

Article

Statistical Learning of the Worst Regional Smog Extremes with Dynamic Conditional Modeling

Lu Deng ^{1,*}, Mengxin Yu ^{2,†} and Zhengjun Zhang ^{3,†}¹ School of Statistics and Mathematics, Central University of Finance and Economics, Beijing 100081, China² Department of Operations Research and Financial Engineering, Princeton University, Princeton, NJ 08540, USA; mengxiny@princeton.edu³ Department of Statistics, University of Wisconsin, Madison, WI 53706, USA; zjz@stat.wisc.edu

* Correspondence: denglu521@sina.com or ludeng@cufe.edu.cn; Tel.: +86-10-62288522

† These authors contributed equally to this work.

Received: 19 May 2020; Accepted: 19 June 2020; Published: 22 June 2020



Abstract: This paper is concerned with the statistical learning of the extreme smog (PM_{2.5}) dynamics of a vast region in China. Differently from classical extreme value modeling approaches, this paper develops a dynamic model of conditional, exponentiated Weibull distribution modeling and analysis of regional smog extremes, particularly for the worst scenarios observed in each day. To gain higher modeling efficiency, weather factors will be introduced in an enhanced model. The proposed model and the enhanced model are illustrated with temporal/spatial maxima of hourly PM_{2.5} observations each day from smog monitoring stations located in the Beijing–Tianjin–Hebei geographical region between 2014 and 2019. The proposed model performs more precisely on fittings compared with other previous models dealing with maxima with autoregressive parameter dynamics, and provides relatively accurate prediction as well. The findings enhance the understanding of how severe extreme smog scenarios can be and provide useful information for the central/local government to conduct coordinated PM_{2.5} control and treatment. For completeness, probabilistic properties of the proposed model were investigated. Statistical estimation based on the conditional maximum likelihood principle is established. To demonstrate the estimation and inference efficiency of studies, extensive simulations were also implemented.

Keywords: nonlinear time series; conditional modeling; conditional maximum likelihood; hazard pollutants; risk control

1. Introduction

Modeling extreme climatic conditions [1,2], extreme weather [3–8], and rather harmful air pollutants, together with their social, economic, political, and human impacts, is a contemporary research topic. It was concluded by [9] that extreme weather is the new normal. References [10,11] studied the air quality in the USA, and air pollution and health effects in a pyramid figure of effects. Smog as a more serious type of harmful air pollutant has been drawing more and more attention recently. Studies on the masses and chemical compositions, as well as the concentrations, formation, and source of smog have been done [12–19]. Considering the significant impact that some meteorological conditions may have on the PM_{2.5} concentrations [20–22], statistical approaches, as well as physical and chemical ones, were also conducted [23–25]. Except for the studies based on annual averaged PM_{2.5} data [26–28], the cases when extreme smog happens are especially concerned [29–33]. For example, public attitudes and responses to the first two red warnings for air pollution in Beijing in 2015 were examined [34].

China's smog problem stands out in its extremely high frequency, long duration, and high concentration. There are eleven long-lasting rounds of severe smog in 2015, which occurred mainly during the last two months of that year. On 30th November 2015, the $PM_{2.5}$ concentration in Beijing and the south of Hebei exceeded $900 \mu\text{g}/\text{m}^3$ and even reached $976 \mu\text{g}/\text{m}^3$ at Liuli River station, Beijing. The worst smog in 2016 started from 16th December and ended on 21st, covering 17 provinces; i.e., over one seventh of the national territory area. Three fourths of the cities are located in the Beijing–Tianjin–Hebei region and its surrounding regions. The $PM_{2.5}$ concentration in the downtown of Shijiazhuang city even broke $1000 \mu\text{g}/\text{m}^3$. These observations clearly direct us to be more concerned with the statistical learning of extreme smog ($PM_{2.5}$) dynamics of a vast region in China. For example, classical extreme value analysis to hourly $PM_{2.5}$ data from 2014 to 2016 in China have been obtained [35]. It is also worth noting that the smog problem is not a unique phenomenon in China. It occurs elsewhere in the world, especially in developing countries; e.g., Delhi, India [36].

Understanding the extreme features of smog problems in China is very important since severe smog levels are more dangerous than ordinary levels and will do greater harm to China's public health [37–42]. For example, in 2013, 83% of the whole population in China was exposed to the air pollution with $PM_{2.5}$ level exceeding $35 \mu\text{g}/\text{m}^3$, which might cause 1.3 million premature mortalities [43]. The smog also causes a significant economic loss for China [44,45] and even for the whole world [46,47].

Using classical extreme value theory, one could fit the generalized extreme value distribution to extreme observations recorded from each of those hundreds of smog monitoring stations. Recently, Dombry [48] studied properties of the maximum likelihood estimators for the extreme value index within the block maxima framework. Studies [49,50] are the most recent ones dealing with spatial extremes and processes. Study [51] proposed a dynamic modeling approach, the autoregressive conditional Fréchet (AcF) model, for maxima of daily negative log returns of 100 stocks in S&P100. These new models and modeling approaches can certainly be applied to the extreme observations in smog extremes in this study. However, static extreme value models do not offer a dynamic view and may be of less interest to administrators; in addition, the AcF model does not have good performance in climate extremes either. Different from published work in the literature, a new study approach is applied to smog extremes in our work, which intends to integrate a new type of extreme value modeling and dynamic modeling into a dynamic conditional distribution modeling and analysis of regional smog extremes, particularly the worst scenarios observed in each day. The results show a significant improvement compared with using existing extreme value modeling approaches. In addition, weather factors will be introduced in the model to gain higher modeling efficiency. The proposed model and the enhanced model are illustrated with real data of hourly $PM_{2.5}$ observations during 2014–2019 from smog monitoring stations located in the Beijing–Tianjin–Hebei geographical region. In particular, we studied the worst smog dynamic scenarios in the vast region of Beijing–Tianjin–Hebei in China. For joint regional extreme event monitoring and control, knowing first what can be the worst scenario in a day clearly will be very useful for administrators to decide whether or not to make warnings and some necessary control treatments. This paper enhances the understanding of how severe extreme smog scenarios can be and provide useful information for the central/local government to conduct coordinated $PM_{2.5}$ control and treatment. In the literature, researchers have paid much attention to extreme values in environmental studies; e.g., extreme temperature [2], precipitation [2,4–6], snowfall [52], and biomedical physics in brain image analysis [53], among many others. Our new model can certainly find applications in these areas. On the other hand, our model can also be applicable to studying systematic risk in financial systems [54], which is a contemporary research topic. Moreover, in terms of systematic risk study in air-quality control and treatment in regional risk and hazard management, our model can be very useful.

Roadmap

The rest of the paper is organized as follows. Section 2 conducts a preliminary analysis of smog in a vast region of Beijing–Tianjin–Hebei. In Section 3, we introduce our dynamic model by specifying a latent independent standard random process of standard Weibull random variables. Probabilistic properties of stationarity and ergodicity of the proposed model are investigated. Statistical estimation based on the conditional maximum likelihood principle is established. To demonstrate the estimation and inference efficiency of studies, extensive simulations are implemented in Section 4. Section 5 contains advanced modeling of smog extremes together with weather data. Section 6 offers conclusions of the paper regarding our applied study of smog extremes. Technical arguments are deferred to Appendix A.

2. Preliminary Analysis of Smog in the Vast Region of Beijing–Tianjin–Hebei

2.1. Which Time Scale of $PM_{2.5}$ Data Is to Be Analyzed?

When considering extreme smog as an extremely harmful air pollutant which may last a few hours in a day or last a couple of days, it becomes rather meaningful to study the hourly characteristics of radical smog movements, instead of daily traits. One can consider an analysis of daily data as well. We will not consider this time scale in this applied project.

To adequately demonstrate the variation of smog in different years, data from 2014 to 2019 are used to fit our proposed model. The first three months' data of 2020 are used to check the predictability of our model. All the data are from the China National Environmental Monitoring Center. Due to technological testing, transition problems, hardware failure, possible delayed updates, etc., some of the data are missing and taken as non-observable smaller values to the observed daily maxima in this study.

2.2. The Geographical Region to Be Focused on

Figure 1 demonstrates 90% quantile maps for January and December of 2014 and 2019 among the monitoring stations in mainland China.

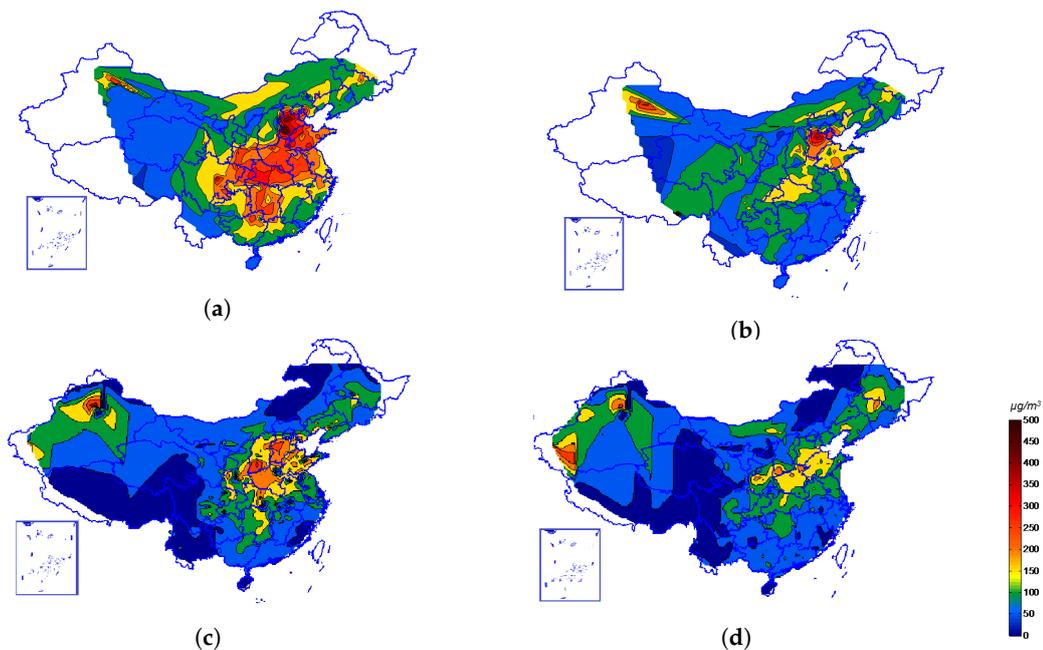


Figure 1. Contour maps of 90% quantiles ($\mu\text{g}/\text{m}^3$). (a) The contour maps of January 2014, (b) the one for December 2014, and (c,d) those of January and December 2019 respectively (the blank areas are due to the absence of stations).

As shown in Figure 1a, in January 2014, most of the monitored areas spanning the north, east, and middle of China (especially those red areas) had quantiles exceeding $150.4 \mu\text{g}/\text{m}^3$, which is the starting point of the very unhealthy level of air quality according to the US standard [55]. There are five intervals of $\text{PM}_{2.5}$ levels according to the US (EPA) standard: (1) $35 \mu\text{g}/\text{m}^3$ is the highest $\text{PM}_{2.5}$ level for the good and moderate category; (2) the range from 35 to $55.4 \mu\text{g}/\text{m}^3$ is unhealthy for sensitive groups; (3) the range from 55.5 to $150.4 \mu\text{g}/\text{m}^3$ is unhealthy; (4) the range above $150.4 \mu\text{g}/\text{m}^3$ is widely viewed as very unhealthy; and (5) it is hazardous when the smog level is above $250.5 \mu\text{g}/\text{m}^3$. We see then that south Hebei, Beijing, and Tianjin showed the highest smog levels with the 90% quantile values being close to an outstanding $500 \mu\text{g}/\text{m}^3$, which is the maximum value given in the US standard. A similar conclusion can be drawn from Figure 1b is that in December 2014, south Hebei, Beijing, and Tianjin were still the worst air quality areas, although the quantiles of the smog values in most areas of China were much lower than their counterpart values in January. In January and December 2019 (as shown in Figure 1c,d respectively), the thresholds among all of China decreased significantly; that shows great improvement in extreme smog problems; south Hebei, Beijing, and Tianjin are still some of the key regions with the severest smog conditions (relatively higher threshold). Monthly 90% quantiles of 2015 and 2018 confirm this conclusion as well.

This study focuses on the smog problem in the Beijing–Tianjin–Hebei region; i.e., not all of mainland China, whose smog issues may be too diversified to draw consensus conclusions about. The importance of focusing the study on the smog problem in the Beijing–Tianjin–Hebei region is due to the fact that the areas are not only geographically connected but also economically significant in their immense contribution to China's GDP (nearly 9% in 2019 with 8% of population). With the implementation of the collaborative development of the Beijing–Tianjin–Hebei region as a national strategy, the coordination among these three areas has been becoming more and more intimate. It has been taken as an indivisible whole, especially in the issues of smog prevention and control.

The smog problem in this region has been recognized on account of health concerns as early as in January 2013, when long-lasting severe smog prevailed and caused a significant increase in the number of patients with respiratory tract infections and allergic symptoms in hospitals and clinics in Beijing, Tianjin, and Shijiazhuang (the capital city of Hebei province). Research also indicated that bad air quality in northern China causes 5.5 years reductions of people's life expectancy [56].

In the Beijing–Tianjin–Hebei region, besides two metropolitan cities, Beijing (the capital of China) and Tianjin (a major port city in northeastern China), there are 11 cities in Hebei province. Among these 11 cities, Zhangjiakou and Chengde are located in north Hebei; Qinghuangdao, Tangshan, and Cangzhou are in east Hebei; Baoding and Langfang are located in the middle of Hebei; Shijiazhuang, Hengshui, Xingtai, and Handan combine south Hebei. Their locations along with Beijing and Tianjin are shown in Figure 2. The number of smog monitoring stations in the region nearly remained the same during 2014–2019 (after deleting stations still under test, the numbers of valid stations are around 80 for all years). Worth noticing is that the station type is the key parameter to explain diurnal variability of atmospheric pollutants due to sources proximity. In this paper, however, since only the daily maxima of $\text{PM}_{2.5}$ across the whole region rather than a spatial model is considered, the station type and geographical information which describe the spatial relations are not used in the study.



Figure 2. Locations of cities in Beijing–Tianjin–Hebei region. The numbers of national monitoring stations in 2016 in each city are listed in the brackets.

2.3. Why Model the Extremes Rather than the Average Levels?

This paper focuses on the extreme smog instead of average levels for two reasons. Firstly, the extreme smog rather than the ordinary smog affects the public health and causes more loss to the welfare of the whole society [36,38,41]. Secondly, when investigating the severity of smog, the results can be very different from the use of the annual averages or the extreme values of PM_{2.5} data. Taking stations of Zhangjiakou and Chengde during 2014–2016 as an example, their annual average PM_{2.5} levels rank as the best 30% among the whole country, showing relatively good air quality on average. However, as far as the means of extreme values (here extreme values refers to those PM_{2.5} levels above the annual 90% quantile) are concerned, those stations belong to the worst third of the whole country. This difference in ranks gives two implications. On the one hand, the poorer grades using more extreme values than those using the annual averages show that the air quality of Zhangjiakou and Chengde may not be as good as one has thought; i.e., a city that performs well on average does not guarantee an absence of extreme smog. On the other hand, for Zhangjiakou and Chengde, while their extreme PM_{2.5} levels were relatively lower most of the time, there were still a few occurrences of very high PM_{2.5} levels to which considerable attention should be paid. The histograms of hourly PM_{2.5} extreme values for those stations also confirm such arguments. One of those histograms at one representative station in Zhangjiakou in 2014 is given in Figure 3 as example.

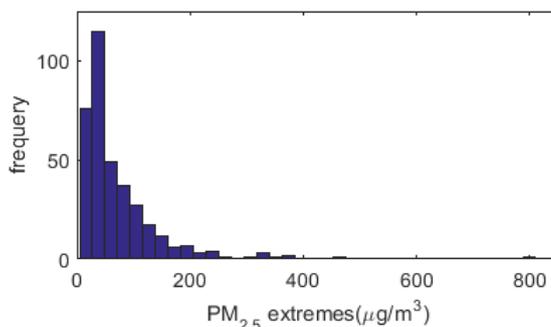


Figure 3. The histogram of hourly PM_{2.5} extremes at one representative station of Zhangjiakou in 2014.

For studies on the average $PM_{2.5}$ level rather than the extreme $PM_{2.5}$ value, we refer readers to [26–28].

2.4. The Study Approach and the Inclusion of Meteorological Variables

With the established arguments in the prior section, this applied study focuses on the extreme values rather than the average values of $PM_{2.5}$ data. Ideally, one should build a model to describe extreme smog level dependencies among monitoring stations in the region. Then, all analyses and inferences, even policy recommendations, are to be based on the constructed model. To achieve this goal, a model builder has to consider many location/spatial parameters and extreme spatial dependencies, which can be very complicated and hardly implementable in a multivariate/spatial extreme value and time series context. We note that in the literature, workable and meaningful air quality models, including a comprehensive air quality model with extensions (CAMx), the community multiscale air quality modeling system (CMAQ), etc., have been developed for various applications and for developing public policies. In this study, we adopt an alternative approach in modeling regional extremes in a time series modeling framework. This paper tries to model the maximum $PM_{2.5}$ values of 24 h a day among all stations in the Beijing–Tianjin–Hebei region. Here, the extreme value represents the highest daily $PM_{2.5}$ concentration of the area as a whole. It is particularly useful when planning smog prevention and control measurements, since such measurements as well as the smog pre-warning systems are integrated and unified in the Beijing–Tianjin–Hebei region. The pre-warnings should be activated and corresponding measurements should also be taken so long as the predicted $PM_{2.5}$ level at one single station inside the region exceeds a specific breakpoint wherever the station is. In this circumstance, only the maximum value of the whole part rather than the particular concentrations of all individual stations is concerned.

The line graphs of the regional daily maximum (based on hourly observations) $PM_{2.5}$ levels from 2014 to 2019 are plotted in Figure 4a. It can be seen that extreme smog levels were much worse in 2014 since its extreme values were mostly much higher compared to the same period of the following five years. It is further demonstrated in Table 1 that the extreme values in 2014 have a much higher sample mean, median, maximum and standard deviation; they improved significantly in recent years with these statistics decreasing in general. More importantly, the maximum $PM_{2.5}$ levels over the years are persistent with an exceptional high value in 2014. Meanwhile, the skewness and kurtosis are persistently larger than those from a Gaussian distribution. The extreme values also show significant seasonality for all six years in that their sample means and standard deviations are much higher in the first and fourth seasons than the other two seasons, as shown in Figure 4b. We shall consider these dynamic characteristics in our model building.

