

Article

Statistical Inference of Dynamic Conditional Generalized Pareto Distribution with Weather and Air Quality Factors

Chunli Huang ¹, Xu Zhao ^{1,*}, Weihu Cheng ¹, Qingqing Ji ^{2,3}, Qiao Duan ⁴ and Yufei Han ⁵

¹ Faculty of Science, Beijing University of Technology, Beijing 100124, China; huangcl@emails.bjut.edu.cn (C.H.); chengweihu@bjut.edu.cn (W.C.)

² University of Chinese Academy of Sciences, Beijing 100049, China; jiqingqing20b@ict.ac.cn

³ Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China

⁴ Faculty of Humanities and Social Sciences, Beijing University of Technology, Beijing 100124, China; duanqiao02@emails.bjut.edu.cn

⁵ School of Computer Science (National Pilot Software Engineering School), Beijing University of Posts and Telecommunications, Beijing 100876, China; hanyufei0630@bupt.edu.cn

* Correspondence: zhaox@bjut.edu.cn

Abstract: Air pollution is a major global problem, closely related to economic and social development and ecological environment construction. Air pollution data for most regions of China have a close correlation with time and seasons and are affected by multidimensional factors such as meteorology and air quality. In contrast with classical peaks-over-threshold modeling approaches, we use a deep learning technique and three new dynamic conditional generalized Pareto distribution (DCP) models with weather and air quality factors for fitting the time-dependence of the air pollutant concentration and make statistical inferences about their application in air quality analysis. Specifically, in the proposed three DCP models, a dynamic autoregressive exponential function mechanism is applied for the time-varying scale parameter and tail index of the conditional generalized Pareto distribution, and a sufficiently high threshold is chosen using two threshold selection procedures. The probabilistic properties of the DCP model and the statistical properties of the maximum likelihood estimation (MLE) are investigated, simulating and showing the stability and sensitivity of the MLE estimations. The three proposed models are applied to fit the PM_{2.5} time series in Beijing from 2015 to 2021. Real data are used to illustrate the advantages of the DCP, especially compared to the estimation volatility of GARCH and AIC or BIC criteria. The DCP model involving both the mixed weather and air quality factors performs better than the other two models with weather factors or air quality factors alone. Finally, a prediction model based on long short-term memory (LSTM) is used to predict PM_{2.5} concentration, achieving ideal results.

Keywords: generalized Pareto distribution; peaks over threshold; dynamic conditional autoregressive modeling; threshold selection; long short-term memory

MSC: 62P12



Citation: Huang, C.; Zhao, X.; Cheng, W.; Ji, Q.; Duan, Q.; Han, Y. Statistical Inference of Dynamic Conditional Generalized Pareto Distribution with Weather and Air Quality Factors. *Mathematics* **2022**, *10*, 1433. <https://doi.org/10.3390/math10091433>

Academic Editor: Gaorong Li

Received: 14 March 2022

Accepted: 21 April 2022

Published: 24 April 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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 (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Air pollution, closely related to economic and social development as well as ecological environment construction, is a global problem that destroys human living environments. In recent years, the Chinese government has attached great importance to the prevention and control of air pollution. The World Air Quality Report 2021 released by the Swiss company IQAir pointed out that air quality in China continued to improve in 2021. Compared to 2020, PM_{2.5} concentrations decreased in 66% of Chinese cities [1]. However, China still faces environmental challenges. Air pollution is mainly composed of harmful gases and particulate matter, which are released into the atmosphere by natural or human activities, and its concentration is far beyond the self-purification capacity of the atmosphere, resulting

