Z_{t-1})))}{\partial Z_{t-1}}$

$= \gamma_2 exp(-\gamma_3 F^{-1}_{Z_{t-1}}(Z_{t-1}))(-\gamma_3)\frac{\partial F^{-1}_{Z_{t-1}}(Z_{t-1})}{\partial Z_{t-1}}$

$= \gamma_2 exp(-\gamma_3 F^{-1}_{Z_{t-1}}(Z_{t-1}))(-\gamma_3)\frac{1}{F'(F^{-1}_{Z_{t-1}}(Z_{t-1}))}$.

Since $F_{Z_{t-1}}(z) = \frac{B(w;p,q_{t-1})}{B(p,q_{t-1})}$, where $B(w; p, q_{t-1})$ is incomplete beta function with parameters (w, p, q_{t-1}) and $w = \frac{(\frac{z}{b_{t-1}})^a}{1+(\frac{z}{b_{t-1}})^a}$.

According to Leibniz rule, we will have:

$F'_{Z_{t-1}}(z)$

$= \frac{1}{B(p,q_{t-1})}(w^{p-1}(1-w)^{q_{t-1}-1}\frac{\partial w}{\partial y} - 0)$

$= \frac{1}{B(p,q_{t-1})}((\frac{(\frac{z}{b_{t-1}})^a}{1+(\frac{z}{b_{t-1}})^a})^{p-1}(\frac{1}{1+(\frac{z}{b_{t-1}})^a})^{q_{t-1}-1}\frac{\frac{a}{b_{t-1}}(\frac{z}{b_{t-1}})^{a-1}}{(1+(\frac{z}{b_{t-1}})^a)^2})$

$= \frac{1}{B(p,q_{t-1})}\frac{\frac{a}{b_{t-1}}(\frac{z}{b_{t-1}})^{ap-1}}{(1+(\frac{z}{b_{t-1}})^a)^{p+q_{t-1}}}$.

So, $\frac{1}{F'(F^{-1}_{Z_{t-1}}(Z_{t-1}))} = \frac{B(p,q_{t-1})(1+(\frac{F^{-1}_{Z_{t-1}}(Z_{t-1})}{b_{t-1}})^a)^{p+q_{t-1}}}{\frac{a}{b_{t-1}}(\frac{F^{-1}_{Z_{t-1}}(Z_{t-1})}{b_{t-1}})^{ap-1}}$.

Then we will have $\frac{\partial f_1(Z_{t-1},Z_{t-2})}{\partial Z_{t-1}} = \gamma_2 exp(-\gamma_3 F^{-1}_{Z_{t-1}}(Z_{t-1}))(-\gamma_3)\frac{B(p,q_{t-1})(1+(\frac{F^{-1}_{Z_{t-1}}(Z_{t-1})}{b_{t-1}})^a)^{p+q_{t-1}}}{\frac{a}{b_{t-1}}(\frac{F^{-1}_{Z_{t-1}}(Z_{t-1})}{b_{t-1}})^{ap-1}}$.

Following similar steps, we can calculate other parts of the Jacobian matrix and get the determinant:

$\gamma_2\gamma_3\lambda_2\lambda_3(\gamma_1 - \lambda_1)exp(-\gamma_3 F^{-1}_{Z_{t-1}}(Z'_{t-1}) - \lambda_3 F^{-1}_{Z_{t-2}}(Z'_{t-2}))B(p, q_{t-1})B(p, q_{t-2})\frac{b_{t-1}b_{t-2}}{a^2} *$

$$\frac{(1+(\frac{F^{-1}_{Z_{t-1}}(Z'_{t-1})}{b_{t-1}})^a)^{p+q_{t-1}}(1+(\frac{F^{-1}_{Z_{t-2}}(Z'_{t-2})}{b_{t-2}})^a)^{p+q_{t-2}}}{(\frac{F^{-1}_{Z_{t-1}}(Z'_{t-1})}{b_{t-1}})^{ap-1}(\frac{F^{-1}_{Z_{t-2}}(Z'_{t-2})}{b_{t-2}})^{ap-1}}.$$

Since,

$$\begin{cases} \gamma_2, \gamma_3, \lambda_2, \lambda_3 > 0, \\ \gamma_1 \neq \lambda_1, \\ b_{t-1}, b_{t-2} > 0, \\ B(p, q_{t-1}), B(p, q_{t-2}) > 0, \\ F^{-1}_{Z_{t-1}}(Z'_{t-1}), F^{-1}_{Z_{t-2}}(Z'_{t-2}) > 0, \end{cases}$$

We will have determinant not equal to 0. So condition (d) proved.

As for conditions (a), (b), (c), (e), there is not much difference from before.