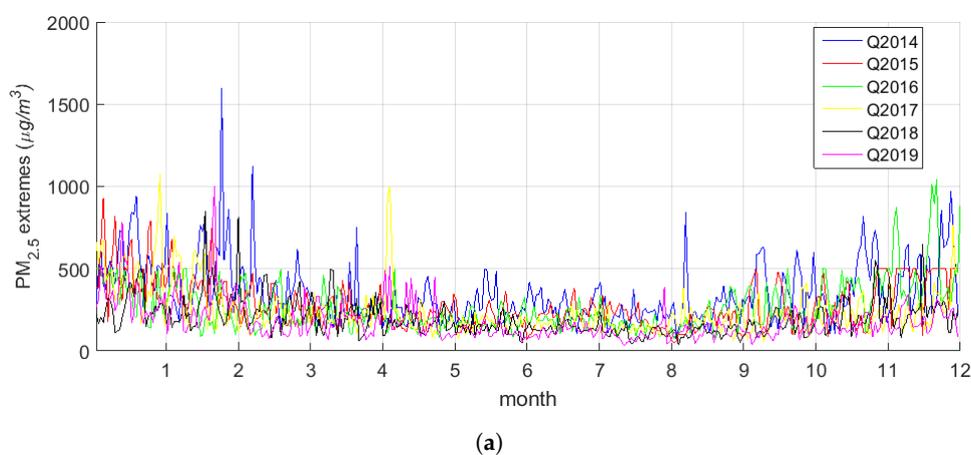
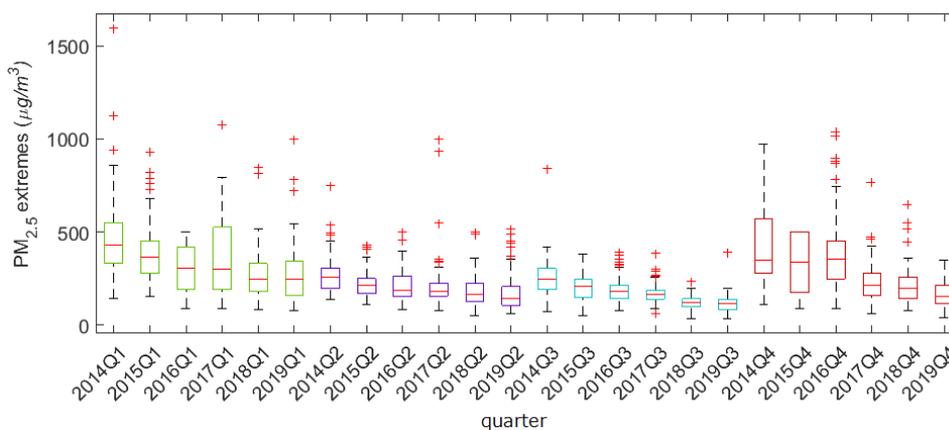


Figure 4. Cont.



(b)

Figure 4. (a) The line graphs of the regional daily extreme values ($\mu\text{g}/\text{m}^3$) from 2014 to 2019 (Q2014 represents those extreme values in 2014 and so on; the same thereafter). (b) The boxplots of seasonal extreme values ($\mu\text{g}/\text{m}^3$) from 2014 to 2019. (2014Q1 represents the first season of 2014 and so on.)

Table 1. Descriptive statistics of extreme values ($\mu\text{g}/\text{m}^3$) from 2014 to 2019.

	Q2014	Q2015	Q2016	Q2017	Q2018	Q2019
Mean	351	281	267	239	195	180
Median	302	241	228	192	163	144
Maximum	1597	929	1040	1076	850	1000
Minimum	74	49	78	58	36	33
Std. Dev.	186	144	149	146	109	117
Skewness	1.8	1.1	1.8	2.3	2.0	2.4
Kurtosis	9.1	4.5	7.9	10.0	9.7	12.6

In the literature, extensive research shows that climate factors may cause a significant impact on the $\text{PM}_{2.5}$ levels [20,21] at one location. These meteorological conditions include temperature, humidity (the humidity mentioned in this paper is relative humidity (%)), wind speed, wind direction, etc.; see [23,24]. Taking Beijing as an example, Figure 5 is the daily maximum $\text{PM}_{2.5}$ level of Beijing in 2018 that is computed using the hourly $\text{PM}_{2.5}$ levels among all monitoring stations in Beijing. The selected meteorological conditions are Beijing’s daily maximum and minimum temperature, daily maximum and minimum humidity, daily maximum wind levels, and its related wind direction in 2018, as shown in Figure 6. All the data are from the National Meteorological Information Center.

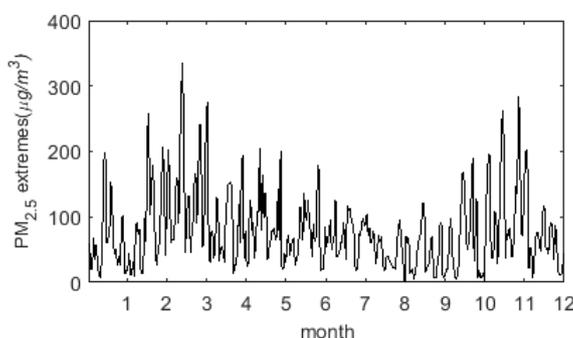


Figure 5. Daily maximum $\text{PM}_{2.5}$ level ($\mu\text{g}/\text{m}^3$) of Beijing in 2018.

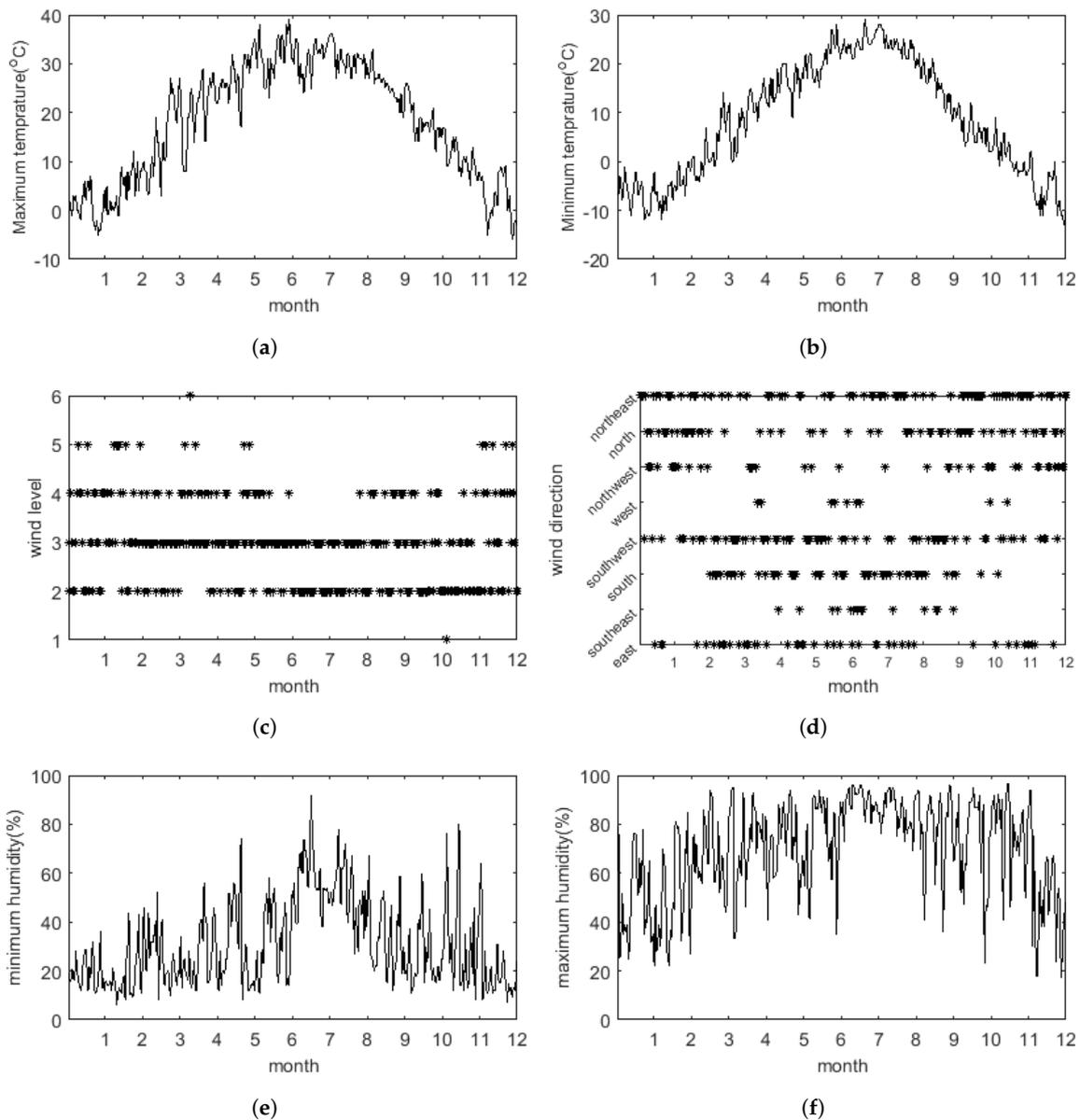


Figure 6. The meteorological condition of Beijing in 2018. (a) The daily maximum temperature (degree centigrade), (b) the daily minimum temperature (degree centigrade), (c) the maximum wind level (force 1 to force 6), (d) the wind direction in the presence of the maximum wind level. (e) The daily minimum humidity (%); (f) the daily maximum humidity (%).

The annual maximum PM_{2.5} level in Beijing is around 350 µg/m³ in 2018, which is much higher than the hazardous breakpoints (250 µg/m³) according to the US standard. Although the annual maximum of the PM_{2.5} values in Beijing is lower than that of the regional extremes given in Table 1 (850 µg/m³), around one tenth of the daily regional extremes are observed from stations in Beijing. Actually, the locations of the highest daily regional extremes are well spread among cities in the Beijing–Tianjin–Hebei region, which exactly demonstrates that modeling the regional extremes instead of local PM_{2.5} dynamics is very important. Besides, the smog in the first and last seasons is much more sever than that in the other two seasons; meanwhile, the daily minimum and maximum temperatures in the first and last seasons are much lower, showing potential negative correlations between the PM_{2.5} level and temperature. As for wind and humidity, they affect the concentration of smog in different ways: it is known that humidity affects the atmospheric chemistry and the formation of secondary

pollutants such as particulate matter and ozone. However, wind influences the concentrations of the trace gases, which react at rates determined by their concentrations. The fourth level wind coming from the north or northeast most often occurs in the first season, which tends to decrease the PM_{2.5} level. However, the smog is still severe when the wind becomes weak. In this season, the differences between the daily maximum and minimum humidity are relatively larger (mainly caused by the lower minimum humidity), which is an adverse diffusion condition for the smog [57]. In the second and third seasons, winds with the third level coming from the south and south-west dominate. In these two seasons, the smog problem is much less severe and both the maximum and minimum humidity are relatively high (resulting in lower humidity difference). In the last season, winds seem to be weaker with the second level coming from the northeast, and both the daily maximum and minimum humidity are barely low (with a relatively smaller minimum humidity), which could be reasons leading to the severe smog. Generally speaking, it is observed that the lower the wind level and the higher the humidity difference (or the lower the minimum humidity), the higher the daily PM_{2.5} extremes. These observations coincide with the general results from other studies [57,58]. More details about the specification of covariates (climate factors) when modeling the regional daily maximum PM_{2.5} levels will be addressed in Section 5.2. The data source is also the National Meteorological Information Center.

3. Model Specification

Suppose X_{tij} is the PM_{2.5} level on day t at time i (hour) of the j th station in a group containing m stations. Define Q_t as

$$Q_t = \max_{1 \leq i \leq 24, j=1, \dots, m} X_{tij}. \quad (1)$$

An illustration of daily regional smog extremes' time series $\{Q_t\}$ is given in Figure 4.

In the extreme value theory, under suitable conditions ([59]), the normalized maximum of a sequence (a group, a block, or an array) of random variables is distributed as a generalized extreme value (GEV) distribution in its limit when the size of the sequence tends to infinity. In view of the definition in Q_t , random variables X_{tij} 's are potentially showing time dependence and spatial dependence, and the distributions can be different. As a result, Q_t can be hardly distributed as a GEV random variable, and the dependence between Q_t and Q_{t-1} cannot be modeled using the models (e.g., [51,60]) developed for GEV distributed random variables.

3.1. The Proposed General Model

Inspired by the AcF model (autoregressive conditional Fréchet model) proposed by [51], we propose the following model for smog extremes modeling:

$$Q_t = \mu_t + \sigma_t Y_t^{1/\alpha_t}, \quad (2)$$

where μ_t is the lowest level of PM_{2.5} in the region on day t , $\sigma_t > 0$ is the scale parameter, $\alpha_t > 0$ is the shape parameter, and $\{Y_t\}_{t \geq 0}$ is a sequence of i.i.d. exponentiated Weibull (unit exponential) random variables.

One notices that the standardized Q_t , $(Q_t - \mu_t)/\sigma_t$ follows an exponentiated Weibull distribution with the shape parameter α_t . By introducing the additional parameter α_t , the distributions of Q_t possess great variability and flexibility. In the literature, the exponentiated Weibull distribution has been applied to model climate extremes; e.g., in flood modeling in ([61]). Unlike the AcF model, wherein unit Fréchet distributions are assumed to capture the heavy-tailedness of financial data, Weibull-distributed $\{Y_t\}$ s are considered here to fit the characteristics of smog data which are not necessarily being heavy-tailed. It has been observed in [35] that the station-wise extremes in the Beijing–Tianjin–Hebei region belong to the maximum domain of attraction of Weibull type, which leads

to a reasonable assumption of Y_t being standard Weibull distributed. To be more rigorous, we also fit the ACF model in Section 5 for comparison.

The next stage is to build a time series model for the $\{Q_t\}$ series. Following the arguments above and those in [51], to reduce the model complexity, we treat μ_t as constant μ and interpret μ as the lowest level of PM_{2.5}fpr all the time across the Beijing–Tianjin–Hebei region, and treat σ_t and α_t as dynamic.

Noticing that the smog levels can be affected by the past smog levels and weather conditions, we assume the following dynamic equations for the time varying parameters:

$$\log \sigma_t = \beta_0 + \beta_1 \log \sigma_{t-1} + \eta_\sigma(Q_{t-1}, Tp_{t-1}, Hu_{t-1}, Wds_{t-1}, Wdd_{t-1}), \tag{3}$$

$$\log \alpha_t = \gamma_0 + \gamma_1 \log \alpha_{t-1} + \eta_\alpha(Q_{t-1}, Tp_{t-1}, Hu_{t-1}, Wds_{t-1}, Wdd_{t-1}), \tag{4}$$

in which Tp_t , Hu_t , Wds_t , and Wdd_t denote the maximum temperature, minimum humidity, mode of wind speed, and mode of wind direction respectively. These four weather condition variables have been commonly discussed in the smog study literature; e.g., [62,63], and the references therein. Other weather factors can also be considered. However, adding too many variables increases the model complexity and weakens the effects of important variables. For this reason, we only focus on the dynamics of variables given in (3) and (4). We call model (2)–(4) the dynamic conditional Weibull (DCW) model. In this paper, we consider functions $\eta_\sigma(\cdot)$ and $\eta_\alpha(\cdot)$ defined in (3) and (4) as exponential functions which are widely adopted in the literature, e.g., [51], due to their flexibilities in functional properties of boundedness, differentiability, and monotonicity. It is worth noting that the functions of $\eta_\sigma(\cdot)$ and $\eta_\alpha(\cdot)$ given in (3) and (4) can also take other forms, as long as they meet some conditions that guarantee the dynamic processes being stationary and ergodic. For clarity and simplicity in the proofs of stationarity and ergodicity, we first omit terms related to weather factors and express model (2)–(4) as:

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

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

$$Q_t = \mu + \sigma_t Y_t^{1/\alpha_t} \tag{7}$$

Moreover, we will discuss the model related to weather factors in Sections 4 and 5.

In the following Theorem 1, we prove that the process (σ_t, α_t) generated from (5)–(7) is stationary and ergodic.

Theorem 1. (Stationarity and ergodicity.) *If we have $0 \leq |\beta_1| \neq |\gamma_1| < 1$, $\beta_0, \gamma_0, \beta_2, \gamma_2, \mu \in R$ and $\beta_3, \gamma_3 \geq 0$ the process $\{\sigma_t, \alpha_t\}$ defined in (5)–(7) is stationary and geometrically ergodic.*

In Section 3.2, we study our parameter estimation procedures and characterize asymptotic behaviors of the estimators.

3.2. Parameter Estimation and Asymptotic Properties

We begin by introducing some notation. We denote $\Theta_s = \{\theta = (\mu, \beta_0, \beta_1, \beta_2, \beta_3, \gamma_0, \gamma_1, \gamma_2, \gamma_3) \mid \beta_0, \gamma_0, \mu, \beta_2, \gamma_2 \in R, -1 < \beta_1, \gamma_1 < 1, \beta_3, \gamma_3 \geq 0\}$ as the parameter space for the estimation problem in (5)–(7) and set the true parameter as $\theta_0 = (\mu_0, \beta_0^0, \beta_1^0, \beta_2^0, \beta_3^0, \gamma_0^0, \gamma_1^0, \gamma_2^0, \gamma_3^0)$. After letting $(\tilde{\sigma}_1, \tilde{\alpha}_1)$ be an arbitrary initial value, we then denote $(\tilde{\sigma}_t(\theta), \tilde{\alpha}_t(\theta))$ as the t -th iterate generated from model with initializer $(\tilde{\sigma}_1, \tilde{\alpha}_1)$ and an arbitrary parameter θ in Θ_s . In addition, (σ_t^0, α_t^0) is denoted as the t -th iterate generated from the model with true (σ_1, α_1) and θ_0 . Moreover, we also denote $(\sigma_t(\theta), \alpha_t(\theta))$ as the

values generated from the true initializer (σ_1, α_1) and an arbitrary θ in Θ_s . With the known $\{\mu, \tilde{\sigma}_t(\theta), \tilde{\alpha}_t(\theta)\}$, the conditional p.d.f. of Q_t is given by

$$f_t(\theta) = f_t(Q_t|\mu, \tilde{\sigma}_t, \tilde{\alpha}_t) = \frac{\tilde{\alpha}_t}{\tilde{\sigma}_t} \left(\frac{Q_t - \mu}{\tilde{\sigma}_t}\right)^{\tilde{\alpha}_t - 1} \exp\left(-\left(\frac{Q_t - \mu}{\tilde{\sigma}_t}\right)^{\tilde{\alpha}_t}\right) \tag{8}$$

for any fixed t . By leveraging the conditional independence property of $\{Q_t\}_{t \geq 0}$, we further write the log-likelihood function with respect to parameter θ as

$$\tilde{L}_n(\theta) = \frac{1}{n} \sum_{t=1}^n \tilde{\ell}_t(\theta) = \frac{1}{n} \sum_{t=1}^n \left[\log \tilde{\alpha}_t - \tilde{\alpha}_t \log \tilde{\sigma}_t + (\tilde{\alpha}_t - 1) \log(Q_t - \mu) - \left(\frac{Q_t - \mu}{\tilde{\sigma}_t}\right)^{\tilde{\alpha}_t} \right], \tag{9}$$

where $\tilde{\ell}_t(\theta) = \log(f_t(\theta))$.

Next, we impose two assumptions of the model that we are investigating in.

Assumption 1. Assume the parameter space Θ is a compact set of Θ_s . Suppose the observations $\{Q_t\}_{t=1}^n$ are generated from a stationary and ergodic DCW process with true parameter θ_0 being an interior point of Θ .

Due to the compactness of Θ , there exists a uniform upper and lower bound of the sequence $(\sigma_t(\theta), \alpha_t(\theta))$ and $(\tilde{\sigma}_t(\theta), \tilde{\alpha}_t(\theta))$ with $\theta \in \Theta$, which are denoted as (σ_U, α_U) and (σ_L, α_L) respectively. We next make an assumption on the lower bound of our sequence α_t .

Assumption 2. The uniform lower bound α_L is larger than 2.

It is noted that [64] studied the consistency and asymptotic normality of irregular MLEs from a group of distributions, including the three-parameter Weibull distribution, using i.i.d. observations. We extend the existing results to a dynamic model with dependent observations. In addition, Assumption 2 coincides with the results given by [64] that the classical asymptotic properties holds only if $\alpha > 2$ under their static settings. We next theoretically characterize the consistency property of the local maximizer of likelihood function $\tilde{L}_n(\theta)$ in the following Theorem 2.

Theorem 2. (Consistency) Under Assumptions 1 and 2, there exists a sequence $\{\hat{\theta}_n\}_{n \geq 1}$ that maximizes $\{\tilde{L}_n(\theta)\}_{n \geq 1}$ and satisfies $\|\hat{\theta}_n - \theta_0\| \leq \tau_n$ with $\tau_n = O_p(n^{-r})$, and $1/\alpha^L < r < 1/2$.