in changes in the composition of the atmosphere, endangering human health and living environments. Smog, as a seriously harmful air pollutant, has received growing attention. Severe smog levels pose a huge threat to China's public health [2–5]. Short-term exposure to air pollution will cause cough, dyspnea, headache, fatigue and other phenomena, while long-term exposure to air pollution will lead to respiratory diseases, cardiovascular damage, nervous system damage and other diseases, and may even lead to birth defects and death [6–9]. Climate warming, sea-level rises, acid rain, the hole in the ozone layer and other particulate pollution directly highlight the environmental problems caused by air pollution, harming human survival and development. To improve air quality and human living environments, air pollution has become a key topic for researchers, and monitoring, assessment, prediction and prevention have become important research directions in the study of air pollution. Smog is also closely linked to China's economic development [10–13]. It is necessary to make full use of multidimensional big data and give full play to the advantages of statistics and artificial intelligence technology. By virtue of interdisciplinary development, researchers have been able to vigorously develop statistical modeling theory and deep learning technology for accurate prediction and effective control of urban air quality. As a result, a solid theoretical foundation and effective technical support can be provided for improving the capabilities and level of ecological and environmental conservation.

Fine particulate matter ($PM_{2.5}$) is the most common object in the studies of pollutant concentration, and the higher the concentration in the air, the more serious the air pollution [14–16]. From a statistical point of view, pollutant concentration prediction has become an important research direction in air pollution forecasting and prevention. The existing research on air pollutant concentration mainly focuses on the sources, concentration distributions, fluctuations, affecting factors, adverse effects on human health and so on. Quantitative prediction of pollutant concentration is the most common statistical method for dealing with air pollution problems, and multivariate regression, cluster analysis and principal component analysis are the most frequently used statistical models. Chen Songxi, an academician of the Chinese Academy of Sciences, applied non-parametric statistics to the national air pollution assessment and prevention research and proposed a method for adjusting spatio-temporal meteorological factors to remove the meteorological confounding effect in atmospheric environmental monitoring, providing a scientific method for accurately measuring pollutant discharge and evaluating air pollution control [17–22]. Wang et al. (2020) [23] established a spatio-temporal O_3 pollution land use regression (LUR) model suitable for large cities based on parametric, non-parametric and semi-parametric classical statistical algorithms combined with meteorological factors, with the ability to monitor O_3 concentration with high spatio-temporal resolution.

In order to improve the prediction accuracy, classical extreme value theory (EVT) has attracted more and more attention. Compared with fine weather, researchers are more concerned with observations of pollutant concentrations exceeding a certain high threshold. Pickands (1975) [24] pointed out that observations above a certain threshold can be approximated well by the generalized Pareto distribution (GPD). As a branch of EVT, GPD plays an important role in many fields. In the field of hydrometeorology, the GPD model is used to analyze and forecast natural phenomena such as floods, wind and rainfall [25,26]. In the financial field, stock yield is non-normal and thick-tailed, which can be well fitted by a GPD [27,28]. In the field of insurance, insurance losses are generally non-negative with a thick tail, and a GPD is usually used to predict the maximum loss [29,30].

Due to the strong time correlation of observations, a traditional model with fixed parameters cannot perfectly fit the time-series observations in reality. To solve this problem, many researchers have conducted in-depth research on the dynamic extreme value distribution model and the dynamic over-threshold GPD model. Using the autoregressive mechanism of the GARCH model, Zhao et al. (2018) [31] established an autoregressive conditional Fréchet model with time-dependent parameters (type II GEV) for the sequence of daily maximum stock returns. They solved the maximum likelihood estimation of the

model parameters and studied the large-sample properties of the parameters. Chavez-Demoulin et al. (2014) [32] applied a Bayesian method to update the time-varying GPD parameters for the UBS stock price, which was a non-parametric method applied to a POT–GPD model. Kelly and Jiang (2014) [33] built a dynamic tail model with POT–GPD for panel data and measured the tail risk of the S&P 500 index. Massacci (2017) [34] studied the time-dependent dynamic parameter estimation of a GPD through the score-based approach, in order to accurately estimate the tail index from U.S. size-sorted decile stock portfolios. Shen et al. (2020) [27] established an autoregressive conditional Pareto (ACP) distribution model via an exponential function. The maximum likelihood estimation of the parameter was given, and its properties were studied. Based on the parameter estimation, they employed the ACP model to the Dow Jones Industrial Average and the S&P 500 index. Deng et al. (2020) [35] applied a dynamic model to air quality management, taking time and meteorological factors into consideration and establishing a dynamic conditional autoregressive Weibull distribution model (type III GEV) via the maximum daily pollutant concentrations. The probabilistic properties of the autoregressive model were investigated in their study. As is well known, it is difficult to fit the $PM_{2.5}$ time series accurately, so the model selection and statistical inferences are the most important and common challenges for real data applications.