□

5.2.2 Proof of Consistency, Asymptotic Normality, and Uniqueness

Proof. For the model with q and b dynamic, the proof of consistency is similar to the proof of Theorem 2.1.2, but with some modifications. Here we have:

$$\frac{\partial L_n(\phi)}{\partial \mu} = \frac{1}{n}\sum_{t=1}^{n}[(-1)(ap-1)\frac{1}{X_t-\mu} - (p+q_t)\frac{1}{1+(\frac{X_t-\mu}{b_t})^a}\frac{a}{-b_t}(\frac{X_t-\mu}{b_t})^{a-1}],$$

$$\frac{\partial^2 L_n(\phi)}{\partial \mu^2} = \frac{1}{n}\sum_{t=1}^{n}\big[(ap-1)(X_t-\mu)^{-2} + (p+q_t)\frac{a(1-a)}{b_t^2}(\frac{X_t-\mu}{b_t})^{a-2}[1+(\frac{X_t-\mu}{b_t})^a)]^{-1} + (p+q_t)\frac{a^2}{b_t^2}(\frac{X_t-\mu}{b_t})^{2a-2}[1+(\frac{X_t-\mu}{b_t})^a)]^{-2}\big].$$

For Part 3, we have $|$ Part 3(1) $| = \frac{1}{n}\sum_{t=1}^{n} \mid (ap-1)(X_t-\mu)^{-2} - (ap-1)(X_t-\mu^0)^{-2} \mid$, which is equal to $\frac{1}{n}\sum_{t=1}^{n} \mid ap-1 \mid\mid (X_t-\mu)^{-2} - (X_t-\mu^0)^{-2}) \mid$ and it's converge to 0.

The original $(p+q)$Part 3(2) (where Part 3(2) $= \frac{1}{n}\sum_{t=1}^{n}\big[\frac{a'_t(1-a'_t)}{b'^2_t}(\frac{X_t-\mu}{b'_t})^{a'_t-2}[1+(\frac{X_t-\mu}{b'_t})^{a'_t})]^{-1} - \frac{a'^0_t(1-a'^0_t)}{b'^{02}_t}(\frac{X_t-\mu^0}{b'^0_t})^{a'^0_t-2}[1+(\frac{X_t-\mu^0}{b'^0_t})^{a'^0_t})]^{-1}\big]$) becomes $\frac{1}{n}\sum_{t=1}^{n}\big[(p+q'_t)\frac{a(1-a)}{b'^2_t}(\frac{X_t-\mu}{b'_t})^{a-2}[1+(\frac{X_t-\mu}{b'_t})^a)]^{-1} - (p+q'^0_t)\frac{a(1-a)}{b'^{02}_t}(\frac{X_t-\mu^0}{b'^0_t})^{a-2}[1+(\frac{X_t-\mu^0}{b'^0_t})^a)]^{-1}\big]$. So the original $(p+q)$ $|$ Part 3(2) $|$ now equals to:

$$\frac{1}{n}\sum_{t=1}^{n} \mid (p+q_t')\frac{a(1-a)}{b_t'^2}(\frac{X_t-\mu}{b_t'})^{a-2}[1+(\frac{X_t-\mu}{b_t'})^a)]^{-1}-(p+q_t'^0)\frac{a(1-a)}{b_t'^{02}}(\frac{X_t-\mu^0}{b_t'^0})^{a-2}[1+(\frac{X_t-\mu^0}{b_t'^0})^a)]^{-1} \mid$$

$$= \frac{1}{n}\sum_{t=1}^{n} \mid (p+q_t')\frac{a(1-a)}{b_t'^2}(\frac{X_t-\mu}{b_t'})^{a-2}[1+(\frac{X_t-\mu}{b_t'})^a)]^{-1}-(p+q_t')\frac{a(1-a)}{b_t'^{02}}(\frac{X_t-\mu^0}{b_t'^0})^{a-2}[1+(\frac{X_t-\mu^0}{b_t'^0})^a)]^{-1}$$
$$+ (p+q_t')\frac{a(1-a)}{b_t'^{02}}(\frac{X_t-\mu^0}{b_t'^0})^{a-2}[1+(\frac{X_t-\mu^0}{b_t'^0})^a)]^{-1} - (p+q_t'^0)\frac{a(1-a)}{b_t'^{02}}(\frac{X_t-\mu^0}{b_t'^0})^{a-2}[1+(\frac{X_t-\mu^0}{b_t'^0})^a)]^{-1} \mid$$
$$\leq \frac{1}{n}\sum_{t=1}^{n}\Bigg[\mid (p+q_t')\frac{a(1-a)}{b_t'^2}(\frac{X_t-\mu}{b_t'})^{a-2}[1+(\frac{X_t-\mu}{b_t'})^a)]^{-1}-(p+q_t')\frac{a(1-a)}{b_t'^{02}}(\frac{X_t-\mu^0}{b_t'^0})^{a-2}[1+(\frac{X_t-\mu^0}{b_t'^0})^a)]^{-1} \mid$$