Theorem 2 shows that there exists a sequence $\{\hat{\theta}_n\}_{n \geq 1}$ which contains not only consistent estimators to θ_0 but local maximizers of $\{\tilde{L}_n(\theta)\}_{n \geq 1}$ as well. Next, we derive the asymptotic distributions of our estimators $\hat{\theta}_n$ in the following Theorem 3.

Theorem 3. (Asymptotic normality) Under the same assumptions in Theorem 2, we have $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, M_0^{-1})$, where $\hat{\theta}_n$ is given in Theorem 2, and M_0 is the Fisher information matrix with the value estimated at θ_0 . Furthermore, the variance of the plugged-in estimated score functions $\{\frac{\partial}{\partial \theta} l_t(\hat{\theta}_n) = \log \tilde{\alpha}_t - \tilde{\alpha}_t \log \tilde{\sigma}_t + (\tilde{\alpha}_t - 1) \log(Q_t - \mu) - \left(\frac{Q_t - \mu}{\tilde{\sigma}_t}\right)^{\tilde{\alpha}_t}\}_{t=1}^n$ is a consistent estimator of M_0 .

Although the existence of the $\hat{\theta}_n$ and their asymptotic distributions are shown in Theorem 2 and Theorem 3 respectively, the uniqueness of MLE remains open. Proposition 1 provides a segmentary answer to the uniqueness of MLE.

Proposition 1. (Asymptotic uniqueness) Define the set $V_n = \{\theta \in \Theta | \mu \leq \mu_0 + \epsilon_n\}$ with Θ given in Theorem 2 and $\epsilon_n = O_p(n^{-\alpha})$, $1/\alpha^L < \alpha < 1/2$. Under the the same assumptions in Theorem 2, there exists a sequence of $\hat{\theta}_n = \arg \max_{\theta \in V_n} \tilde{L}_n(\theta)$ such that we have $\|\hat{\theta}_n - \theta_0\| \leq \tau_n$, and $\mathbb{P}(\hat{\theta}_n \text{ is the unique global maximizer of } \tilde{L}_n(\theta) \text{ over } V_n) \rightarrow 1$, where $\tau_n = O_p(n^{-r})$, $1/\alpha^L < \alpha < r < 1/2$.

Note that, given observations $\{Q_t\}_{t \geq 1}$, the parameter space of θ is defined as $\Theta_n = \{\theta \in \Theta \mid \mu < Q_{n,1}\}$ after ranking $\{Q_t\}_{t \geq 0}$. One is able to see that $V_n \subseteq \Theta_n$ since we have $Q_{n,1} - \mu_0 \geq O_p(n^{-1/\alpha_L})$. In addition, Proposition 1 states that with the probability tending to 1, $\hat{\theta}_n = \arg \max_{\theta \in V_n} \tilde{L}_n(\theta)$ is a unique consistent estimator of θ_0 over V_n . The formal proof of this Proposition 1 can be found in Appendix A.5.

In Section 4, we use simulation examples to illustrate the numerical evidence of the established theory, and then in Section 5, we study smog dynamics in the Beijing–Tianjin–Hebei region.

4. Numerical Studies Using Simulations

In this section, we present two simulation examples. Example 1 does not involve weather variables. Example 2 contains weather variables. In each simulation, we generate the process using parameter values from real applications and i.i.d. standard Weibull random variates with lengths of 2000 and 5000 respectively. Then we fit the parameters by conditional MLE, starting with the paired values which we used to get the true MLE. The codes for simulation studies in this Section 4 and for model estimation and prediction in Section 5, which were developed in R software, have been put in GitHub for free assessments. (URL is <https://github.com/MaxineYu/DynamicConditionalWeibullModel-DCWcodes>.)

Example 1. *Simulations without weather factors: In Table 2, we list all parameter values (taken from the estimated values for a real application based on model (10)–(12) using data from 2014–2016, and report the estimated mean values and standard deviations from 500 repetitions of time series with lengths of 2000 (scenario SC1) and 5000 (scenario SC2) respectively.*

The following Table 2 presents our estimation results for simulations without considering weather factors.

Table 2. Performance of MLE with the sample sizes of 2000 (SC1) and 5000 (SC2). Mean and SD are samples’ means and standard derivations of 500 estimated corresponding parameter values respectively. Ratio denotes the ratio between the standard errors with $n = 2000$ and standard errors with $n = 5000$.

Parameter	True Value	Mean (SC1)	SD (SC1)	Mean (SC2)	SD (SC2)	Ratio
μ	4.677×10^1	4.781×10^1	3.022	4.719×10^1	1.961	0.649
β_0	5.387	5.394	1.837×10^{-1}	5.392	1.266×10^{-1}	0.689
β_1	1.912×10^{-1}	1.890×10^{-1}	2.595×10^{-2}	1.900×10^{-1}	1.795×10^{-2}	0.692
β_2	−2.219	−2.230	7.863×10^{-2}	−2.224	4.951×10^{-2}	0.630
β_3	3.439×10^{-3}	3.479×10^{-3}	2.668×10^{-4}	3.455×10^{-3}	1.687×10^{-4}	0.632
γ_0	2.398	2.387	6.065×10^{-2}	2.394	3.825×10^{-2}	0.631

One can see that the mean values of 500 estimates are very close to their corresponding true values. The corresponding standard deviations also illustrate the significance of our estimation.

Table 2 also presents the ratios between the standard errors with $n = 2000$ and those corresponding ones with $n = 5000$. It is noticed that the ratios are close to $0.632 = \sqrt{2000/5000}$, which is consistent with our theoretical justification in Theorem 3 with $\sqrt{n}(\hat{\theta}_n - \theta_0) \rightarrow_d N(0, M_0^{-1})$.

Next, we study simulations involving weather factors.

Example 2. *Simulations with weather factors. In Table 3, we list all parameter values (taken from the estimated values for a real application based on model (16)–(18) using data from 2014–2016), and report the estimated mean values and standard deviations from 500 repetitions of time series with lengths of 2000 and 5000 respectively.*

We note that in the simulations, we used the observed weather values of the maximum temperature, minimum humidity, mode of wind-level, and modes of wind-direction (see Section 5 for more details) of a day (24 h).

Table 3. Performance of MLE with the sample sizes of 2000 (SC1) and 5000 (SC2). Mean and SD are samples means and standard derivations of 500 estimated corresponding parameter values respectively.

Parameter	True Value	Mean (SC1)	SD (SC1)	Mean (SC2)	SD (SC2)
μ	4.624×10^1	4.695×10^1	3.499	4.642×10^1	2.237
β_0	6.010	6.021	2.337×10^{-1}	6.016	1.443×10^{-1}
β_1	1.257×10^{-1}	1.235×10^{-1}	3.090×10^{-2}	1.250×10^{-1}	2.004×10^{-2}
β_2	-1.437	-1.434	1.133×10^{-1}	-1.440	6.263×10^{-2}
β_3	2.237×10^{-3}	2.256×10^{-3}	1.998×10^{-4}	2.240×10^{-3}	1.111×10^{-4}
β_4	9.283×10^{-3}	9.442×10^{-2}	1.011×10^{-3}	9.291×10^{-3}	6.166×10^{-4}
β_5	2.622×10^{-3}	2.649×10^{-3}	5.506×10^{-4}	2.627×10^{-3}	3.143×10^{-4}
β_6	1.021×10^{-1}	1.035×10^{-1}	1.730×10^{-2}	1.022×10^{-1}	9.577×10^{-3}
$c_{\sigma,1}$	7.165×10^{-3}	6.783×10^{-3}	3.045×10^{-2}	7.566×10^{-3}	1.956×10^{-2}
$c_{\sigma,2}$	-1.488×10^{-1}	-1.511×10^{-1}	3.131×10^{-2}	-1.490×10^{-1}	1.867×10^{-1}
$c_{\sigma,3}$	-1.042×10^{-1}	-1.066×10^{-1}	2.770×10^{-1}	-1.045×10^{-1}	1.701×10^{-1}
$c_{\sigma,4}$	-8.835×10^{-2}	-8.782×10^{-2}	3.093×10^{-3}	-8.797×10^{-2}	1.713×10^{-2}
$c_{\sigma,5}$	-4.190×10^{-2}	-4.270×10^{-2}	2.890×10^{-2}	-4.253×10^{-2}	1.759×10^{-2}
$c_{\sigma,6}$	1.992×10^{-2}	2.320×10^{-2}	3.699×10^{-2}	1.949×10^{-2}	2.289×10^{-2}
$c_{\sigma,7}$	4.577×10^{-2}	4.703×10^{-2}	2.594×10^{-2}	4.575×10^{-2}	1.520×10^{-2}
γ_0	2.572	2.574	6.617×10^{-2}	2.575	4.271×10^{-2}

Comparing the results related to parameter estimation and standard errors in Tables 2 and 3, we can see that the common parameters like $\theta = (\mu, \beta_0, \beta_1, \beta_2, \beta_3, \gamma_0)$ in Examples 1 and 2 have comparable estimation accuracy. Moreover, the estimated parameter values associated with the weather variables are close to their corresponding true values, while the Monte Carlo standard deviations of the parameters associated with modes of wind-directions are relatively large with $n = 2000$, which suggests low estimation accuracy, though they are improved when $n = 5000$. This phenomenon is understandable, as the corresponding observations are zero inflated.

In summary, two simulation examples confirm that our estimation procedure (the conditional MLE) works for our proposed model parameter estimation, which provides an empirical support in our theoretical results and real data applications. The computational times of Examples 1 and 2 along with the computing environment are listed in Appendix B.

5. Real Data Inferences

5.1. Inference without Weather Factors

With the established model and notation in Section 3, we first fit model (5)–(7) using the data from 2014–2019. Given the fitted parameters in the above model, we generate and plot the fitted σ_t and α_t dynamics in Figure 7. In this process, for every $t \geq 1$, we apply real values of smog data Q_{t-1} in Equations (5) and (6) to generate fitted (σ_t, α_t) . In addition, initial values of (σ_1, α_1) can be arbitrarily chosen (greater than zero), as their effects on the generated sequences are negligible by our theoretical justification given in Appendix A.

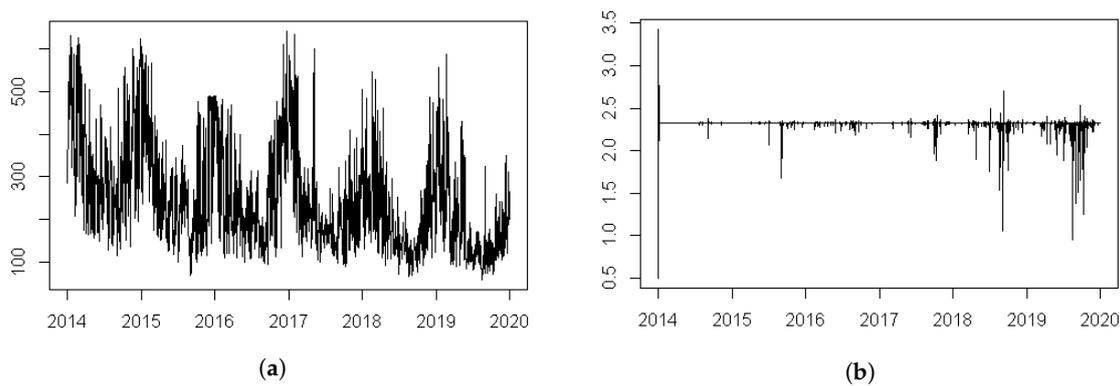


Figure 7. (a) The graph of fitted σ_t and (b) the graph of fitted α_t resulted from the DCW model.

Figure 7a depicts dynamics of σ_t which represents the scale of the maximum extreme smog. It is shown that σ_t enjoys smaller values during the second and third season and enhances greatly during the first and fourth seasons. This seasonality matches one of the extreme smog values, which means that when the extreme smog goes severe, the maximum value among the whole region fluctuates more. The observed phenomenon is similar to the volatility sequence in GARCH (generalized autoregressive conditional heteroscedasticity) model ([65]) which reflects the local volatilities. The σ_t in DCW plays the same role as the volatility in GARCH model. From Figure 7b, one can see that values of α_t converge to a constant very quickly once a fluctuation occurs. Moreover, considering the fact that the parameter γ_2 is neither significant nor stable in Monte Carlo simulations, the parameter γ_2 is set as zero here. As a result, the equation of α_t , which only contains γ_0 and γ_1 this time, also converges to a constant. Based on the above arguments, we constrain $\gamma_1 = \gamma_2 = 0$ in Equation (6), which simplified to the following model (11), and our final model without the weather factor becomes (10)–(12):

$$\log \sigma_t = \beta_0 + \beta_1 \log(\sigma_{t-1}) + \beta_2 \exp(-\beta_3 Q_{t-1}), \tag{10}$$

$$\alpha_t = \gamma_0, \tag{11}$$

$$Q_t = \mu + \sigma_t Y_t^{1/\alpha_t}. \tag{12}$$

The estimated results are presented in Table 4.

Table 4. Estimated parameters for model (10)–(12).

Parameter	Fitted Value	SD
μ	3.218×10^1	3.457×10^{-1}
β_0	4.885	1.081×10^{-1}
β_1	2.461×10^{-1}	1.682×10^{-2}
β_2	-2.195	4.443×10^{-2}
β_3	4.613×10^{-3}	1.549×10^{-4}
γ_0	2.320	1.814×10^{-2}

It is worth noting that the simplified model (10)–(12) and the model (5)–(7) have very closed maximum likelihood values (-5.84223 and -5.84348 respectively). The likelihood ratio statistic testing the two constraints ($\gamma_1 = \gamma_2 = 0$) equals 0.0025, which is much smaller than the critical value $\chi^2(2) = 5.991$. This confirms the validity of the two constraints; i.e., the simplified model should be used. Using the estimated parameters given in Table 4, we generate a sequence of fitted Q_t (notated as \tilde{Q}_t thereafter) based on Equation (12), in which σ_t and α_t are obtained by Equations (10) and (11) with the same procedure we used to plot Figure 7. QQ-plot(quantile–quantile plot) of the real data Q_t against the generated one \tilde{Q}_t is presented in Figure 8a and the line graphs of these two series are given in Figure 8b.

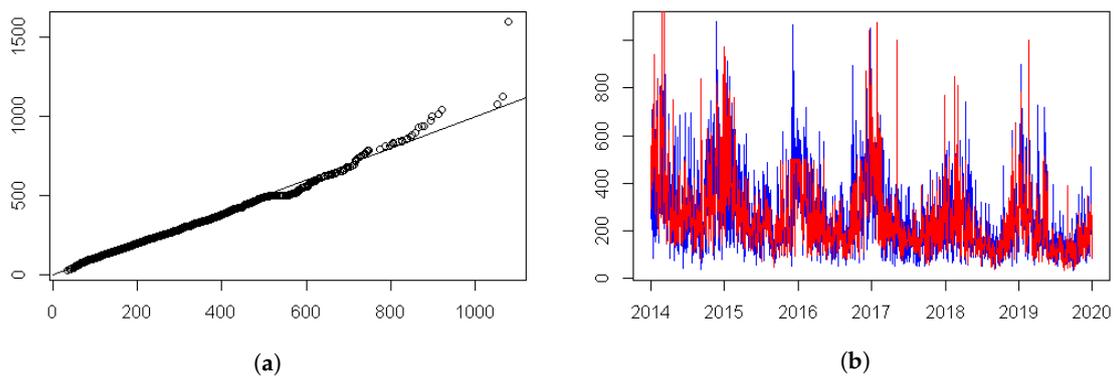


Figure 8. (a) The quantile-quantile plot of \tilde{Q}_t ($\mu\text{g}/\text{m}^3$) using our DCW model (x -axis) with real values Q_t (y -axis, $\mu\text{g}/\text{m}^3$), and (b) time series plot and the comparison between \tilde{Q}_t (blue lines, $\mu\text{g}/\text{m}^3$) and real values Q_t (red lines, $\mu\text{g}/\text{m}^3$).

It is observed that points in Figure 8a distribute around the line of 45-degree, implying that the distribution of fitted data \tilde{Q}_t is close to the one of real Q_t . Tendencies of \tilde{Q}_t also fits well with Q_t in Figure 8b.

An alternative approach is also considered here. As discussed in Section 3, one may apply the AcF model proposed in [51] to capture the heavy-tailedness of financial data. To illustrate the validity of the AcF model in its application on smog data, it is also fitted here then the estimated parameters are used to generate a fitted sequence \tilde{Q}_t . The density curves of the generated values \tilde{Q}_t from both the AcF model and our DCW model combined with the histogram of real data Q_t are depicted in Figure 9a. The QQ-plot based on the AcF model are also given in Figure 9b.

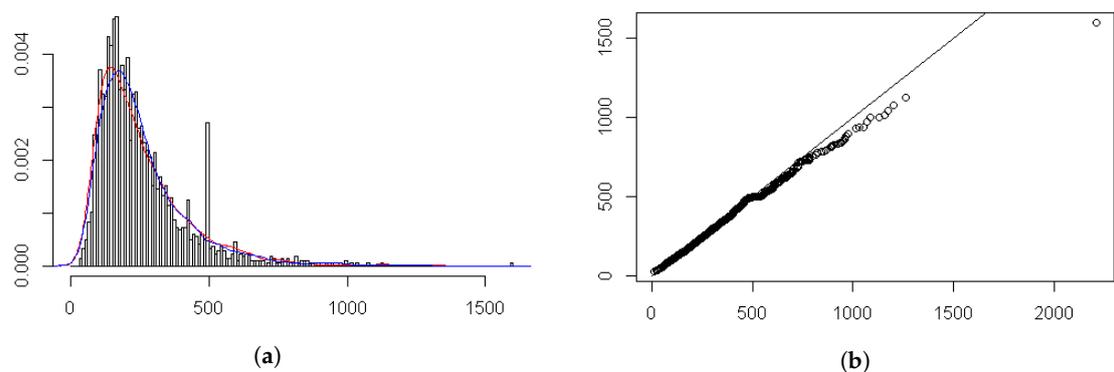


Figure 9. (a) The histogram of the real values Q_t ($\mu\text{g}/\text{m}^3$) and \tilde{Q}_t ($\mu\text{g}/\text{m}^3$); the red curve represents the density curve of fitted values from the DCW model, and the blue one comes from AcF model; and (b) is the QQ-plot of Q_t (y -axis, $\mu\text{g}/\text{m}^3$), and the \tilde{Q}_t ($\mu\text{g}/\text{m}^3$) from AcF model (x -axis)

As shown in Figure 9a, the density curve of generated values \tilde{Q}_t from the DCW model approximate that of real values better than that from the AcF model in both the tail length and location of the peak. Taking the tail as an example, simulated values from the AcF model which exceed 1500 or even 2000 $\mu\text{g}/\text{m}^3$ show much more frequency than that of real data. The same argument can also be put forward according to QQ-plot in Figure 9b, in which the points are obviously lower than the 45-degree line. This bias in fitting extreme high values does not happen in our DCW model in Figure 8a.

It is also worth noting that in Figure 9a, the histogram shows a salient at value 500 $\mu\text{g}/\text{m}^3$, which is caused by the recording mechanism. Under some unknown circumstances, during the period from the end of 2015 to the start of 2016, the $\text{PM}_{2.5}$ values larger than 500 are truncated to 500; that causes an unusually high frequency at 500 in the histogram. Even so, the overall fitting efficiency of our DCW model looks reasonably acceptable using those truncated values 500 $\mu\text{g}/\text{m}^3$ by seeing the generated values \tilde{Q}_t displayed by the red density curve in Figure 9a.