Threshold selection is a critical issue for fitting the autoregressive conditional generalized Pareto distribution model. In practice, the threshold should be chosen in advance. If the threshold is too large, the sample size of observations exceeding the threshold will be too small, which may increase the variance of the parameter estimation and affect the estimation effect. If the threshold is too small, the sample size can be increased, but the estimator is prone to bias. Choulakian and Stephens (2001) [36] transformed the threshold selection into the goodness-of-fit test of the model. Through the selection method, an appropriate threshold was chosen, allowing the exceedance to follow the GPD, and the threshold selection was carried out at the same time as testing the model. Bermudez et al. (2001) [37] used a Bayesian predictive approach to the peaks-over-threshold (POT) method, which can also be applied to small-sample situations. Bader et al. (2018) [38] proposed an automated threshold selection procedure based on a sequence of goodness-of-fit tests, and attained automatic threshold selection by applying stopping rules, which transform the results of ordered, sequentially tested hypotheses to control the false discovery rate. Yang et al. (2018) [39] developed an empirical threshold selection method based on the relationship between eigenvalues and thresholds. Schneider et al. (2021) [40] proposed selecting the threshold by minimizing the asymptotic mean squared error of the Hill estimator.

With the continuous development of artificial intelligence and machine learning, an increasing number of scholars have applied traditional machine learning methods to statistical prediction models in recent years and achieved good results in terms of accuracy and time efficiency. Boznar et al. (1993) [41] compared prediction results based on the three-layer neural network perceptron with the results generated by a traditional atmospheric diffusion model. Neagu et al. (2002) [42] used a fuzzy neural network model to predict the concentration of nitrogen oxide pollutants, achieving good results. Esfandani and Nematzadeh (2016) [43] proposed a prediction model for air quality in Tehran based on a feedback neural network. Amarpuri et al. (2019) [44] established a convolutional long short-term memory network to predict carbon dioxide emissions and achieved ideal results. An air pollution prediction model based on LSTM is a good choice for predicting $PM_{2.5}$ concentrations. García et al. (2020) [45] analyzed the concentrations of nitrogen dioxide (NO_2), nitrogen oxides (NO_x), particulate matter (PM_{10}) and toluene (C_7H_8) at eight sites in Madrid (Spain) through seven regression-based machine learning models and time-series models. Sánchez-Pérez et al. (2020) [46] established a complete spatio-temporal dispersion model for pollutants through a network simulation method, to obtain the concentrations of pollutants released at any time in a given space. Sayeed et al. (2021) [47] used a generalized deep convolutional neural network (CNN) model to predict air pollutants, which could predict the hourly pollutant concentration within 7 days with relatively high accuracy.

In this paper, dynamic autoregressive mechanisms are applied, and weather and air quality factors are also involved in our model. The framework of this paper follows that of [27]. The three main contributions of this paper are as follows. First, we construct a dynamic conditional generalized Pareto distribution (DCP) with both weather and air quality factors to fit the smog observations, considering the time-dependency of the scale parameter and the tail index of the GPD. Secondly, the threshold is chosen using a threshold selection program rather than by specifying a quantile. Thirdly, the PM_{2.5} time series are predicted by combining the DCP model with deep learning technology.