$$+ \mid (p+q_t')\frac{a(1-a)}{b_t'^{02}}(\frac{X_t-\mu^0}{b_t'^0})^{a-2}[1+(\frac{X_t-\mu^0}{b_t'^0})^a)]^{-1}-(p+q_t'^0)\frac{a(1-a)}{b_t'^{02}}(\frac{X_t-\mu^0}{b_t'^0})^{a-2}[1+(\frac{X_t-\mu^0}{b_t'^0})^a)]^{-1} \mid \Bigg]$$
$$\leq \frac{1}{n}\sum_{t=1}^{n} \mid (p+q_t')\frac{a(1-a)}{b_t'^2}(\frac{X_t-\mu}{b_t'})^{a-2}[1+(\frac{X_t-\mu}{b_t'})^a)]^{-1}-(p+q_t')\frac{a(1-a)}{b_t'^{02}}(\frac{X_t-\mu^0}{b_t'^0})^{a-2}[1+(\frac{X_t-\mu^0}{b_t'^0})^a)]^{-1} \mid$$

$$+\frac{1}{n}\sum_{t=1}^{n} \mid (p+q_t')\frac{a(1-a)}{b_t'^{02}}(\frac{X_t-\mu^0}{b_t'^0})^{a-2}[1+(\frac{X_t-\mu^0}{b_t'^0})^a)]^{-1}-(p+q_t'^0)\frac{a(1-a)}{b_t'^{02}}(\frac{X_t-\mu^0}{b_t'^0})^{a-2}[1+(\frac{X_t-\mu^0}{b_t'^0})^a)]^{-1} \mid$$
$$= \frac{1}{n}\sum_{t=1}^{n} \mid (p+q_t')\Bigg[\frac{a(1-a)}{b_t'^2}(\frac{X_t-\mu}{b_t'})^{a-2}[1+(\frac{X_t-\mu}{b_t'})^a)]^{-1} - \frac{a(1-a)}{b_t'^{02}}(\frac{X_t-\mu^0}{b_t'^0})^{a-2}[1+(\frac{X_t-\mu^0}{b_t'^0})^a)]^{-1} \mid \Bigg]$$
$$+ \frac{1}{n}\sum_{t=1}^{n} \mid (q_t'-q_t'^0)\Bigg[\frac{a(1-a)}{b_t'^{02}}(\frac{X_t-\mu^0}{b_t'^0})^{a-2}[1+(\frac{X_t-\mu^0}{b_t'^0})^a)]^{-1}\Bigg] \mid.$$

The proof of
$\frac{1}{n}\sum_{t=1}^{n} \mid (p+q_t')\Bigg[\frac{a(1-a)}{b_t'^2}(\frac{X_t-\mu}{b_t'})^{a-2}[1+(\frac{X_t-\mu}{b_t'})^a)]^{-1} - \frac{a(1-a)}{b_t'^{02}}(\frac{X_t-\mu^0}{b_t'^0})^{a-2}[1+(\frac{X_t-\mu^0}{b_t'^0})^a)]^{-1} \mid \Bigg]$
converges to 0 is similar and easier than that of Part 3(2). And as for proof of $\frac{1}{n}\sum_{t=1}^{n} \mid (q_t'-q_t'^0)\Bigg[\frac{a(1-a)}{b_t'^{02}}(\frac{X_t-\mu^0}{b_t'^0})^{a-2}[1+(\frac{X_t-\mu^0}{b_t'^0})^a)]^{-1}\Bigg] \mid$ converges to 0, firstly $(q_t'-q_t'^0)$ converges to 0 can be shown as the proof of Lemma 3. Then, the rest is also similar within the proof in Part 3(2).

For original Part 3 $(p+q) \mid$ Part 3(3) $\mid$, we can apply the same idea of dealing with $(p+q_t)$ term as above. So we have $\frac{\partial^2 L_n^0(\phi)}{\partial \mu^2} - \frac{\partial^2 L_n^0(\phi^0)}{\partial \mu^2}$ in Part 3 converges to 0.

For part 1, the only place need to be modified is also the $(p+q_t)$ term, the method is also the same as proof above. There is nothing different in proof of Part 2. So we can complete the proof of consistency.

The proof of asymptotic normality and uniqueness is same as the proof of Theorem 2.1.3 and Theorem 2.1.4. □

Bibliography

Bali, T. and Weinbaum, D. (2007). A conditional extreme value volatility estimator based on high-frequency returns. *Journal of Economic Dynamics and Control*, 31(2):361–397.