5.2. Inference with Weather Factors

As mentioned before, the smog extremes Q_t often relate to weather factors. We extend Equation (5) to the following (13); then model (5)–(7) becomes (13)–(15):

$$\log \sigma_t = \beta_0 + \beta_1 \log \sigma_{t-1} + \beta_2 \exp(X_{t-1}), \tag{13}$$

$$\log \alpha_t = \gamma_0 + \gamma_1 \log(\alpha_{t-1}) + \gamma_2 \exp(-\gamma_3 Q_{t-1}), \tag{14}$$

$$Q_t = \mu + \sigma_t Y_t^{1/\alpha_t}. \tag{15}$$

where $X_{t-1} = -\beta_3 Q_{t-1} + \beta_4 T p_{t-1} + \beta_5 H u_{t-1} + \beta_6 W d s_{t-1} + \sum_{i=1}^7 c_{\sigma,i} W d d_{i,t-1}$ and $T p_t$, $H u_t$, $W d s_t$, and $W d d_{i,t}$ represent the maximum temperature, the minimum humidity, the mode of wind-level (taking values: 1, 2, 3, 4, ...), and the mode of wind-directions of a day (24 h) of the related city respectively. Seven dummy variables are used to represent eight different wind-directions. Considering the lagged effects and non-negligible impacts that weather factors have on the scale parameter of smog data, the values of weather factors from the last day are used to generate current σ_t . For simplicity, the linear function of weather variables is used here. It can also be extended to non-linear functions to gain higher modeling efficiency. Adding weather factors in model Equation (6) can also be studied, but may lead to a much more difficult parameter estimation process due to the increase of model complexity and dimension of the parameter space. The advanced modeling and analysis will be deferred to our future projects. After fitting model (13)–(15), the fitted parameter values are used to generate (σ_t, α_t) by plugging in real values of the $(t - 1)$ th day in model (13) and (14), where initial values of (σ_1, α_1) are chosen to be greater than zero.

In Figure 10a, the estimated σ_t shows similar seasonality as mentioned in Section 5.1. Besides, from the tendency of estimated σ_t from 2014–2019, it can be seen that both the value and fluctuation of σ_t decrease after 2017 (except that the values larger than 500 during the period from the end of 2015 to the start of 2016 are truncated), showing improvements in air quality. Figure 10b shows that the sequence of α_t shrinks to a constant value quickly. Based on this observation, we set both γ_1 and γ_2 to be zero in Equation (14), which is simplified to the following Equation (17) (the same to Equation (11)). The final model with weather factors is specified as follows:

$$\log \sigma_t = \beta_0 + \beta_1 \log \sigma_{t-1} + \beta_2 \exp(X_{t-1}), \tag{16}$$

$$\alpha_t = \gamma_0, \tag{17}$$

$$Q_t = \mu + \sigma_t Y_t^{1/\alpha_t}, \tag{18}$$

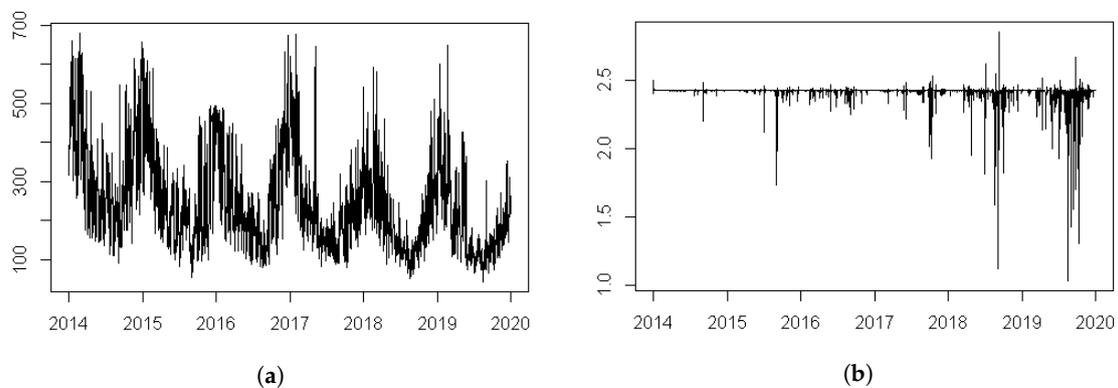


Figure 10. (a) The sequence of fitted σ_t generated from the model, including weather factors; (b) the graph of fitted α_t .

The estimates of model (16)–(18) are listed in Table 5 (The computational times of Tables 4 and 5 along with the computing environment are listed in Appendix B). The last term in Equation (16) is an increasing function of Q_{t-1} and a decreasing function of $(Tp_{t-1}, Hu_{t-1}, Wds_{t-1})$, which coincides with the fact (see Section 2.4) that the scale of the smog tends to be higher when the prior day has higher extreme PM_{2.5} values, together with lower temperature, lower minimum humidity, and lower wind-speed than a normal day.

Table 5. Estimated parameters for model (16)–(18).

Parameter	Fitted Value	SD
μ	3.299×10^1	4.216×10^{-3}
β_0	5.420	1.174×10^{-1}
β_1	1.731×10^{-1}	1.770×10^{-2}
β_2	−1.786	8.367×10^{-2}
β_3	3.724×10^{-3}	1.467×10^{-4}
β_4	7.685×10^{-3}	5.057×10^{-4}
β_5	2.761×10^{-3}	7.391×10^{-4}
β_6	1.531×10^{-2}	6.882×10^{-3}
$c_{\sigma,1}$	1.392×10^{-2}	2.335×10^{-2}
$c_{\sigma,2}$	-1.230×10^{-1}	1.846×10^{-2}
$c_{\sigma,3}$	-1.660×10^{-1}	2.203×10^{-2}
$c_{\sigma,4}$	-1.239×10^{-1}	2.478×10^{-2}
$c_{\sigma,5}$	-8.593×10^{-2}	2.371×10^{-2}
$c_{\sigma,6}$	-1.864×10^{-2}	3.288×10^{-2}
$c_{\sigma,7}$	1.769×10^{-2}	1.957×10^{-2}
γ_0	2.408	1.835×10^{-2}

The estimated parameter values are also used to generate a sequence of fitted values \tilde{Q}_t using the same procedure as is the case without weather factors; then \tilde{Q}_t and real Q_t are plotted in Figure 11 with the left panel being the QQ-plot and the right panel being the line graph.

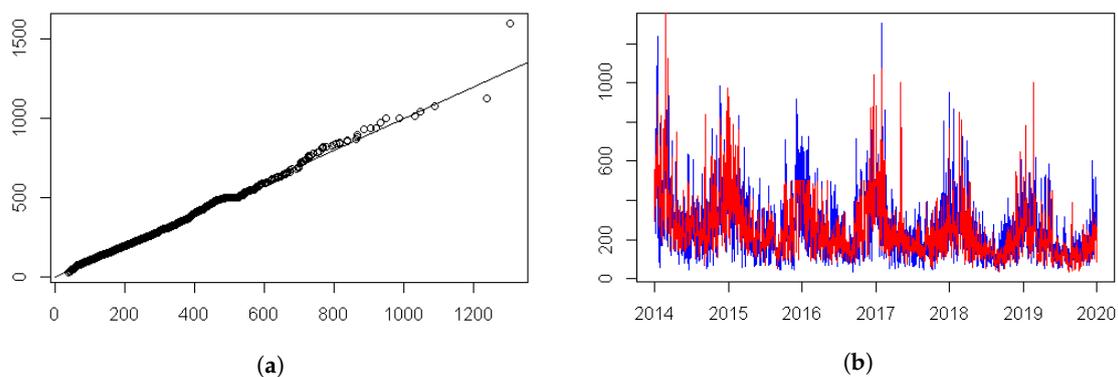


Figure 11. (a) The QQ-plot of the \tilde{Q}_t ($\mu\text{g}/\text{m}^3$) with weather factors (x -axis) and real Q_t (y -axis, $\mu\text{g}/\text{m}^3$); (b) the line graphs of \tilde{Q}_t (blue lines, $\mu\text{g}/\text{m}^3$) and real Q_t (red lines, $\mu\text{g}/\text{m}^3$).

The QQ-plot in Figure 11a shows that the simulated sequence and its true values almost distribute in a line of 45-degrees. Moreover, the scale of \tilde{Q}_t in Figure 11b fits the real values better than it does in Figure 8b, showing more precise description on the overall pattern of the real scenarios. The root mean squared errors between \tilde{Q}_t and real values are also computed in order to compare the two models. They are 155 for the model without weather factors, and 153 for the model with weather factors. From the above, it can be concluded that the model, including weather factors, performs better than the model without weather factors. Considering the truncated extremes during the period from the end of 2015 to the start of 2016, the fitted values from the models with weather factors may be good choices for

those missing values. It is also interesting to notice that the fitted σ_t in Figure 10a and \tilde{Q}_t in Figure 11b have very similar variation tendencies.

Finally, the predictability of our introduced models is explored. Two estimated models (10)–(12) and (16)–(18) are used respectively to forecast the regional daily PM_{2.5} extreme values from January to March in 2020.

Similar results are obtained from the two models. The results from the second model are used to illustrate the prediction process. For the given t -th day, the predicted σ_t (see Figure 12a) is generated via Equation (16) with the one-step-ahead prediction method (the $(t - 1)$ -th day's real Q_{t-1} and weather factors are used). Then we predict the t -th day's smog value Q_t based on our obtained scale σ_t and fitted parameters μ, γ_0 according to Equation (18). The mean values of the 500 repetitions based on the simulated standard Weibull random variables of Y_t are taken as our final predicted values, shown as the red line in Figure 12b.

Compared Figure 12a with the fitted σ_t from 2014–2019 (Figures 7a and 10a); both the value and fluctuation of predicted σ_t in 2020 decrease to some extent, showing that the fluctuations of extreme PM_{2.5} become smaller. Figure 12b shows that our results give a relatively good prediction of the future variation of extreme smog in that the real values almost lie in the 95% prediction intervals. Specifically, our predictions capture the characteristics of extreme smog lying in 200 $\mu\text{g}/\text{m}^3$ –500 $\mu\text{g}/\text{m}^3$ well, although under certain circumstances, the real values exceed this bound due to irregular random factors. These two figures show a similar tendency; i.e., the predicted σ_t and the regional extremes co-move to some extent. Moreover, the values and fluctuations of both series decline more in February and March than in January 2020.

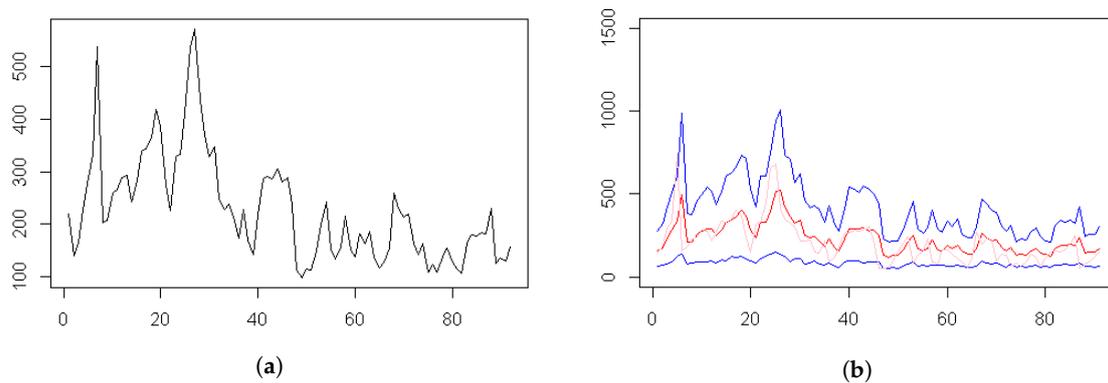


Figure 12. Prediction for the first season of 2020. (a) The predicted σ_t . (b) The regional smog extremes, where the light red line denotes real values of Q_t ($\mu\text{g}/\text{m}^3$), the dark red line denotes predicted values, and the topmost and lowest blue lines are 95% prediction intervals.

6. Conclusions and Discussion

The Beijing–Tianjin–Hebei region is one of the key regions suffering from the severest extreme smog in China. However, it is difficult to model all the PM_{2.5} monitoring stations at the same time; more importantly, the regional extremes are what really matters for this region in order to conduct a joint control strategy. To describe the potential dynamic variation of the regional extremes, this paper integrated classical extreme value modeling and dynamic modeling into a dynamic conditional Weibull distribution modeling and analysis framework, in which the worst scenarios observed among multiple locations in each day during 2014–2019 are described. In addition, weather factors were introduced in the model to gain higher modeling efficiency. The proposed model performs more precisely on fittings compared with other previous models dealing with maxima with autoregressive parameter dynamics (taking the AcF model as an example).

Using the proposed model, the fitted scale parameter and fitted regional smog extremes are obtained and given to show the variation tendency during 2014–2019. It indicates that the extreme smog in the Beijing–Tianjin–Hebei region shows strong seasonality that both its extreme values and

fluctuation are higher in the first and fourth season than the second and the third season. It can also be seen that the extreme values and fluctuations during 2014–2017 are almost the same; i.e., the regional extremes did not improve much during this period, although other works show that a lot cities in the Beijing–Tianjin–Hebei region have experienced lower PM_{2.5} levels since 2014 [35]. However, the extreme values and fluctuations decrease after 2017, showing some improvements in air quality. These findings imply that if the central/local government wants to conduct coordinated joint PM_{2.5} control, the strict treatment strategy must be maintained as long as the regional extremes remain at a high level. It is worth noting that although the regional extremes larger than 500 µg/m³ during the period from the end of 2015 to the start of 2016 are truncated in the original data, they can be fitted by our model, which might be a promising choice for estimating the missing data.

The proposed models can be used to predict the maximum PM_{2.5} level among all monitoring stations in the Beijing–Tianjin–Hebei region. The results of one-step-ahead prediction show that the real values almost lie in the 95% prediction intervals and our predictions capture the variation tendency of extreme smog lying in 200 µg/m³–500 µg/m³ well. Considering that the widely-used meteorologic models for forecasting often require more computation complexity, the one-step-ahead prediction of our model is especially suitable for the short-term forecast and quick response due to its simplicity, operability, and accuracy. More importantly, the predictions of regional extremes rather than single stations are greatly useful to the regional joint early warning system. Under this mechanism, once the future PM_{2.5} level of a single station rather than the average level of the whole region exceeds the breakpoints of a certain grade, the early warning will be activated and the corresponding treatment measures will be taken. In fact, in practical applications, when the current PM_{2.5} values rise rapidly, it will attract public attention. We suggest that most strict measures should be taken when our prediction is close to 500 µg/m³. In this relatively stricter way, the regional coordinate PM_{2.5} prevention and control can be better performed.

This paper gives some theoretical implications as well. Our DCW model can be extended to contain $\log \sigma_{t-q_1}$ and $\log \alpha_{t-q_2}$, which is very similar to the GARCH(q_1, q_2) model. In our application, we have already fit the data very well; there is no need to increase the number of the parameters in our model, which may increase the instability. The advanced method may be more useful in some other situations.

There also exist some needs to investigate dynamics of the sequence of μ , and to add some smooth penalty functions onto the equation in order to guarantee that μ_t is always below the value of Q_t .

It is worth mentioning that existing non-parametric techniques based on GAMs (generalized additive models) or MARS (multivariate adaptive regression splines) discussed in [66–71] are also widely used for fitting nonlinear time series data. In addition, a discussion about joint modeling of the scale and shape parameters of the Weibull distribution with GAMs methodology can be found in [68,71]. Certainly, some comparisons of our model with the existing non-parametric techniques based on GAMs or MARS can give the readers more information about the model suitability in applications. We will implement such comparisons in future projects.

In our simplified model, we did not consider the interactions between the weather factors. However, in reality, interaction effects can exist. Sometimes, they also play an important part. There still remains more work to be done to study the interaction effects. We only added the weather factors onto the equation of scale parameter because they always play a key role. It would be interesting to add the weather factors onto the dynamic tail parameter equation. We shall explore this idea in a different project.

Author Contributions: Conceptualization, Z.Z.; methodology, Z.Z. and M.Y.; software, M.Y. and L.D.; validation, Z.Z., M.Y., and L.D.; formal analysis, M.Y. and L.D.; investigation, Z.Z., M.Y., and L.D.; resources, L.D.; data curation, M.Y. and L.D.; writing—original draft preparation, Z.Z., M.Y., and L.D.; writing—review and editing, Z.Z., M.Y., and L.D.; project administration, Z.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the MOE (Ministry of Education in China) Project of Humanities and Social Sciences (grant number 19YJA790004), the Young Talents Supporting Program of Central University of

Finance and Economics (grant number QYP2005), and the Discipline Funds of Central University of Finance and Economics.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Technical Arguments

Appendix A.1. Proof of Theorem 1

In this subsection, we would like to give our proof for Theorem 1 and our demonstration is built upon some conclusions in [72].

Proof. Without loss of generality, the parameter μ is set as 0, and our model (DCW) defined in (5)–(7) (up to the sign filippings of parameters which does not affect our conclusion) becomes:

$$\log \sigma_t = \beta_0 + \beta_1 \log(\sigma_{t-1}) - \beta_2 \exp(-\beta_3(\sigma_{t-1} Y_{t-1}^{1/\alpha_{t-1}})), \tag{A1}$$

$$\log \alpha_t = \gamma_0 - \gamma_1 \log(\alpha_{t-1}) - \gamma_2 \exp(-\gamma_3(\sigma_{t-1} Y_{t-1}^{1/\alpha_{t-1}})). \tag{A2}$$

Under assumptions given in Theorem 1, without loss of generality, we further assume parameters in Equations (A1) and (A2) satisfy $\beta_1, \gamma_1, \beta_2, \gamma_2 > 0$. We then have that

$$\log \sigma_t = [\beta_0 - z_1 + \beta_1 \log(\sigma_{t-1})] + [z_1 - \beta_2 \exp(-\beta_3(\sigma_{t-1} Y_{t-1}^{1/\alpha_{t-1}}))], \tag{A3}$$

$$\log \alpha_t = [\gamma_0 - z_2 - \gamma_1 \log(\alpha_{t-1})] + [z_2 - \gamma_2 \exp(-\gamma_3(\sigma_{t-1} Y_{t-1}^{1/\alpha_{t-1}}))] \tag{A4}$$

which holds for any z_1, z_2 that satisfies $0 < z_1 < \beta_2$ and $0 < z_2 < \gamma_2$. Next, after denoting $\mathbf{X}_t = (\log \sigma_t, \log \alpha_t)$, we then define:

$$T(\mathbf{X}_{t-1}) = [\beta_0 - z_1 + \beta_1 \log(\sigma_{t-1}), \gamma_0 - z_2 - \gamma_1 \log(\alpha_{t-1})], \tag{A5}$$

$$S(\mathbf{X}_{t-1}, Y_{t-1}) = [z_1 - \beta_2 \exp(-\beta_3(\sigma_{t-1} Y_{t-1}^{1/\alpha_{t-1}})), z_2 - \gamma_2 \exp(-\gamma_3(\sigma_{t-1} Y_{t-1}^{1/\alpha_{t-1}}))]. \tag{A6}$$

Hence, we obtain

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

where $\{Y_t\}_{t \geq 0}$ is a sequence of i.i.d. unit Weibull random variables. Following the terminologies of [72], we obtain that $T(\cdot)$ has a compact attractor $\Lambda = (\frac{\beta_0 - z_1}{1 - \beta_1}, \frac{\gamma_0 - z_2}{1 + \gamma_1})$. In other words, for any $x \in \mathbb{R}^2$, we have $T^n(x) \rightarrow \Lambda$ as $n \rightarrow \infty$. Further, we set G_0 as the area $G_0 = (\frac{\beta_0 - \beta_2}{1 - \beta_1}, \frac{\beta_0}{1 - \beta_1}) \times (\frac{\gamma_0 - \frac{\gamma_2}{1 + \gamma_1}}{1 + \gamma_1}, \frac{\gamma_0 + \frac{\gamma_1 \gamma_2}{1 + \gamma_1}}{1 + \gamma_1})$, which is an open area in \mathbb{R}^2 .