2. DCP Model

2.1. Conditional Distribution

The cumulative distribution function of a three-parameter GPD is defined as

$$G_{\mu,\xi,\sigma}(x) = \begin{cases} 1 - \left(1 + \frac{\xi}{\sigma}(x - \mu)\right)^{-\frac{1}{\xi}}, & \xi \neq 0 \\ 1 - \exp\left(-\frac{x-\mu}{\sigma}\right), & \xi = 0 \end{cases}, \tag{1}$$

where $\mu \in (-\infty, \infty)$ is the location parameter, $\xi \in (-\infty, \infty)$ is the shape parameter and $\sigma \in (0, \infty)$ the scale parameter. In Equation (1), $\mu \leq x < \infty$ when $\xi \geq 0$, $\mu < x < \mu - \sigma/\xi$ when $\xi < 0$. In particular, the GPD is an exponential distribution when $\xi = 0$. Additionally, the two-parameter GPD should be mentioned here for its extensive use, especially in parameter estimation. The classical two-parameter GPD(ξ, σ) is obtained by taking $\mu = 0$ in (1) [48]. The GPD has an important property that if X is a random variable (r.v.) distributed according to a GPD(ξ, σ), then the r.v. $Y = X - u | X > u$ has a GPD($\xi, \sigma + \xi u$) for a threshold u . This means that the shape parameter does not alter in the “excess over the threshold” operation.

Let $\{Q_t\}_{t=1}^n$ be the time sequence of the daily moving average concentration of PM_{2.5} at time t for smog occurrences, where n denotes the size of the observations. Let $F_t(q_t | \mathcal{F}_{t-1})$ be the conditional cumulative distribution of Q_t , where \mathcal{F}_{t-1} denotes the available information set until time $t - 1$. In practice, the underlying distribution $F_t(q_t | \mathcal{F}_{t-1})$ of the dataset is unknown. Based on the famous Pickands–Balkema–de Haan theorem [24,49], the standard practice is to employ GPD modeling for the tail region if the dataset under the POT framework of the original distribution is in the maximum domain of attraction. The obvious limitation is that the time characteristics of Q_t are totally ignored, which may result in the loss of some sample information and cause the statistical inference result to be inaccurate if the dataset depends strongly on time. To solve this problem, Massacci (2017) [34] and Shen et al. (2020) [27] proposed a dynamic GPD framework under the parameters (u, ζ_t, α_t) as follows:

$$G_t^u(q_t | \mathcal{F}_{t-1}) = 1 - \left(1 + \frac{q_t - u}{\alpha_t}\right)^{-\zeta_t}, \quad q_t > u > 0, \quad \zeta_t > 0, \quad \alpha_t > 0,$$

where the parameters α_t and ζ_t are time varying, and ζ_t is the tail index and u is the selected threshold. When the POT approach is employed, based on the Pickands–Balkema–de Haan theorem [24,49], the conditional distribution of the positive excess $F_t^u(q_t | \mathcal{F}_{t-1})$ can be approximated by the dynamic GPD if the distribution satisfies the condition of the theorem. That is,

$$\lim_{u \rightarrow +\infty} \sup_{u \leq q_t < +\infty} |F_t^u(q_t | \mathcal{F}_{t-1}) - G_t^u(q_t | \mathcal{F}_{t-1})| = 0,$$

where

$$\begin{aligned} F_t^u(q_t | \mathcal{F}_{t-1}) &= P(u < Q_t \leq q_t | Q_t > u; \mathcal{F}_{t-1}) \\ &= \frac{F_t(q_t | \mathcal{F}_{t-1}) - F_t(u | \mathcal{F}_{t-1})}{1 - F_t(u | \mathcal{F}_{t-1})}, 0 < u \leq q_t, \end{aligned}$$

$F_t^u(q_t|\mathcal{F}_{t-1})$ is assumed to be the dynamic GPD, so we can rewrite $F_t(q_t|\mathcal{F}_{t-1})$ as

$$\begin{aligned} F_t(q_t|\mathcal{F}_{t-1}) &= P(Q_t \leq q_t|\mathcal{F}_{t-1}) \\ &= [1 - F_t(u|\mathcal{F}_{t-1})]F_t^u(q_t|\mathcal{F}_{t-1}) + F_t(u|\mathcal{F}_{t-1}) \\ &= P_t G_t^u(q_t|\mathcal{F}_{t-1}) + 1 - P_t, \end{aligned} \tag{2}$$

where $P_t = P(Q_t > u|\mathcal{F}_{t-1})$.