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

Chakravarti, I., Laha, R., and Roy, J. (1967). *Handbook of Methods of Applied Statistics, Volume I*. Wiley series in probability and mathematical statistics. John Wiley and Sons.

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

Chavez-Demoulin, V., Embrechts, P., and Sardy, S. (2014). Extreme-quantile tracking for financial time series. *Journal of Econometrics*, 188(1):44–52.

Chen, L. and Singh, V. (2017). Generalized beta distribution of the second kind for flood frequency analysis. *Entropy*, 19(6):254.

Corrado, C. (2001). Option pricing based on the generalized lambda distribution. *Journal of Futures Market*, 21(3):213–236.

Cui, Q. and Ma, Y. (2014). Pricing synthetic cdo with mgb2 distribution. *Statistics and Its Interface*, 7(3):309–318.

Cunnane, c. (1973). A particular comparison of annual maxima and partial duration series methods of flood frequency prediction. *Journal of Hydrology*, 18:257–271.

Diebold, F., Schuermann, T., and Stroughair, J. (1998). Pitfalls and opportunities in the use of extreme value theory in risk management. In Refenes, A., Moody, J., and Burgess, A., editors, *Decision Technologies for Computational Finance*, volume 2 of *Advances in Computational Management Science*, chapter 1, pages 3–12. Springer US.

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.

Ferreira, A. and de Haan, L. (2015). On the block maxima method in extreme value theory: PWM estimators. *Annals of Statistics*, 43(1):276–298.

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(3):165–176.

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.

Frees, E. and Valdez, E. (2008). Hierarchical insurance claims modeling. *Journal of the American Statistical Association*, 103(484):1457–1469.

Freĭdlin, M. and Wentzell, A. (1988). *Random Perturbations of Dynamical Systems*. Springer-Verlag, New York.

Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of Mathematics*, 44(3):423–453.

Goda, Y. (2000). *Random Seas and Design of Maritime Structures*. World Scientific.

Grimshaw, S. (2017). Computing maximum likelihood estimates for the generalized pareto distribution. *Technometrics*, 35(2):185–191.

Gumbel, E. (1958). *Statistics of Extremes*. New York: Colombia University Press.

Hall, S., Swamy, P., and Tavlas, G. (2016). Time-varying coefficient models: A proposal for selecting the coefficient driver sets. *Macroeconomic Dynamics*, pages 1–17.

Hosking, J. and Wallis, J. (1987). Parameter and quantile estimation for the generalized pareto distribution. *Techmometrics*, 29(3):339–349.

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.

Kelly, B. and Jiang, H. (2014). Tail risk and asset prices. *The Review of Financial Studies*, 27(10):2841–2871.

Kleiber, C. and Kotz, S. (2003). *Statistical Size Distributions in Economics and Actuarial Sciences*. Wiley, J.

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.

Lechner, J., Simiu, E., and N., H. (1993). Assessment of 'peaks over threshold' methods for estimating extreme value distribution tails. *Structural Safety*, 12(4):305–314.

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.

Lundbergh, S., Terasvirta, T., and Van Dijk, D. (2003). Time-varying smooth transition autoregressive models. *Journal of Business & Economic Statistics*, 21(1):104–128.

Makelainen, T., Schmidt, K., and Styan, G. (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.

Mao, G. and Zhang, Z. (2018). Stochastic tail index model for high frequency financial data with bayesian analysis. *Journal of Econometrics*, 205(2):470–487.

McDonald, J. and Xu, Y. (1995). A generalizeation of the beta distribution with applications. *Journal of Econometrics*, 66(1-2):133–152.

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.

Rootzen, H. and Finkenstadt, B. (2003). *Extreme Values in Finance, Telecommunications, and the Environment*. Taylor Francis Inc.

Shuway, R. and Stoffer, D. (2017). *Time Series Analysis and Its Applications, Fourth Edition*. Springer Texts in Statistics. Springer.

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

Teyssière, G. and Kirman, A. (2007). *Long Memory in Economics*. Springer, Berlin, Heidelberg.

Yang, X., Frees, E., and Zhang, Z. (2011). A generalized beta copula with applications in modeling multivariate long-tailed data. *Insurance: Mathmatics & Economics*, 49(2):265–284.

Zhang, X. and Schwaab, B. (2017). Tail risk in government bond markets and ECB asset purchases. *Working paper*.

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

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(5):1–16.

Zhou, C. (2010). Existence and consistency of the maximum likelihood estimators for the extreme value index. *Journal of Multivariate Analysis*, 101(4):971–983.