Then we are able to prove that the process $\{X_t\}$ meets five conditions given in Theorem 1 of [72]. Condition (a: Λ has a dense orbit) is proved, since for any x in \mathbb{R}^2 , we have $T^n(x) \rightarrow \Lambda$ as $n \rightarrow \infty$ by our argument above. Conditions (c: Lipschitz continuous over G_0) and (e: $E[S(\mathbf{X}_n, Y_n) | \mathbf{X}_n = x]$ is uniformly bounded on G_0) are satisfied because the area G_0 is bounded and $S(\mathbf{X}_n, Y_n)$ is continuous in Y_n . Next, we will verify the condition (b: exponentially attracting) by leveraging our conclusion from the following Lemma A1. □

Lemma A1. *The area G_0 we have constructed above is an absorbing area for \mathbf{X}_t .*

Proof. The proof of Lemma A1 is deferred to Appendix A.1.1. □

For the remaining part, we need to check the condition (d), which is demonstrated by our Lemma A2 below.

Lemma A2. *For all $x \in G_0$, 0 is in the support set of $\|S(x, Y_{t-1})\|$. Then for all $x \in G_0$, there exists a positive constant r , s.t. the second step transition probability of \mathbf{X}_t , $P^2(x, dy)$ has an absolutely continuous component,*

of which the probability density function is positive over $B(T^2(x), r)$ with $B(x, r)$ being the open ball in G_0 with center x and radius r .

Proof. The detailed proof of Lemma A2 is given in Appendix A.1.2. □

Now we have proved those five conditions given in [72]. Therefore, the Markov chain $\{\mathbf{X}_t = (\log \sigma_t, \log \alpha_t)\}_{t \geq 0}$ defined in (A1) and (A2) possesses the characteristic of geometrically ergodic. Moreover, $\{\mathbf{X}_t\}_{t \geq 0}$ defined in G_0 is irreducible; the Markov chain $\{\mathbf{X}_t\}_{t \geq 0}$ is also stationary in G_0 . Thus, we claim our conclusion of Theorem 1.

Next, we prove the following Lemmas A1 and A2, which are two major ingredients in proving our Theorem 1.

Appendix A.1.1. Proof of Lemma A1

Proof. It is easy to verify this conclusion, if we let $\log \alpha_{t-1} < \frac{\gamma_0 + \frac{\gamma_1 \gamma_2}{1 - \gamma_1}}{1 + \gamma_1}$; then we have $\log(\alpha_t) = \gamma_0 - \gamma_1 \log(\alpha_{t-1}) - \gamma_2 \exp(-\gamma_3 Q_{t-1}) > \gamma_0 - \gamma_1 \frac{\gamma_0 + \frac{\gamma_1 \gamma_2}{1 - \gamma_1}}{1 + \gamma_1} - \gamma_2 = \frac{\gamma_0 - \frac{\gamma_2}{1 + \gamma_1}}{1 + \gamma_1}$. On the other hand, if we set $\log(\alpha_{t-1}) > \frac{\gamma_0 - \frac{\gamma_2}{1 + \gamma_1}}{1 + \gamma_1}$, we can also prove $\log(\alpha_t) < \frac{\gamma_0 + \frac{\gamma_1 \gamma_2}{1 - \gamma_1}}{1 + \gamma_1}$. □

Appendix A.1.2. Proof of Lemma A2

Proof. We set $0 < z_1 < \beta_2, 0 < z_2 = \gamma_2 \exp(\frac{\gamma_3}{\beta_3} \log(\frac{z_1}{\beta_2})) < \gamma_2 < \frac{\gamma_2}{1 - \gamma_1}$. We have for any fixed \mathbf{X}_t , that there always exists a $0 < Y'_t < \infty$ s.t. $Q'_t := \sigma_t Y_t'^{1/\alpha_t} = -\frac{1}{\beta_3} \log \frac{z_1}{\beta_2}$. Considering the value of z_1 and z_2 we have set above, we obtain that given \mathbf{X}_t, Y'_t is the only solution to $S(\mathbf{X}_t, Y_t) = 0$. Therefore, for any $x \in G_0, 0$ is always in the support of $|S(x, Y_{t-1})|$. Thus, we have completed the first part of Lemma A2.

In the following paragraph, we verify that there exists a positive constant r s.t. $P^2(x, dy)$ has an absolutely continuous component whose probability density function is positive over $B(T^2(x), r)$. In addition, we denote Q' as $Q' = -\frac{1}{\beta_3} \log(\frac{z_1}{\beta_2})$ for convenience. Given \mathbf{X}_{t-1} , for $\mathbf{X}_{t+1} = (\log \sigma_{t+1}, \log \alpha_{t+1})$, we obtain the following equations:

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

We notice that given $\mathbf{X}_{t-1}, \mathbf{X}_{t+1}$ is a function of (Q_{t-1}, Q_t) ; we then denote \mathbf{X}_{t+1} as $\mathbf{X}_{t+1} = F_{\mathbf{X}_{t-1}}(Q_{t-1}, Q_t)$. It can be seen that we have $\mathbf{X}_{t+1} = F_{\mathbf{X}_{t-1}}(Q', Q') = T^2(\mathbf{X}_{t-1})$ and the determinant of the Jacobean matrix at point (Q', Q') is given by $(\gamma_1 + \beta_1)\beta_2\beta_3\gamma_2\gamma_3 \exp(-(\beta_3 + \gamma_3)Q')$ which is not zero if $\gamma_1 \neq -\beta_1$ and $\beta_2, \beta_3, \gamma_2, \gamma_3 \neq 0$.

By the inverse function theorem, one knows that there exists an open neighborhood at $\mathbf{X}_{t+1} = F_{\mathbf{X}_{t-1}}(Q', Q') = T^2(\mathbf{X}_{t-1})$, denoted by $B(T^2(\mathbf{X}_{t-1}), r(\mathbf{X}_{t-1}))$, and another open neighborhood at (Q', Q') s.t. there is a bijection between them. It is worth noting that the value of \mathbf{X}_{t-1} does not affect the radius of that open neighborhood, since it is determined by the landscape around point (Q', Q') . Thus, for all \mathbf{X}_{t-1} , we can write $B(T^2(\mathbf{X}_{t-1}), r(\mathbf{X}_{t-1}))$ as $B(T^2(\mathbf{X}_{t-1}), r)$ for some constant r .

Then given $\mathbf{X}_{t-1} \in G_0$, we prove that there exist one-to-one maps between every value in $B(\mathbf{X}_{t+1}, r)$ with values in some open neighborhood of (Q', Q') , so that it can also form a bijection with the neighborhood of (Y'_{t-1}, Y'_t) . According to the formulation of the density function of Weibull distribution, we see that the function P^2 has a positive density over the neighborhood of (Y'_{t-1}, Y'_t) given \mathbf{X}_{t-1} , and due to the existence of the bijection, $P^2(x, dy)$ has an absolutely continuous component whose probability density function is positive over $B(T^2(x), r)$. Then we complete our proof of the second part of Lemma A2. □

Before proceeding our proof for Theorems 2 and 3 and Proposition 1, we would like to remind the readers of some assumptions and notation given in Section 3.2. We assume the true parameter θ_0 is the interior point in Θ , a compact subset of Θ_s ; and the observations are generated from a stationary and

ergodic DCW with the true parameter θ_0 . In addition, we denote $(\sigma_t(\theta), \alpha_t(\theta))$ as the sequence which is generated from the true initial value (σ_1^0, α_1^0) and $(\tilde{\sigma}_t(\theta), \tilde{\alpha}_t(\theta))$ as the one generated from an arbitrary initial value $(\tilde{\sigma}_1, \tilde{\alpha}_1)$ and $\theta \in \Theta$. Moreover, due to the compactness of Θ , there exist uniform upper and lower bounds of the sequence (σ_t, α_t) which are denoted as (σ_U, α_U) and (σ_L, α_L) respectively. Next, in the following part of our proof, we use notation $L_n(\theta)$ to denote the conditional likelihood of θ given $\{(\sigma_t, \alpha_t)\}_{t \geq 0}$

$$L_n(\theta) = \frac{1}{n} \sum_{t=1}^n l_t(\theta) = \frac{1}{n} \sum_{t=1}^n \left[\log \alpha_t - \alpha_t \log \sigma_t + (\alpha_t - 1) \log(Q_t - \mu) - \left(\frac{Q_t - \mu}{\sigma_t} \right)^{\alpha_t} \right] \tag{A7}$$

and notation $\tilde{L}(\theta)$ defined in Equation (9). In the next Appendix A.2, we will first prove several technical lemmas which build blocks for proving Theorems 2 and 3 and Proposition 1.

Appendix A.2. Technical Lemmas

First, we will illustrate the identifiability of our model by the following Lemma A3.

Lemma A3. (Identifiability) *If $Q_t(\theta) = Q_t(\theta_0)$ a.s. for all t , then $\theta = \theta_0$. Here a.s. is for the infinite product space generated by $\{\dots, Y_{-1}, Y_0, Y_1, Y_2, \dots\}$, in which Y_t 's are i.i.d unit Weibull random variables.*

Proof. The proof of Lemma A3 is the same as the proof of Lemma 3 in [51]. □

In the following Lemma A4, we will discuss some characteristics of the score function as well as the Fisher information matrix, based on the true initial value (σ_1^0, α_1^0) , at the true parameter θ_0 .

Lemma A4. *Under the conditions in Theorem 2, we get $E_{\theta_0}[\frac{\partial}{\partial \theta} l_t(\theta_0)] = 0$ and $M_0 = Var_{\theta_0}(\frac{\partial}{\partial \theta} l_t(\theta_0)) = -E_{\theta_0}[\frac{\partial^2}{\partial \theta \partial \theta^T} l_t(\theta_0)]$, in which M_0 is the Fisher information matrix at θ_0 . M_0 is also well defined and positive definite.*

Proof. For the first part: $E_{\theta_0}[\frac{\partial}{\partial \theta} l_t(\theta_0)] = 0$, after interchanging the integration operator with differential operator we obtain

$$\begin{aligned} E_{\theta_0} \left[\frac{\partial \log f_t(Q_t, \theta_0 | \sigma_t, \alpha_t)}{\partial \theta} \right] &= \int \frac{\partial \log f_t(q_t, \theta_0 | \sigma_t, \alpha_t)}{\partial \theta} f(q_t, \theta_0 | \sigma_t, \alpha_t) dq_t \\ &= \int \frac{1}{f_t(q_t, \theta_0 | \sigma_t, \alpha_t)} \frac{\partial f_t(q_t, \theta_0 | \sigma_t, \alpha_t)}{\partial \theta} f(q_t, \theta_0 | \sigma_t, \alpha_t) dq_t = \int \frac{\partial f_t(q_t, \theta_0 | \sigma_t, \alpha_t)}{\partial \theta} dq_t. \end{aligned}$$

Note that for any $x \in (0, 1)$ and $\alpha_L > 2$ (we will assume this in the next lemma) there exists a $c > 0$ s.t. $|\log(x)| \leq \frac{c}{x}$; then it is easy to find a $g(q_t)$ s.t. $|\frac{\partial f_t(q_t, \theta_0 | \sigma_t, \alpha_t)}{\partial \theta}| \leq g(q_t)$ and $\int g(q_t) dq_t < \infty$ for all $\theta \in (\theta_0 - \epsilon, \theta_0 + \epsilon)$ and some $\epsilon > 0$. Then we get $\int \frac{\partial f_t(q_t, \theta_0 | \sigma_t, \alpha_t)}{\partial \theta} dq_t = \frac{\partial}{\partial \theta} \int f_t(q_t, \theta_0 | \sigma_t, \alpha_t) dq_t = 0$ by dominate convergence theorem, which gives: $E_{\theta_0}[\frac{\partial}{\partial \theta} l_t(\theta_0)] = 0$.

Next, after assuming those regular conditions (we can change the integration with the second derivative for the p.d.f) are satisfied by our model, we get:

$$\begin{aligned} E_{\theta_0} \left[\frac{\partial^2 \log f_t(q_t, \theta_0 | \sigma_t, \alpha_t)}{\partial \theta \partial \theta'} \right] &= \int \frac{\partial}{\partial \theta} \left(\frac{1}{f_t(q_t, \theta_0 | \sigma_t, \alpha_t)} \frac{\partial f_t(q_t, \theta_0 | \sigma_t, \alpha_t)}{\partial \theta'} \right) f_t(q_t, \theta_0 | \sigma_t, \alpha_t) dq_t \\ &= - \int \left[\frac{1}{f_t(q_t, \theta_0 | \sigma_t, \alpha_t)} \frac{\partial f_t(q_t, \theta_0 | \sigma_t, \alpha_t)}{\partial \theta} \right] \left[\frac{1}{f_t(q_t, \theta_0 | \sigma_t, \alpha_t)} \frac{\partial f_t(q_t, \theta_0 | \sigma_t, \alpha_t)}{\partial \theta'} \right]^T f_t(q_t, \theta_0 | \sigma_t, \alpha_t) dq_t \\ &\quad + \int \frac{\partial^2 f_t(q_t, \theta_0 | \sigma_t, \alpha_t)}{\partial \theta \partial \theta'} dq_t, \end{aligned}$$

in which we have

$$\int \frac{\partial^2 f_t(q_t, \theta_0 | \sigma_t, \alpha_t)}{\partial \theta \partial \theta'} dq_t = \frac{\partial^2}{\partial \theta \partial \theta'} \int f_t(q_t, \theta_0 | \sigma_t, \alpha_t) dq_t = 0.$$

Then, we have $M_0 = \text{Var}_{\theta_0}(\frac{\partial}{\partial \theta} l_t(\theta_0)) = -E_{\theta_0}[\frac{\partial^2}{\partial \theta \partial \theta'} l_t(\theta_0)]$. As the sequence $\frac{\partial^2}{\partial \theta \partial \theta'} l_t(\theta_0)$ is strictly stationary when t tends to infinity, M_0 is independent with t and is also well defined ($M_0 < \infty$). In order to prove that M_0 is positive definite, we observe that there does not exist a $c \in R^9$ s.t. $c^T \frac{\partial}{\partial \theta} l_t(\theta_0) = 0$ a.s. by following our Lemma A3. \square

According to our expression of the first and second order derivatives of $L_n(\theta_0)$ given in Appendix A.6, the following Lemma A5 shows that their expectations exist.

Lemma A5. Under the assumptions given in Theorem 2, we have (a) for any $\alpha > 0$, $\frac{1}{n} \sum_{t=1}^n (Q_t - \mu_0)^\alpha \rightarrow_p E_{\theta_0}[(Q_1 - \mu_0)^\alpha] < \infty$, (b) For any positive k , $\frac{1}{n} \sum_{t=1}^n [\log(Q_t - \mu_0)]^k \rightarrow_p E_{\theta_0}[\log(Q_1 - \mu_0)]^k < \infty$. Further, if we set $1/\alpha_L < 1/2$, we further obtain (c), $\frac{1}{n} \sum_{t=1}^n (Q_t - \mu_0)^{-\alpha} \rightarrow_p E_{\theta_0}[(Q_1 - \mu_0)^{-\alpha}] < \infty$ with any $0 \leq \alpha \leq 2$.

Proof. Here we leverage the fact that the scale sequence $\{\sigma_t\}$ and the tail sequence $\{\alpha_t\}$ enjoy the boundedness condition stated in Section 3.2. We obtain $E_{\theta_0}[(Q_t - \mu_0)^\alpha] < \infty$ for any α since Y_t follows standard Weibull distribution and $Q_t - \mu_0 < \sigma_U \max(Y_t^{1/\alpha_L}, Y_t^{1/\alpha_U})$. Then (a) is established by the pointwise ergodicity Theorem [73].

As for (b), we have $|\log(Q_t - \mu_0)|^k = |\log \sigma_t + 1/\alpha_t \log Y_t|^k \leq 2^{k-1}(C + 1/\alpha_L |\log Y_t|^k)$ by the convexity of $x^k, k \geq 1$. We also know that $-\log(Y_t)$ follows a Gumbel distribution so we have $E_{\theta_0}[|\log Y_t|^k] < \infty$ for any positive integer k . Then (b) can also be proved by using pointwise ergodicity Theorem [73].

Next, we will prove (c). We know that $(Q_t - \mu_0) > \sigma_L \min(Y_t^{1/\alpha_L}, Y_t^{1/\alpha_U})$ holds and Y_t^{-1} follows the unit Fréchet distribution. After restricting $2 < \alpha_L \leq \alpha_U$, we obtain our conclusion by utilizing the fact that $E[Y_t^{-r}]$ exists when $0 < r < 1$. \square

The following Lemma A6 states that the convergence rate of $\min(Q_t) - \mu_0$ is larger than $n^{-r}, r > 1/\alpha_L$.

Lemma A6. Under the conditions given in Theorem 2, we have $Q_{n,1} - \mu_0 \geq O_p((n)^{-1/\alpha_L})$.

Proof. The conclusion follows directly from $nY_{n,1} \rightarrow_p 1$ and $Q_{n,1} - \mu_0 \geq \sigma_L Y_{n,1}^{1/\alpha_L}$ a.s. when $n \rightarrow \infty$. \square

The next Lemmas A7–A10 will build blocks for proving $\|L_n(\hat{\theta}_n) - L_n(\theta_0)\| \rightarrow_p 0$ in which $\hat{\theta}_n$ denotes the local maximizer of $\tilde{L}_n(\theta)$.

Lemma A7. We denote (a) $S_n^\alpha(\mu) = n^{-1} \sum_{k=1}^n (Q_{n,k} - \mu)^\alpha, \alpha > 0$ or (b) $S_n^\alpha(\mu) = n^{-1} \sum_{k=1}^n \log(Q_{n,k} - \mu)$ or (c) $S_n^\alpha(\mu) = n^{-1} \sum_{k=1}^n (Q_{n,k} - \mu)^\alpha [\log(Q_{n,k} - \mu)]^m$ for $\alpha \geq 0$ and $m = 1, 2, 3$. Under the conditions in Theorem 2, given positive sequence τ_n , s.t. $\tau_n \sim n^{-r}, r > 1/\alpha^L$, the following result holds uniformly over $|\mu_n - \mu_0| < \tau_n$,

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

Remark A1. For $X = \frac{1}{n} \sum_{k=1}^n (Q_{n,k} - \mu_n)^\alpha, \alpha \geq 0$ we have $X \leq \frac{1}{n} \sum_{k=1}^n (Q_{n,k} - \mu_0 + \tau_n)^\alpha$. Since we have obtained $Q_{n,1} - \mu_0 \geq O_p((n)^{-1/\alpha_L})$ in Lemma A6, for any fixed $0 < \rho < 1, \tau_n \sim n^{-r}, 1/\alpha_L < r < 1/2$, we are able to verify that $P(\rho(Q_{n,1} - \mu_0) > \tau_n) \rightarrow_p 1$ holds. Then, as we have $P(\rho(Q_{n,k} - \mu_0) > \tau_n, 1 \leq k \leq n) \rightarrow_p 1$, we obtain $X \leq \frac{1}{n} \sum_{k=1}^n ((1 + \rho)(Q_{n,k} - \mu_0))^\alpha$. When we have $-2 \leq \alpha < 0, X \leq \frac{1}{n} \sum_{k=1}^n (Q_{n,k} - \mu_0 - \tau_n)^\alpha \leq \frac{1}{n} \sum_{k=1}^n ((1 - \rho)(Q_{n,k} - \mu_0))^\alpha$ can also be proved by using similar arguments.

Proof. For the proof of (a), we obtain $|S_n^\alpha(\mu_n) - S_n^\alpha(\mu_0)| \leq \frac{1}{n} \sum_{k=1}^n |(Q_{n,k} - \mu_n)^\alpha - (Q_{n,k} - \mu_0)^\alpha| \leq \frac{1}{n} \sum_{k=1}^n [\alpha(Q_{n,k} - \mu_n)^{\alpha-1} + \alpha(Q_{n,k} - \mu_0)^{\alpha-1}] |\mu_n - \mu_0| \leq \frac{\tau_n}{n} \sum_{k=1}^n [\alpha(Q_{n,k} - \mu_n)^{\alpha-1} + \alpha(Q_{n,k} - \mu_0)^{\alpha-1}] \leq$

$\frac{2\tau_n}{n} \sum_{k=1}^n \alpha \max\{(Q_{n,k} - \mu_n)^{\alpha-1}, (Q_{n,k} - \mu_0)^{\alpha-1}\}$, from the remark given above. Thus, we conclude that (a) $|S_n^\alpha(\mu_n) - S_n^\alpha(\mu_0)| \leq O_p(\tau_n)$ holds with probability going to 1.