For a given threshold value u , we focus on the exceedance $Q_t > u$ and define $Y_t = \max(Q_t - u, 0)$. Based on Equation (2), the corresponding conditional cumulative distribution function $P(Y_t \leq y_t|\mathcal{F}_{t-1}) = H_t(y_t|\mathcal{F}_{t-1})$ of Y_t is as follows [27,34]:

$$\begin{aligned} H_t(y_t|\mathcal{F}_{t-1}) &= \mathbb{I}(y_t = 0)(1 - P_t) + \mathbb{I}(y_t > 0)F_t(y_t + u|\mathcal{F}_{t-1}) \\ &= \mathbb{I}(y_t = 0)(1 - P_t) + \mathbb{I}(y_t > 0)\left[P_t G_t^u(y_t + u|\mathcal{F}_{t-1}) + 1 - P_t\right] \\ &= \mathbb{I}(y_t = 0)(1 - P_t) + \mathbb{I}(y_t > 0)\left[1 - P_t\left(1 + \frac{y_t}{\alpha_t}\right)^{-\zeta_t}\right], \end{aligned}$$

where $\mathbb{I}(\cdot)$ is an indicator function. P_t is approximated as a power law multiplied by a time-varying function slowly varying at infinity. Massacci (2017) [34] parameterized the function and obtained the following formula for P_t :

$$P_t = (1 + u)^{-\zeta_t}. \tag{3}$$

From Equation (3), the cumulative distribution function $H_t(Y_t|\mathcal{F}_{t-1})$ of Y_t becomes

$$H_t(y_t|\mathcal{F}_{t-1}) = \mathbb{I}(y_t = 0)[1 - (1 + u)^{-\zeta_t}] + \mathbb{I}(y_t > 0)\left[1 - (1 + u)^{-\zeta_t}\left(1 + \frac{y_t}{\alpha_t}\right)^{-\zeta_t}\right], u > 0, \alpha_t > 0, \zeta_t > 0. \tag{4}$$

By solving the inverse function of Y_t , we obtain

$$Y_t = \alpha_t \mathbb{I}(P_t > Z_t) \left[\left(\frac{P_t}{Z_t} \right)^{\frac{1}{\zeta_t}} - 1 \right], \tag{5}$$

where Z_t follows a uniform distribution in (0,1) and P_t is as given in (3). Equation (4) contains three distributions from EVT: the POT framework for the GPD, the power law for the conditional probability of Y_t greater than 0 and the uniform distribution of Z_t .

2.2. Model Specification

Shen et al. (2022) [27] assumed that $\zeta_t/\alpha_t = b$ for simplicity, and the form of α_t was modeled as follows:

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

where $0 \leq \beta_1 \leq 1, \beta_2 > 0, \beta_3 > 0$.

In this paper, after threshold selection for the dynamic conditional generalized Pareto (DCP) model, we concentrate on the autoregression of both α_t and ζ_t , which are the critical parameters reflecting the tail behavior. We impose a dynamic structure on the time-dependent parameters (α_t, ζ_t) and consider weather and air quality factors.

Specifically, the DCP model with weather and air quality factors assumes the form

$$\log \alpha_t = \beta_0 + \beta_1 \log \alpha_{t-1} + \eta_1(Q_{t-1}, T_{t-1}, H_{t-1}, W_{t-1}, SO2_{t-1}, NO2_{t-1}, CO_{t-1}), \tag{6}$$

$$\log \zeta_t = \gamma_0 + \gamma_1 \log \zeta_{t-1} + \eta_2(Q_{t-1}, T_{t-1}, H_{t-1}, W_{t-1}, SO2_{t-1}, NO2_{t-1}, CO_{t-1}), \tag{7}$$

where $\beta_1, \gamma_1 \in (0, 1), \beta_0, \gamma_0 \in R, \eta_1(\cdot)$ and $\eta_2(\cdot)$ are the observation-driven functions for $\log \alpha_t$ and $\log \zeta_t$. T_t, H_t, W_t denote daily average temperature, average relative humidity and average wind speed, and $SO2_t, NO2_t, CO_t$ denote daily moving average concentrations

of sulfur dioxide, nitrogen dioxide and carbon monoxide on day t , respectively. These three weather factors and three air quality factors are commonly considered in studies on smog. Other weather and air quality factors could also be considered, but the model complexity may be increased, and the effect may be weakened by adding too many factors.