The Proof of (b) is similar with the corresponding part in Lemma 7 given in [51].

For (c), when $m = 1$, first, we separate the inequality into two parts, $|S_n^\alpha(\mu_n) - S_n^\alpha(\mu_0)| \leq \frac{1}{n} \sum_{k=1}^n (Q_{n,k} - \mu_n)^\alpha |\log(Q_{n,k} - \mu_n) - \log(Q_{n,k} - \mu_0)| + \frac{1}{n} \sum_{k=1}^n |(Q_{n,k} - \mu_n)^\alpha - (Q_{n,k} - \mu_0)^\alpha| |\log(Q_{n,k} - \mu_0)|$. Then for the first part, we get $\frac{1}{n} \sum_{k=1}^n (Q_{n,k} - \mu_n)^\alpha |\log(Q_{n,k} - \mu_n) - \log(Q_{n,k} - \mu_0)| \leq \frac{1}{n} \sum_{k=1}^n (Q_{n,k} - \mu_n)^\alpha \log(1 + \frac{\mu_n - \mu_0}{Q_{n,k} - \mu_n}) \leq \frac{\tau_n}{n} \sum_{k=1}^n (Q_{n,k} - \mu_n)^{\alpha-1}$. As for the second part, we obtain $\frac{1}{n} \sum_{k=1}^n |(Q_{n,k} - \mu_n)^\alpha - (Q_{n,k} - \mu_0)^\alpha| |\log(Q_{n,k} - \mu_0)| \leq \frac{2\tau_n}{n} \sum_{k=1}^n \alpha \max\{(Q_{n,k} - \mu_n)^{\alpha-1}, (Q_{n,k} - \mu_0)^{\alpha-1}\} |\log(Q_{n,k} - \mu_0)| \leq 2\tau_n (\frac{1}{n} \sum_{k=1}^n \alpha^2 \max\{(Q_{n,k} - \mu_n)^{2\alpha-2}, (Q_{n,k} - \mu_0)^{2\alpha-2}\})^{1/2} (\frac{1}{n} \sum_{k=1}^n |\log(Q_{n,k} - \mu_0)|^2)^{1/2} = O_p(\tau_n)$. This is the case when $\mu_n \geq \mu_0$. For $\mu_n < \mu_0$, the process of our proof is similar so we just omit the details here. For $m = 2$, we get $|S_n^\alpha(\mu_n) - S_n^\alpha(\mu_0)| \leq \frac{1}{n} \sum_{k=1}^n (Q_{n,k} - \mu_n)^\alpha \log(Q_{n,k} - \mu_n) |\log(Q_{n,k} - \mu_n) - \log(Q_{n,k} - \mu_0)| + \frac{1}{n} \sum_{k=1}^n |(Q_{n,k} - \mu_n)^\alpha \log(Q_{n,k} - \mu_n) - (Q_{n,k} - \mu_0)^\alpha \log(Q_{n,k} - \mu_0)| \log(Q_{n,k} - \mu_0)$. Similarly, for the first part, we obtain $\frac{1}{n} \sum_{k=1}^n (Q_{n,k} - \mu_n)^\alpha \log(Q_{n,k} - \mu_n) |\log(Q_{n,k} - \mu_n) - \log(Q_{n,k} - \mu_0)| \leq \frac{\tau_n}{n} \sum_{k=1}^n (Q_{n,k} - \mu_n)^{\alpha-1} |\log(Q_{n,k} - \mu_n)| \leq \tau_n (\frac{1}{n} \sum_{k=1}^n (Q_{n,k} - \mu_n)^{2\alpha-2})^{1/2} (\frac{1}{n} \sum_{k=1}^n |\log(Q_{n,k} - \mu_n)|^2)^{1/2} = O_p(\tau_n)$. For the second part, we next use some conclusions that we got in the process of proving the case of $m = 1$. So the second part is proved as follows:

$$\begin{aligned} & \frac{1}{n} \sum_{k=1}^n |(Q_{n,k} - \mu_n)^\alpha \log(Q_{n,k} - \mu_n) - (Q_{n,k} - \mu_0)^\alpha \log(Q_{n,k} - \mu_0)| \log(Q_{n,k} - \mu_0) \\ & \leq \frac{1}{n} \sum_{k=1}^n [\tau_n (Q_{n,k} - \mu_n)^{\alpha-1} + 2\tau_n \alpha \max\{(Q_{n,k} - \mu_n)^{\alpha-1}, (Q_{n,k} - \mu_0)^{\alpha-1}\} \times \\ & \quad |\log(Q_{n,k} - \mu_0)|] |\log(Q_{n,k} - \mu_0)| \\ & \leq \frac{\tau_n}{n} \sum_{k=1}^n \max\{(Q_{n,k} - \mu_n)^{\alpha-1}, (Q_{n,k} - \mu_0)^{\alpha-1}\} P_2(|\log(Q_{n,k} - \mu_0)|) \\ & \leq \tau_n (\frac{1}{n} \sum_{k=1}^n \max\{(Q_{n,k} - \mu_n)^{2\alpha-2}, (Q_{n,k} - \mu_0)^{2\alpha-2}\})^{1/2} (\frac{1}{n} \sum_{k=1}^n P_2^2(|\log(Q_{n,k} - \mu_0)|))^{1/2}, \end{aligned}$$

in which $P_j(x)$ denotes a polynomial of order j .

For $m = 3$, we have $|S_n^\alpha(\mu_n) - S_n^\alpha(\mu_0)| \leq \frac{1}{n} \sum_{k=1}^n (Q_{n,k} - \mu_n)^\alpha \log^2(Q_{n,k} - \mu_n) |\log(Q_{n,k} - \mu_n) - \log(Q_{n,k} - \mu_0)| + \frac{1}{n} \sum_{k=1}^n |(Q_{n,k} - \mu_n)^\alpha \log^2(Q_{n,k} - \mu_n) - (Q_{n,k} - \mu_0)^\alpha \log^2(Q_{n,k} - \mu_0)| \log(Q_{n,k} - \mu_0)$. It is easy to see that the first part can be regarded as $O_p(\tau_n)$. As for the second part, we have $\leq \frac{\tau_n}{n} \sum_{k=1}^n \max\{(Q_{n,k} - \mu_n)^{\alpha-1}, (Q_{n,k} - \mu_0)^{\alpha-1}\} P_3(|\log(Q_{n,k} - \mu_0)|)$. Then the proof for Lemma A7 is completed after applying the Holder's inequality. \square

In the next Lemmas A8 and A9 we will prove that the supremum of the difference between the first n values of σ_t and σ_t^0 (so as α_t and α_t^0), which are generated by using arbitrary parameter θ in the neighborhood of the true value and the true one θ_0 respectively, converges at the rate of τ_n . This convergence rate also holds for their partial derivatives.

Lemma A8. Denote $\Phi = (\gamma_0, \gamma_1, \gamma_2, \gamma_3)$ and $\Phi_0 = (\gamma_0^0, \gamma_1^0, \gamma_2^0, \gamma_3^0)$, if $\|\Phi - \Phi_0\| < \tau_n$ and $\tau_n \searrow 0$. under the conditions in Theorem 2, we have:

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

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

Proof. Here we briefly illustrate our proof of (a), the proofs of (b), (c) are almost the same. The domain of α_t is bounded so the function $\exp(\cdot)$ defined on a compact set is Lipschitz continuous. Then it is equivalent to prove: $\sup_{1 \leq t \leq n} |\log \alpha_t - \log \alpha_t^0| = O(\tau_n)$. As we can express $\log \alpha_t$ as

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

we further obtain

$$\begin{aligned} |\log \alpha_t - \log \alpha_t^0| &\leq \left| \gamma_0 \sum_{k=1}^{t-1} (-\gamma_1)^{k-1} - \gamma_0^0 \sum_{k=1}^{t-1} (-\gamma_1^0)^{k-1} \right| + \left| (-\gamma_1)^{t-1} \log \alpha_1^0 - (-\gamma_1^0)^{t-1} \log \alpha_1^0 \right| \\ &+ \left| \gamma_2 \sum_{k=1}^{t-1} (-\gamma_1)^{k-1} \exp(-\gamma_3 Q_{t-k}) - \gamma_2^0 \sum_{k=1}^{t-1} (-\gamma_1^0)^{k-1} \exp(-\gamma_3^0 Q_{t-k}) \right|. \end{aligned}$$

The rest parts are similar with the corresponding parts in Lemma 8 [51]. □

Lemma A9. We denote $\Psi = (\beta_0, \beta_1, \beta_2, \beta_3)$ and $\Psi_0 = (\beta_0^0, \beta_1^0, \beta_2^0, \beta_3^0)$. Under the conditions in Theorem 2, if we have $\|\Psi - \Psi_0\| < \tau_n$ and $\tau_n \searrow 0$, we obtain

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

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

Proof. The proof of this Lemma is almost the same of the proof of Lemma A8. □

The next Lemma A10 will build blocks for the proof of Lemma A11.

Lemma A10. Suppose we have $\tau_n \sim n^{-r}$ and $\sup_{1 \leq t \leq n} |\alpha_t - \alpha_t^0| = O(\tau_n)$, where $\{\alpha_t\}$ and $\{\alpha_t^0\}$ represent two different sequences of tail index that are generated basing on different parameters (Φ, Φ_0) and $\|\Phi - \Phi_0\| < \tau_n$ with the true initial value. Under the conditions in Theorem 2, we have

$$\frac{1}{n} \sum_{t=1}^n |(Q_t - \mu_n)^{\alpha_t} - (Q_t - \mu_n)^{\alpha_t^0}| = O_p(\tau_n),$$

uniformly over $|\mu_n - \mu_0| < \tau_n$. The same result also holds for

$$\frac{1}{n} \sum_{t=1}^n |(Q_t - \mu_n)^{\alpha_t} - (Q_t - \mu_n)^{\alpha_t^0}| [\log(Q_t - \mu_n)]^k, k = 1, 2.$$

Proof. We will only give our proof for the case of $\frac{1}{n} \sum_{t=1}^n |(Q_t - \mu_n)^{\alpha_t} - (Q_t - \mu_n)^{\alpha_t^0}|$ here, the proofs of other two cases are similar. Without loss of generality, we assume $\alpha_t^0 > \alpha_t$, then we obtain

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^n |(Q_t - \mu_n)^{\alpha_t} - (Q_t - \mu_n)^{\alpha_t^0}| &\leq \frac{C}{n} \sum_{t=1}^n (Q_t - \mu_n)^{\alpha^*} |\log(Q_t - \mu_n)| \tau_n \\ &\leq \frac{C\tau_n}{n} \sum_{t=1}^n ((Q_t - \mu_n)^{\alpha_L} + (Q_t - \mu_n)^{\alpha_U}) |\log(Q_t - \mu_n)| = O_p(\tau_n), \end{aligned}$$

in which $\alpha^* \in (\alpha_t, \alpha_t^0)$. □

Together with our conclusions from Lemma A7–A10, we prove $\|L_n(\hat{\theta}_n) - L_n(\theta_0)\| \rightarrow_p 0$ when $\|\hat{\theta}_n - \theta_0\| = O_p(\tau_n)$ in the following Lemma A11.

Lemma A11. We denote $m_{\theta_i, \theta_j}(\theta_0) = -E_{\theta_0}[\frac{\partial^2}{\partial \theta_i \partial \theta_j} l_1(\theta_0)]$. Under the conditions in Theorem 2, for all second order derivatives of $L_n(\hat{\theta}_n)$ we have $\frac{\partial^2}{\partial \theta_i \partial \theta_j} L_n(\hat{\theta}_n) \rightarrow_p -m_{\theta_i, \theta_j}(\theta_0)$, uniformly over $\|\hat{\theta}_n - \theta_0\| < \tau_n$, where $\tau_n \sim n^{-r}, 1/\alpha_L < r < 1/2$.

Proof. Here we just give our proof for the case of $\frac{\partial}{\partial \mu^2} L_n(\hat{\theta}_n)$; the proofs of remaining cases are similar. Note that the first and second order of the partial derivatives of $L_n(\cdot)$ are measurable functions of the stationary and ergodic series $\{Q_t\}_{t \geq 0}$, so they are also ergodic and strictly stationary. By the pointwise ergodicity Theorem [73], we have $\frac{\partial}{\partial \theta_i \partial \theta_j} L_n(\theta_0) \rightarrow_p -m_{\theta_i, \theta_j}(\theta_0)$. Thus, we next need to prove that $\frac{\partial}{\partial \mu^2} L_n(\hat{\theta}_n) - \frac{\partial}{\partial \mu^2} L_n(\theta_0) \rightarrow_p 0$ holds. By definition, we obtain

$$\begin{aligned} \frac{\partial}{\partial \mu^2} L_n(\hat{\theta}_n) - \frac{\partial}{\partial \mu^2} L_n(\theta_0) &= \frac{1}{n} \sum_{t=1}^n [-(\alpha_t - 1)(Q_t - \mu_n)^{-2} + (\alpha_t^0 - 1)(Q_t - \mu_0)^{-2}] \\ &\quad - \frac{1}{n} \sum_{t=1}^n [-\alpha_t(\alpha_t - 1)\sigma_t^{-\alpha_t}(Q_t - \mu_n)^{\alpha_t - 2} + \alpha_t^0(\alpha_t^0 - 1)(\sigma_t^0)^{-\alpha_t^0}(Q_t - \mu_0)^{\alpha_t^0 - 2}] \\ &:= \mathbf{I} + \mathbf{II}. \end{aligned}$$

If **I** is greater than zero, we have

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

When $\mu_0 \leq \mu_n$, we get

$$= (Q_t - \mu_n)^{-2} \left[(\alpha_t^0 - 1) \left(1 + \frac{\mu_0 - \mu_n}{Q_t - \mu_0} \right)^2 - (\alpha_t - 1) \right] \leq (Q_t - \mu_n)^{-2} \sup [(\alpha_t^0 - 1) - (\alpha_t - 1)].$$

From Lemma A6, we have $(nY_{n,1})^{1/\alpha_L} \rightarrow_p 1$ and $\tau_n \sim n^{-r}, 1/\alpha_L < r < 1/2$, so it holds that $\tau_n / (\alpha_L Y_{n,1}^{1/\alpha_L}) = O_p(n^{-\alpha}) \rightarrow_p 0$, with $0 < \alpha < r - 1/\alpha_L$. Thus, if $\mu_0 > \mu_n$, we have

$$\begin{aligned} (Q_t - \mu_n)^{-2} \left[(\alpha_t^0 - 1) \left(1 + \frac{\mu_0 - \mu_n}{Q_t - \mu_0} \right)^2 - (\alpha_t - 1) \right] &\leq (Q_t - \mu_n)^{-2} \left[(\alpha_t^0 - 1) \left(1 + \frac{\tau_n}{\alpha_L Y_{n,1}^{1/\alpha_L}} \right)^2 - (\alpha_t - 1) \right] \\ &\leq (Q_t - \mu_n)^{-2} [\sup |\alpha_t^0 - \alpha_t| + \sup |\alpha_t^0 - 1| O_p(n^{-\alpha})]. \end{aligned}$$

When I is less than zero, we have:

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

If $\mu_0 > \mu_n$, we get:

$$(Q_t - \mu_n)^{-2} [(\alpha_t - 1) - (\alpha_t^0 - 1) \left(1 + \frac{\mu_0 - \mu_n}{Q_t - \mu_0}\right)^2] \leq (Q_t - \mu_n)^{-2} \sup |\alpha_t - \alpha_t^0|.$$

Then if $\mu_0 \leq \mu_n$, we have:

$$\begin{aligned} & (Q_t - \mu_n)^{-2} [(\alpha_t - 1) - (\alpha_t^0 - 1) \left(1 + \frac{\mu_0 - \mu_n}{Q_t - \mu_0}\right)^2] \\ & \leq (Q_t - \mu_n)^{-2} [(\alpha_t - 1) - (\alpha_t^0 - 1) \left(1 - \frac{\tau_n}{\alpha_L Y_{n,1}^{1/\alpha_L}}\right)^2] \\ & \leq (Q_t - \mu_n)^{-2} [\sup |\alpha_t - \alpha_t^0| + \sup |\alpha_t^0 - 1| O_p(n^{-\alpha})]. \end{aligned}$$

So it can be seen that the first part $\leq O_p(\tau_n) + O_p(n^{-\alpha}) \rightarrow_p 0, 0 < \alpha < r - 1/\alpha_L$.

Further, for the second term II, we have

$$\begin{aligned} \text{II} &= \left| \frac{1}{n} \sum_{t=1}^n [-\alpha_t(\alpha_t - 1)\sigma_t^{-\alpha_t}(Q_t - \mu_n)^{\alpha_t-2} + \alpha_t^0(\alpha_t^0 - 1)(\sigma_t^0)^{-\alpha_t^0}(Q_t - \mu_0)^{\alpha_t^0-2}] \right| \\ &\leq \frac{1}{n} \sum_{t=1}^n \alpha_t^0(\alpha_t^0 - 1)(\sigma_t^0)^{-\alpha_t^0} |(Q_t - \mu_n)^{\alpha_t-2} - (Q_t - \mu_0)^{\alpha_t^0-2}| \\ &\quad + \frac{1}{n} \sum_{t=1}^n \alpha_t^0(\alpha_t^0 - 1)(\sigma_t^0)^{-\alpha_t^0} |(Q_t - \mu_n)^{\alpha_t^0-2} - (Q_t - \mu_n)^{\alpha_t-2}| \\ &\quad + \frac{1}{n} \sum_{t=1}^n |\alpha_t^0(\alpha_t^0 - 1)(\sigma_t^0)^{-\alpha_t^0} - \alpha_t(\alpha_t - 1)(\sigma_t)^{-\alpha_t}| (Q_t - \mu_n)^{\alpha_t-2} := \text{i} + \text{ii} + \text{iii}. \end{aligned}$$

The first term (i) goes to zero by Lemma A7 (a), and the second term (ii) goes to zero by Lemma A10. Due to the boundedness of the α_t, σ_t , by utilizing the differential mean value Theorem, we prove that the third term (iii) converges to 0 in probability. \square

We have already proved $\|L_n(\hat{\theta}_n) - L_n(\theta_0)\| \rightarrow_p 0$ when $\|\hat{\theta}_n - \theta_0\| = O_p(\tau_n)$. And note that verifying $\|\check{L}_n(\hat{\theta}_n) - L_n(\theta_0)\| \rightarrow_p 0$ is one of our final goals. Thus, we will prove the next Lemma A12 as well as Lemma A13 in order to demonstrate $\|\check{L}_n(\hat{\theta}_n) - L_n(\hat{\theta})\| \rightarrow_p 0$.

Lemma A12. Under the conditions in Theorem 2, there exists a positive constant C and $0 < C_b < 1$ s.t. for all $\theta \in \Theta$ and $t \geq 1$.

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

$$(d) |\sigma_t - \check{\sigma}_t| \leq C \cdot C_b^{t-1}, (e) \left| \frac{\partial \sigma_t}{\partial \Psi_i} - \frac{\partial \check{\sigma}_t}{\partial \Psi_i} \right| \leq C \cdot t C_b^{t-1}, (f) \left| \frac{\partial^2 \sigma_t}{\partial \Psi_i \partial \Psi_j} - \frac{\partial^2 \check{\sigma}_t}{\partial \Psi_i \partial \Psi_j} \right| \leq C \cdot t^2 C_b^{t-1}.$$

Proof. The proof of this lemma follows from direct calculation, so we just omit the detailed parts here. \square

Lemma A13. Under the conditions of Theorem 2, we have $\frac{1}{n} \sum_{t=1}^n |(Q_t - \mu_n)^{\alpha_t} - (Q_t - \mu_n)^{\tilde{\alpha}_t}| \rightarrow_p 0$, uniformly over $|\mu_n - \mu_0| < \tau_n$, where $\tau_n \sim n^{-r}, r > 0$. The same result holds for

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

Proof. By mean value theorem, we have

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^n |(Q_t - \mu_n)^{\alpha_t} - (Q_t - \mu_n)^{\tilde{\alpha}_t}| &\leq \frac{C}{n} \sum_{t=1}^n (Q_t - \mu_n)^{\alpha_t^*} |\log(Q_t - \mu_n)| C_b^{t-1} \\ &\leq \frac{C}{n} \sum_{t=1}^n [(Q_t - \mu_n)^{\alpha_L} + (Q_t - \mu_n)^{\alpha_U}] |\log(Q_t - \mu_n)| C_b^{t-1} \rightarrow_p 0. \end{aligned}$$

Thus, we claim our conclusion for Lemma A13. □

Next, we would like to discuss Lemma A14 which can be utilized to prove Theorems 2 and 3.

Lemma A14. Under the conditions in Theorem 2, we have (a) : for all second order derivatives of $\tilde{L}_n(\hat{\theta}_n)$, we have $\frac{\partial^2}{\partial \theta_i \partial \theta_j} \tilde{L}_n(\hat{\theta}_n) \rightarrow_p -m_{\theta_i \theta_j}(\theta_0)$, uniformly over $\|\hat{\theta}_n - \theta_0\| < \tau_n$, where $\tau_n \sim n^{-r}, 1/\alpha_L < r < 1/2$. (b) : for the score function of $\tilde{L}_n(\theta)$, we have $(\tau_n^*)^{-1} (\frac{\partial}{\partial \mu} \tilde{L}_n(\theta_0) - \frac{\partial}{\partial \mu} L_n(\theta_0)) \rightarrow_p 0$ if $\tau_n^* n \rightarrow \infty$.

Proof. For the proof of part (a), first, we see that $\frac{\partial^2}{\partial \theta_i \partial \theta_j} \tilde{L}_n(\theta) - \frac{\partial^2}{\partial \theta_i \partial \theta_j} L_n(\theta) \rightarrow_p 0$ holds for any θ by using our conclusions from Lemmas A12 and A13. In addition, we also know that $\frac{\partial^2}{\partial \theta_i \partial \theta_j} L_n(\hat{\theta}_n) - \frac{\partial^2}{\partial \theta_i \partial \theta_j} L_n(\theta_0) \rightarrow_p 0$ holds uniformly over $\|\hat{\theta}_n - \theta_0\| \leq \tau_n$ by Lemma A11. Thus, we get $\frac{\partial^2}{\partial \theta_i \partial \theta_j} \tilde{L}_n(\hat{\theta}_n) - \frac{\partial^2}{\partial \theta_i \partial \theta_j} L_n(\theta_0) \rightarrow_p 0$ over $\|\hat{\theta}_n - \theta_0\| \leq \tau_n$, which claims the first part of Lemma A14.

For the proof of part (b), here we just prove the case of $\frac{\partial}{\partial \mu} \tilde{L}_n(\theta_0)$ and the proof of remaining parts are similar. For convenience, we set $g(\sigma_t, \alpha_t) = \frac{\alpha_t}{\sigma_t^{\alpha_t}}$, then we are able to verify that $|g(\sigma_t, \alpha_t) - g(\tilde{\sigma}_t, \tilde{\alpha}_t)| \leq C \cdot C_b^{t-1}$ holds by utilizing Lemma A12 after assuming σ_L is greater than zero. Next, we obtain

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

The rest parts of our proof of Lemma A14 are similar with the corresponding proof (Lemma 14) in [51], so we omit the details. □

We will leverage on the martingale difference to prove the following Lemma A15 which displays the asymptotic distribution of the score function.

Lemma A15. Under the conditions in Theorem 2,

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

where M_0 is the Fisher information matrix valued at θ_0 .

Proof. We use CLT for martingale difference [74], then we have:

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

So for any $\lambda \in R^9, \{\lambda \frac{\partial l_t(\theta_0)}{\partial \theta}, \mathcal{F}_t\}_t$ is a square-integrable stationary martingale difference. Note that the sequences σ_t, α_t and Q_t are both stationary and ergodic, so the sequences $\frac{\partial \alpha_t}{\partial \Phi}, \frac{\partial \sigma_t}{\partial \Psi}$ are also strictly stationary and ergodic. We also know that $\frac{\partial l_t(\theta_0)}{\partial \theta_i}$ (for $i = 1, \dots, 9$) are generated from $\alpha_t, \sigma_t, Q_t, \frac{\partial \alpha_t}{\partial \Phi}, \frac{\partial \sigma_t}{\partial \Psi}$ so they also follow the properties of strict stationarity and ergodicity. Then by CLT and Wold-Cramer device in [74], we can finally get the conclusion that Lemma A15 holds. \square

Given these technical Lemmas, we next give our proof of our Theorems 2 and 3 and Proposition 1.

Appendix A.3. Proof of Theorem 2

Proof. We let $\{\tau_n\}_{n \in Z^+}$ be any sequence with $\tau_n \sim n^{-r}, 1/\alpha_L < r < 1/2$. Next we set $t \in R, y \in R^8$ and define $f_n(t, y) = \tau_n^{-2} \tilde{L}_n(\mu_0 + \tau_n t, \phi_0 + \tau_n y)$, in which ϕ_0 is defined as $\phi_0 = (\beta_0^0, \beta_1^0, \beta_2^0, \beta_3^0, \gamma_0^0, \gamma_1^0, \gamma_2^0, \gamma_3^0)$. By Taylor Expansion, we get

$$\begin{aligned} \frac{\partial}{\partial t} f_n(t, y) &= \tau_n^{-1} \frac{\partial \tilde{L}_n(\mu_0 + \tau_n t, \phi_0 + \tau_n y)}{\partial \mu} \\ &= \tau_n^{-1} \frac{\partial \tilde{L}_n(\mu_0, \phi_0)}{\partial \mu} + \frac{\partial^2 \tilde{L}_n(\mu^*, \phi^*)}{\partial \mu^2} t + \sum_{i=1}^8 \frac{\partial^2 \tilde{L}_n(\mu^*, \phi^*)}{\partial \mu \partial \phi_i} y_i \\ &= \tau_n^{-1} \left(\frac{\partial \tilde{L}_n(\mu_0, \phi_0)}{\partial \mu} - \frac{\partial L_n(\mu_0, \phi_0)}{\partial \mu} \right) + \tau_n^{-1} \frac{\partial L_n(\mu_0, \phi_0)}{\partial \mu} + \frac{\partial^2 \tilde{L}_n(\mu^*, \phi^*)}{\partial \mu^2} t \\ &\quad + \sum_{i=1}^8 \frac{\partial^2 \tilde{L}_n(\mu^*, \phi^*)}{\partial \mu \partial \phi_i} y_i. \end{aligned}$$

It can be observed that $|\mu^* - \mu_0| < \tau_n t$ and $\|\phi^* - \phi_0\| < \tau_n \|y\|$ due to mean value theorem. Hence, the first term goes to 0 according to Lemma A14 (b) and the second term goes to 0 by Lemma A15. Further, the last two terms converge in probability by utilizing Lemma A14 (a) i.e.,

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

in which $m_{\mu\phi_i}(\theta_0) = -E_{\theta_0}[\frac{\partial^2}{\partial \mu \partial \phi_i} l_1(\theta_0)]$ and $m_{\mu\mu}(\theta_0) = -E_{\theta_0}[\frac{\partial^2}{\partial \mu^2} l_1(\theta_0)]$. Then we obtain $\frac{\partial}{\partial t} f_n(t, y) = -m_{\mu\mu}(\theta_0)t - \sum_{i=1}^8 m_{\mu\phi_i}(\theta_0)y_i + o_p(1)$. Similarly, we also have $\frac{\partial}{\partial y_i} f_n(t, y) = -m_{\phi_i\mu}(\theta_0)t - \sum_{j=1}^8 m_{\phi_i\phi_j}(\theta_0)y_j + o_p(1)$ for $i = 1, \dots, 8$, where $o_p(1)$'s are decaying uniformly over $t^2 + \|y\|^2 \leq 1$.

Let $t^2 + \|y\|^2 = 1$, as the Fisher information matrix M_0 at θ_0 is positive definite according to our Lemma A4, we have

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

According to the Lemma 5 given in [64], we obtain that there is a local maximum over the open area $t^2 + \|y\|^2 < 1$ with probability going to 1. Thus, there exists a sequence of local maximizer $\hat{\theta}_n$ of $\tilde{L}_n(\theta)$ s.t. $\hat{\theta}_n \rightarrow_p \theta_0$ and $\|\hat{\theta}_n - \theta_0\| \leq \tau_n$, where $\tau_n \sim n^{-r}$ and $1/\alpha_L < r < 1/2$. \square

Appendix A.4. Proof of Theorem 3

Proof. By Taylor expansion we have

$$\frac{\partial \tilde{L}_n(\hat{\theta}_n)}{\partial \theta} = \frac{\partial \tilde{L}_n(\theta_0)}{\partial \theta} + \frac{\partial^2 \tilde{L}_n(\theta^*)}{\partial \theta \partial \theta'} (\hat{\theta}_n - \theta_0)$$

where we have $\theta^* = \lambda \hat{\theta}_n + (1 - \lambda)\theta_0$ with $0 \leq \lambda \leq 1$. Therefore, we further obtain

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = -\left(\frac{\partial^2 \tilde{L}_n(\theta^*)}{\partial \theta \partial \theta'}\right)^{-1} \sqrt{n} \frac{\partial \tilde{L}_n(\theta_0)}{\partial \theta},$$

in which we have $-\left(\frac{\partial^2 \tilde{L}_n(\theta^*)}{\partial \theta \partial \theta'}\right) \rightarrow_p I(\theta_0) = -E_{\theta_0}\left[\frac{\partial^2}{\partial \theta \partial \theta'} l_t(\theta_0)\right]$ by Lemma A14 (a). In addition, $\sqrt{n} \frac{\partial \tilde{L}_n(\theta_0)}{\partial \theta}$ converges to $N(0, I(\theta_0))$ in distribution by our conclusions from Lemma A14 (b) and Lemma A15. In the end, after utilizing Slutsky theorem we conclude that $\sqrt{n}(\hat{\theta}_n - \theta_0)$ converges to $N(0, M_0^{-1})$ in distribution which claims our conclusion of Theorem 3. \square

Appendix A.5. Proof of Proposition 1

Proof. We denote V_n as $V_n = \{\theta \in \Theta \mid \mu \leq \mu_0 + \epsilon_n\}$, in which $\epsilon_n \sim n^{-\alpha}$, with $n^{-1/2} < \tau_n < \epsilon_n < n^{-1/\alpha_L}$. We obtain $\mu_0 + \epsilon_n \searrow \mu_0$.

We further define $\Theta_n^\delta = \{\theta \in V_n \mid \|\theta - \theta_0\| \geq \delta\}$, $\Theta_n^\mu = \{\theta \in V_n \mid \|\theta - \theta_0\| \geq \delta, \mu > \mu_0\}$, and $\Theta^\delta = \{\theta \in V_n \mid \|\theta - \theta_0\| \geq \delta, \mu \leq \mu_0\}$. We then have $\Theta_n^\delta = \Theta_n^\mu \cup \Theta^\delta$.

(I) First, we want to prove, for any $\delta > 0$,

$$P(\sup_{\Theta_n^\delta} \tilde{L}_n(\theta) \geq \tilde{L}_n(\theta_0)) \rightarrow 0, (n \rightarrow \infty).$$

By Lemmas A7 and A13, we have $\sup_{\Theta_n^\delta} |\tilde{L}_n(\theta) - L_n(\theta)| \rightarrow_p 0$ as $n \rightarrow \infty$. In addition, by Lemma A7, we further have $\sup_{\Theta_n^\mu} |L_n(\mu, \phi) - L_n(\mu_0, \phi)| \rightarrow_p 0$ as $n \rightarrow \infty$.

We then obtain

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

The last inequality follows from the fact $\epsilon_n \searrow 0$, then with probability going to 1, we have $\{\phi \mid \phi \in \Theta_n^\mu\} \subseteq \{\phi \mid \phi \in \Theta^{\delta/2}\}$. Following similar proof procedures given in Lemmas A13 and A14, we are able to see that $\tilde{L}_n(\theta_0) = L_n(\theta_0) + o_p(1) \rightarrow_p E_{\theta_0}[l_1(\theta_0)]$ holds. Then the rest proof of the first part follows from the proof of Proposition 2 in [48].

For the second part, we define $\Theta_n^{\delta c} = \{\theta \in V_n \mid \|\theta - \theta_0\| < \delta\}$, $\Theta_n^{\mu c} = \{\theta \in V_n \mid \|\theta - \theta_0\| < \delta, \mu > \mu_0\}$, and $\Theta^{\delta c} = \{\theta \in V_n \mid \|\theta - \theta_0\| < \delta, \mu \leq \mu_0\}$. Note that we have $\Theta_n^{\delta c} = \Theta_n^{\mu c} \cup \Theta^{\delta c}$. Next we want to prove that there exists a $\delta^* > 0$ s.t.

(II) $P(\text{All Hessian matrices } \frac{\partial^2}{\partial \theta_i \partial \theta_j} \tilde{L}_n(\theta) \text{ over } \theta \in \Theta_n^{\delta^* c} \text{ is negative}) \rightarrow 1$, as n tends to infinity. According to our results given in Lemmas A7 and A13, we obtain

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

and

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

Note that there exists a function $g(x)$ s.t. $l_t(x, \theta) \leq g(x)$ for all the $\theta \in \Theta^{\delta c}$, with $E_{\theta_0}(g(x)) < \infty$. Therefore, by the properties of stationarity and ergodicity and the uniform law of large numbers, we have

$$\sup_{\Theta^{\delta c}} \left| \frac{\partial^2}{\partial \theta_i \partial \theta_j} L_n(\theta) - E_{\theta_0} \left[\frac{\partial^2}{\partial \theta_i \partial \theta_j} l_1(\theta) \right] \right| \rightarrow_p 0, (n \rightarrow \infty),$$

with $E_{\theta_0} \left[\frac{\partial^2}{\partial \theta_i \partial \theta_j} l_1(\theta_0) \right] = -M_0$, in which M_0 is positive definite by Lemma A4. Moreover, the function $E_{\theta_0} \left[\frac{\partial^2}{\partial \theta_i \partial \theta_j} l_1(\theta) \right]$ is continuous, so there exists a $\delta^* > 0$ s.t. $E_{\theta_0} \left[\frac{\partial^2}{\partial \theta_i \partial \theta_j} l_1(\theta) \right]$ is negative definite for all $\theta \in \Theta^{\delta^* c}$. Combining the demonstrations given above, we finish our proof of (II).

Utilizing our conclusion from (I), we obtain that the global maximizer of $\tilde{L}_n(\theta)$ over V_n is located in $\Theta_n^{\delta^* c}$. It is known from Theorem 2 that with probability going to 1, there exists a sequence of $\hat{\theta}_n$ of local maximizer of $\tilde{L}_n(\theta)$ s.t. $\|\hat{\theta}_n - \theta_0\| \leq \tau_n$, where $\tau_n = O_p(n^{-r})$, and $1/\alpha_L < \alpha < r < 1/2$. So we have $P(\hat{\theta}_n \in \Theta_n^{\delta^* c}) \rightarrow 1$. Combining with our proof of (II) and using the conclusion of Theorem 2.6 in [75], we finally claim the conclusion of this proposition. \square

In the next subsection we will illustrate expressions of the first order as well as the second order derivatives of our likelihood function.

Appendix A.6. First and the Second Order Partial Derivatives of $l_t(\theta)$

In this section we denote $\Phi = (\gamma_0, \gamma_1, \gamma_2, \gamma_3)$ and similarly, we set $\Psi = (\beta_0, \beta_1, \beta_2, \beta_3)$ and the likelihood function as follows:

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

The first order partial derivatives of $l_t(\theta)$ are give by:

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

The second order partial derivatives of $l_t(\theta)$ are given by:

$$\begin{aligned} \frac{\partial^2 l_t(\theta)}{\partial \mu^2} &= -\frac{\alpha_t(\alpha_t - 1)}{\sigma_t^2} \left(\frac{Q_t - \mu}{\sigma_t} \right)^{\alpha_t - 2} - \frac{\alpha_t - 1}{(Q_t - \mu)^2}, \\ \frac{\partial^2 l_t(\theta)}{\partial \mu \partial \Phi} &= \left[-\frac{1}{Q_t - \mu} + \frac{1}{\sigma_t} \left(\frac{Q_t - \mu}{\sigma_t} \right)^{\alpha_t - 1} + \frac{\alpha_t}{\sigma_t} \left(\frac{Q_t - \mu}{\sigma_t} \right)^{\alpha_t - 1} \log \left(\frac{Q_t - \mu}{\sigma_t} \right) \right] \frac{\partial \alpha_t}{\partial \Phi}, \\ \frac{\partial^2 l_t(\theta)}{\partial \mu \partial \Psi} &= \left[-\frac{\alpha_t^2}{\sigma_t^2} \left(\frac{Q_t - \mu}{\sigma_t} \right)^{\alpha_t - 1} \right] \frac{\partial \sigma_t}{\partial \Psi}, \\ \frac{\partial^2 l_t(\theta)}{\partial \Phi \partial \Psi} &= \left[-\frac{1}{\sigma_t} + \frac{\alpha_t}{\sigma_t} \left(\frac{Q_t - \mu}{\sigma_t} \right)^{\alpha_t} \log \left(\frac{Q_t - \mu}{\sigma_t} \right) + \frac{1}{\sigma_t} \left(\frac{Q_t - \mu}{\sigma_t} \right)^{\alpha_t} \right] \frac{\partial \sigma_t}{\partial \Psi} \frac{\partial \alpha_t}{\partial \Phi}, \\ \frac{\partial^2 l_t(\theta)}{\partial \Phi_i \partial \Phi_j} &= \left[\frac{\alpha_t}{\sigma_t^2} - \frac{\alpha_t(\alpha_t + 1)}{\sigma_t^2} \left(\frac{Q_t - \mu}{\sigma_t} \right)^{\alpha_t} \right] \frac{\partial \alpha_t}{\partial \Phi_i} \frac{\partial \alpha_t}{\partial \Phi_j} + \left[-\frac{\alpha_t}{\sigma_t} + \frac{\alpha_t}{\sigma_t} \left(\frac{Q_t - \mu}{\sigma_t} \right)^{\alpha_t} \right] \frac{\partial^2 \alpha_t}{\partial \Phi_i \partial \Phi_j}, \\ \frac{\partial^2 l_t(\theta)}{\partial \Psi_i \partial \Psi_j} &= \left[-\frac{1}{\alpha_t^2} - \left(\frac{Q_t - \mu}{\sigma_t} \right)^{\alpha_t} \left(\log \left(\frac{Q_t - \mu}{\sigma_t} \right) \right)^2 \right] \frac{\partial \sigma_t}{\partial \Psi_i} \frac{\partial \sigma_t}{\partial \Psi_j} \\ &\quad + \left[\frac{1}{\alpha_t} + \log \left(\frac{Q_t - \mu}{\sigma_t} \right) - \left(\frac{Q_t - \mu}{\sigma_t} \right)^{\alpha_t} \log \left(\frac{Q_t - \mu}{\sigma_t} \right) \right] \frac{\partial^2 \sigma_t}{\partial \Psi_i \partial \Psi_j}. \end{aligned}$$

Appendix B. Algorithms Computation Details

The computational time for representative algorithms are listed in Table A1.

Table A1. Computational time.

Algorithms	Algorithms for Example 1/SC1	Algorithms for Example 1/SC2	Algorithms for Example 2/SC1	Algorithms for Example 2/SC2	Algorithms for Table 4	Algorithms for Table 5
Computational time (hour)	97.48	269.45	759.66	3030.24	0.29	5.30

All the algorithms are running on a server with Xeon 2.4GHz and 16GB of memory. Especially, algorithms for Tables 4 and 5 are running using parallel computing of 16 processes.

References

- Smith, R.L. Extreme value analysis of environmental time series: An application to trend detection in ground-level ozone. *Stat. Sci.* **1989**, *4*, 367–377. [\[CrossRef\]](#)
- Yiou, P.; Goubanova, K.; Li, Z.X.; Nogaj, M. Weather regime dependence of extreme value statistics for summer temperature and precipitation. *Nonlinear Process. Geophys.* **2008**, *15*, 365–378. [\[CrossRef\]](#)
- Smith, R.L.; Grady, A.M.; Hegerl, G.C. Extreme precipitation trends over the continental United States. In Proceedings of the 15th Aha Hulikoa Hawaiian Winter Workshop, Honolulu, HI, USA, 5 November 2007.
- Reich, B.J.; Shaby, B.A. A hierarchical max-stable spatial model for extreme precipitation. *Ann. Appl. Stat.* **2012**, *6*, 1430–1451. [\[CrossRef\]](#) [\[PubMed\]](#)
- Cooley, D.; Nychka, D.; Naveau, P. Bayesian spatial modeling of extreme precipitation return levels. *J. Am. Stat. Assoc.* **2017**, *102*, 824–840. [\[CrossRef\]](#)
- Zhang, Z.; Zhang, C.; Cui, Q. Random threshold driven tail dependence measures with application to precipitation data analysis. *Stat. Sin.* **2017**, *27*, 685–709.
- Naveau, P.; Nogaj, M.; Ammann, C.; Yiou, P.; Cooley, D.; Jomelli, V. Statistical methods for the analysis of climate extremes. *C. R. Geosci.* **2005**, *337*, 1013–1022. [\[CrossRef\]](#)
- Gilleland, E.; Brown, B.G.; Ammann, C.M. Spatial extreme value analysis to project extremes of large-scale indicators for severe weather. *Environmetrics* **2013**, *24*, 418–432. [\[CrossRef\]](#)
- Kempton, G.; Wild, K. Extreme weather is the new normal. *Electr. Light Power* **2013**, *91*, 20–22.
- Mannshardt, E.; Benedict, K.; Jenkins, S.; Keating, M.; Mintz, D. Analysis of short-term ozone and PM2.5 measurements: Characteristics and relationships for air sensor messaging. *J. Air Waste Manag.* **2017**, *67*, 462–474. [\[CrossRef\]](#)
- Mannshardt, E.; Naess, L. Air quality in the USA. *Significance* **2018**, *15*, 24–27. [\[CrossRef\]](#)
- Xu, H.; Bi, X.H.; Zheng, W.W.; Wu, J.H.; Feng, Y.C. Particulate matter mass and chemical component concentrations over four Chinese cities along the western Pacific coast. *Environ. Sci. Pollut. Res. Int.* **2015**, *22*, 1940–1953. [\[CrossRef\]](#) [\[PubMed\]](#)
- Chang, S.Y. The Characteristics of PM2.5 and Its Chemical Compositions between Different Prevailing Wind Patterns in Guangzhou. *Aerosol Air. Qual. Res.* **2013**, *13*, 1373–1383. [\[CrossRef\]](#)
- Li, J.; Song, Y.; Mao, Y.; Mao, Z.; Wu, Y.; Li, M.; Huang, X.; He, Q.; Hu, M. Chemical characteristics and source apportionment of PM2.5 during the harvest season in eastern China's agricultural regions. *Atmos. Environ.* **2014**, *92*, 442–448. [\[CrossRef\]](#)
- Yang, F.; Tan, J.; Zhao, Q.; Du, Z.; He, K.; Ma, Y.; Duan, F.; Chen, G.; Zhao, Q. Characteristics of PM2.5 speciation in representative megacities and across China. *Atmos. Chem. Phys.* **2011**, *11*, 1025–1051. [\[CrossRef\]](#)
- Huang, R.J.; Zhang, Y.; Bozzetti, C.; Ho, K.F.; Cao, J.J.; Han, Y.; Daellenbach, K.R.; Slowik, J.G.; Platt, S.M.; Canonaco, F. High secondary aerosol contribution to particulate pollution during haze events in China. *Nature* **2014**, *514*, 218–222. [\[CrossRef\]](#)
- Zhang, R.; Wang, G.; Song, G.; Zamora, M.L.; Qi, Y.; Yun, L.; Wang, W.; Min, H.; Yuan, W. Formation of Urban Fine Particulate Matter. *Chem. Rev.* **2015**, *115*, 3803–3855. [\[CrossRef\]](#)
- Wang, X.; Chen, J.; Cheng, T.; Zhang, R.; Wang, X. Particle number concentration, size distribution and chemical composition during haze and photochemical smog episodes in Shanghai. *J. Environ. Sci.* **2014**, *26*, 1894–1902. [\[CrossRef\]](#)
- Liu, J.; Li, J.; Zhang, Y.; Liu, D.; Ding, P.; Shen, C.; Shen, K.; He, Q.; Ding, X.; Wang, X. Source Apportionment Using Radiocarbon and Organic Tracers for PM2.5 Carbonaceous Aerosols in Guangzhou, South China: Contrasting Local- and Regional-Scale Haze Events. *Environ. Sci. Technol.* **2014**, *48*, 12002–12011. [\[CrossRef\]](#)

20. Huang, K.; Zhuang, G.; Wang, Q.; Fu, J.S.; Lin, Y.; Liu, T.; Han, L.; Deng, C. Extreme haze pollution in Beijing during January 2013: Chemical characteristics, formation mechanism and role of fog processing. *Atmos. Chem. Phys.* **2014**, *14*, 479–486. [[CrossRef](#)]
21. Wang, L.; Zhang, N.; Liu, Z.; Sun, Y.; Ji, D.; Wang, Y. The Influence of Climate Factors, Meteorological Conditions, and Boundary-Layer Structure on Severe Haze Pollution in the Beijing-Tianjin-Hebei Region during January 2013. *Adv. Meteorol.* **2015**, *2014*, 1–14. [[CrossRef](#)]
22. Zhang, X.; Xu, X.; Ding, Y.; Liu, Y.; Zhang, H.; Wang, Y.; Zhong, J. The impact of meteorological changes from 2013 to 2017 on PM_{2.5} mass reduction in key regions in China. *Sci. China Earth Sci.* **2019**, *62*, 1885–1902. [[CrossRef](#)]
23. Chen, J. Impact of Relative Humidity and Water Soluble Constituents of PM_{2.5} on Visibility Impairment in Beijing, China. *Aerosol Air. Qual. Res.* **2014**, *14*, 260–268. [[CrossRef](#)]
24. Liang, X.; Zou, T.; Guo, B.; Li, S.; Zhang, H.; Zhang, S.; Huang, H.; Chen, S.X. Assessing Beijing's PM_{2.5} pollution: Severity, weather impact, APEC and winter heating. *Proc. R. Soc. A* **2015**, *471*. [[CrossRef](#)]
25. Requia, W.J.; Jhun, I.; Coull, B.A.; Koutrakis, P. Climate impact on ambient PM_{2.5} elemental concentration in the United States: A trend analysis over the last 30 years. *Environ. Int.* **2019**, *131*, 104888. [[CrossRef](#)]
26. Lin, G.; Fu, J.; Jiang, D.; Hu, W.; Dong, D.; Huang, Y.; Zhao, M. Spatio-Temporal Variation of PM_{2.5} Concentrations and Their Relationship with Geographic and Socioeconomic Factors in China. *Int. J. Environ. Res. Public Health* **2013**, *11*, 173–186. [[CrossRef](#)]
27. Donkelaar, A.V.; Martin, R.V.; Brauer, M.; Kahn, R.; Levy, R.; Verduzco, C.; Villeneuve, P.J. Global Estimates of Ambient Fine Particulate Matter Concentrations from Satellite-Based Aerosol Optical Depth: Development and Application. *Environ. Health Perspect.* **2010**, *118*, 847–855. [[CrossRef](#)]
28. Donkelaar, A.V.; Martin, R.V.; Brauer, M.; Boys, B.L. Use of Satellite Observations for Long-Term Exposure Assessment of Global Concentrations of Fine Particulate Matter. *Environ. Health Perspect.* **2015**, *123*, 135–143. [[CrossRef](#)]
29. Cao, C.; Jiang, W.; Wang, B.; Fang, J.; Lang, J.; Tian, G.; Jiang, J.; Zhu, T.F. Inhalable microorganisms in Beijing's PM_{2.5} and PM₁₀ pollutants during a severe smog event. *Environ. Sci. Technol.* **2014**, *48*, 1499–1507. [[CrossRef](#)]
30. Guo, S.; Hu, M.; Zamora, M.L.; Peng, J.; Shang, D.; Zheng, J.; Du, Z.; Wu, Z.; Shao, M.; Zeng, L. Elucidating severe urban haze formation in China. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 17373–17378. [[CrossRef](#)]
31. He, H.; Tie, X.; Zhang, Q.; Liu, X.; Gao, Q.; Li, X.; Gao, Y. Analysis of the causes of heavy aerosol pollution in Beijing, China: A case study with the WRF-Chem model. *Particulology* **2015**, *20*, 32–40. [[CrossRef](#)]
32. Uno, I.; Sugimoto, N.; Shimizu, A.; Yumimoto, K.; Hara, Y.; Wang, Z. Record Heavy PM_{2.5} Air Pollution over China in January 2013: Vertical and Horizontal Dimensions. *Sci. Online Lett. Atmos. Sola* **2014**, *10*, 136–140. [[CrossRef](#)]
33. Sun, Y.; Jiang, Q.; Wang, Z.; Fu, P.; Li, J.; Yang, T.; Yin, Y. Investigation of the sources and evolution processes of severe haze pollution in Beijing in January 2013. *J. Geophys. Res. Atmos.* **2014**, *119*, 4380–4398. [[CrossRef](#)]
34. Zhao, H.; Wang, F.; Niu, C.; Wang, H.; Zhang, X. Red warning for air pollution in China: Exploring residents' perceptions of the first two red warnings in Beijing. *Environ. Res.* **2018**, *161*, 540–545. [[CrossRef](#)] [[PubMed](#)]
35. Deng, L.; Zhang, Z. Assessing the features of extreme smog in China and the differentiated treatment strategy. *Proc. R. Soc. A* **2018**, *474*. [[CrossRef](#)]
36. Wu, H. *Breathing in Delhi Air Equivalent to Smoking 44 Cigarettes a Day*; CNN: New Delhi, India, 2017. Available online: <https://www.cnn.com/2017/11/10/health/delhi-pollution-equivalent-cigarettes-a-day/index.html> (accessed on 10 May 2020).
37. Bell, J.E.; Brown, C.L.; Conlon, K.; Herring, S.; Kunkel, K.E.; Lawrimore, J.; Luber, G.; Schreck, C.; Smith, A.; Uejio, C. Changes in extreme events and the potential impacts on human health. *J. Air. Waste. Manag.* **2018**, *68*, 265–287. [[CrossRef](#)]
38. Chen, S.M.; He, L.Y. Welfare loss of China's air pollution: How to make personal vehicle transportation policy. *China Econ. Rev.* **2014**, *31*, 106–118. [[CrossRef](#)]
39. Pearson, J.F.; Bachireddy, C.; Shyamprasad, S.; Goldfine, A.B.; Brownstein, J.S. Association Between Fine Particulate Matter and Diabetes Prevalence in the U.S. *Diabetes Care* **2010**, *33*, 2196–2201. [[CrossRef](#)]
40. Yang, G.H.; Zhong, N.S. Effect on health from smoking and use of solid fuel in China. *Lancet* **2008**, *372*, 1445–1446. [[CrossRef](#)]
41. Watts, J. China: The air pollution capital of the world. *Lancet* **2005**, *366*, 1761–1762. [[CrossRef](#)]

42. Fang, X.; Fang, B.; Wang, C.; Xia, T.; Bottai, M.; Fang, F.; Cao, Y. Relationship between fine particulate matter, weather condition and daily non-accidental mortality in Shanghai, China: A Bayesian approach. *PLoS ONE* **2017**, *12*, e0187933. [CrossRef]
43. Liu, J.; Han, Y.; Tang, X.; Zhu, J.; Zhu, T. Estimating adult mortality attributable to PM_{2.5} exposure in China with assimilated PM_{2.5} concentrations based on a ground monitoring network. *Sci. Total Environ.* **2016**, *568*, 1253–1262. [CrossRef] [PubMed]
44. Huo, H.; Zhang, Q.; Guan, D.; Su, X.; Zhao, H.; He, K. Examining air pollution in China using production- and consumption-based emissions accounting approaches. *Environ. Sci. Technol.* **2014**, *48*, 14139–14147. [CrossRef] [PubMed]
45. Zhao, H.Y.; Zhang, Q.; Davis, S.J.; Guan, D.; Liu, Z.; Huo, H.; Lin, J.T.; Liu, W.D.; He, K.B. Assessment of China's virtual air pollution transport embodied in trade by a consumption-based emission inventory. *Atmos. Chem. Phys.* **2015**, *15*, 6815. [CrossRef]
46. Muller, N.Z.; Mendelsohn, R.; Nordhaus, W. Environmental Accounting for Pollution in the United States Economy. *Am. Econ. Rev.* **2011**, *101*, 1649–1675. [CrossRef]
47. Lee, J.H. The Sociological Analysis on the Smog of China: The Perspective of Complex Risk Society. *J. North-East Asian Cult.* **2014**, *1*, 211–225. [CrossRef]
48. Dombry, C. Existence and consistency of the maximum likelihood estimators for the extreme value index within the block maxima framework. *Bernoulli* **2015**, *21*, 420–436. [CrossRef]
49. Davison, A.C.; Huser, R.; Thibaud, E. Spatial extremes. In *Handbook of Environmental and Ecological Statistics*; Gelfand, A.E., Fuentes, M., Smith, R.L., Eds.; CRC Press: Boca Raton, FL, USA, 2018.
50. Huser, R.G.; Wadsworth, J.L. Modeling spatial processes with unknown extremal dependence class. *J. Am. Stat. Assoc.* **2018**. [CrossRef]
51. Zhao, Z.; Zhang, Z.; Chen, R. Modeling maxima with autoregressive conditional Fréchet model. *J. Econ.* **2018**, *207*, 325–351. [CrossRef]
52. Kunkel, K.E.; Palecki, M.A.; Hubbard, K.G.; Robinson, D.A.; Redmond, K.T.; Easterling, D.R. Trend identification in twentieth-century U.S. snowfall: The challenges. *J. Atmos. Ocean. Technol.* **2007**, *24*, 64–73. [CrossRef]
53. Guo, R.; Zhang, C.; Zhang, Z. Maximum independent component analysis with application to EEG data. *Stat. Sci.* **2020**, *35*, 145–157. [CrossRef]
54. Gavronski, P.G.; Ziegelmann, F.A. Measuring Systemic Risk via GAS models and Extreme Value Theory: Revisiting the 2007 Financial Crisis. *Financ. Res. Lett.* **2020**. [CrossRef]
55. U.S. EPA. Revised Air Quality Standards for Particle Pollution and Updates to the Air Quality Index. 2012. Available online: https://www.epa.gov/sites/production/files/2016-04/documents/overview_factsheet.pdf (accessed on 10 May 2020).
56. Chen, Y.; Ebenstein, A.; Greenstone, M.; Li, H. Evidence on the impact of sustained exposure to air pollution on life expectancy from China's Huai River policy. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 12936–12941. [CrossRef]
57. Wang, Y.; Jiang, H.; Zhang, S.; Xu, J.; Lu, X.; Jin, J.; Wang, C. Estimating and source analysis of surface pm_{2.5} concentration in the beijing-tianjin-hebei region based on modis data and air trajectories. *Int. J. Remote Sens.* **2016**, *37*, 4799–4817. [CrossRef]
58. Cheng, Y.; He, K.; Du, Z.; Zheng, M.; Duan, F.; Ma, Y. Humidity plays an important role in the PM_{2.5} pollution in Beijing. *Environ. Pollut.* **2015**, *197*, 68–75. [CrossRef] [PubMed]
59. Leadbetter, M.R.; Lindgren, G.; Rootzén, H. *Extremes and Related Properties of Random Sequences and Processes*; Springer Science & Business Media: Berlin, Germany, 1983.
60. Mao, G.; Zhang, Z. Stochastic tail index model for high frequency financial data with Bayesian analysis. *J. Econ.* **2018**, *205*, 470–487. [CrossRef]
61. Mudholkar, G.S.; Hutson, A.D. The exponentiated Weibull family: Some properties and a flood data application. *Commun. Stat. Theor. Methods* **1996**, *25*, 3059–3083. [CrossRef]
62. Hawkins, T.W.; Holland, L.A. Synoptic and local weather conditions associated with PM_{2.5} concentration in Carlisle, Pennsylvania. *Middle States Geogr.* **2010**, *43*, 72–84.
63. Chen, Z.; Xie, X.; Cai, J.; Chen, D.; Gao, B.; He, B.; Cheng, N.; Xu, B. Understanding meteorological influences on PM_{2.5} concentrations across China: A temporal and spatial perspective. *Atmos. Chem. Phys.* **2018**, *18*, 5343–5358. [CrossRef]

64. Smith, R. Maximum likelihood estimation in a class of nonregular cases. *Biometrika* **1985**, *72*, 67–90. [[CrossRef](#)]
65. Bollerslev, T. Generalized autoregressive conditional heteroskedasticity. *J. Econ.* **1986**, *31*, 307–327. [[CrossRef](#)]
66. De Gooijer, J.G. *Elements of Nonlinear Time Series Analysis and Forecasting*; Springer: Berlin, Germany, 2017.
67. Hastie, T.J.; Tibshirani, R.J. *Generalized Additive Models*; Chapman & Hall: Boca Raton, FL, USA, 1990.
68. Thomas, W.Y. *Vector Generalized Linear and Additive Models: With an Implementation in R*; Springer: New York, NY, USA, 2015.
69. Nieto, P.G.; Antón, J.Á.; Vilán, J.V.; García-Gonzalo, E. Air quality modeling in the Oviedo urban area (NW Spain) by using multivariate adaptive regression splines. *Environ. Sci. Pollut. Res.* **2015**, *22*, 6642–6659. [[CrossRef](#)]
70. Shahraiyini, H.T.; Shahsavani, D.; Sargazi, S.; Habibi Nokhandan, M. Evaluation of MARS for the spatial distribution modeling of carbon monoxide in an urban area. *Atmos. Pollut. Res.* **2015**, *6*, 581–588. [[CrossRef](#)]
71. Stasinopoulos, M.D.; Rigby, R.A.; Heller, G.Z.; Voudouris, V.; De Bastiani, F. *Flexible Regression and Smoothing Using GAMLSS in R*; Chapman and Hall: London, UK, 2017.
72. Chan, K.; Tong, H. A Note on Noisy Chaos. *J. R. Stat. Soc. B* **1994**, *56*, 301–311. [[CrossRef](#)]
73. Birkhoff, G.D. Proof of the ergodic theorem. *Proc. Nat. Acad. Sci. USA* **1931**, *17*, 656–660. [[CrossRef](#)] [[PubMed](#)]
74. Billingsley, P. The Lindeberg-Levy theorem for martingales. *Proc. Am. Math. Soc.* **1961**, *12*, 788–792.
75. Makelainen, T.; Schmidt, K.; Styan, G. On the existence and uniqueness of the maximum likelihood estimate of a vector-valued parameter in fixed-size samples. *Ann. Stat.* **1981**, *9*, 758–767. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).