We use continuous monotonic exponential functions in $\eta_1(\cdot)$ and $\eta_2(\cdot)$, as in other studies in the literature [27,31], for simplicity, flexibility and easy interpretation. From Equation (5), there exists a positive association between α_t and Y_t , while ζ_t and Y_t are negatively correlated. An increasing $\eta_1(\cdot)$ and a decreasing $\eta_2(\cdot)$ ensure that a large Y_t is followed by a large α_t and small ζ_t , so we choose the autoregressive process with weather factors as

$$\log \alpha_t = \beta_0 + \beta_1 \log \alpha_{t-1} - \beta_2 \exp(-\beta_3 Y_{t-1} + \beta_4 T_{t-1} + \beta_5 W_{t-1} + \beta_6 H_{t-1}), \quad (8)$$

$$\log \zeta_t = \gamma_0 + \gamma_1 \log \zeta_{t-1} + \gamma_2 \exp(-\gamma_3 Y_{t-1} + \gamma_4 T_{t-1} + \gamma_5 W_{t-1} + \gamma_6 H_{t-1}), \quad (9)$$

where Y_t is given in (5), and $\{Z_t\} \stackrel{i.i.d}{\sim} U(0,1)$, $\beta_i, \gamma_i \in R, i = 0,4,5,6, 0 \leq \beta_1 \neq \gamma_1 < 1, \beta_j, \gamma_j > 0, j = 2,3$. Equations (8) and (9) mean that an extreme event observed at time $t - 1$ (large Y_{t-1}) causes the distribution of Y_t to have a larger scale (large α_t) and a heavier tail (small ζ_t). That is why the exceedances tend to occur at around the same period in our examples.

In addition, α_t and ζ_t of the DCP model with air quality factors and the model with mixed weather and air quality factors are expressed in the same way as in Equations (10)–(13), respectively.

$$\log \alpha_t = \beta_0 + \beta_1 \log \alpha_{t-1} - \beta_2 \exp(-\beta_3 Y_{t-1} - \beta_4 SO2_{t-1} - \beta_5 NO2_{t-1} - \beta_6 CO_{t-1}), \quad (10)$$

$$\log \zeta_t = \gamma_0 + \gamma_1 \log \zeta_{t-1} + \gamma_2 \exp(-\gamma_3 Y_{t-1} - \gamma_4 SO2_{t-1} - \gamma_5 NO2_{t-1} - \gamma_6 CO_{t-1}), \quad (11)$$

where $\beta_i, \gamma_i \in R, i = 0,4,5,6, 0 \leq \beta_1 \neq \gamma_1 < 1, \beta_j, \gamma_j > 0, j = 2,3$.

$$\log \alpha_t = \beta_0 + \beta_1 \log \alpha_{t-1} - \beta_2 \exp(-\beta_3 Y_{t-1} - \beta_4 SO2_{t-1} - \beta_5 CO_{t-1} + \beta_6 W_{t-1} + \beta_7 H_{t-1}), \quad (12)$$

$$\log \zeta_t = \gamma_0 + \gamma_1 \log \zeta_{t-1} + \gamma_2 \exp(-\gamma_3 Y_{t-1} - \gamma_4 SO2_{t-1} - \gamma_5 CO_{t-1} + \gamma_6 W_{t-1} + \gamma_7 H_{t-1}), \quad (13)$$

where $\beta_i, \gamma_i \in R, i = 0,4,5,6,7, 0 \leq \beta_1 \neq \gamma_1 < 1, \beta_j, \gamma_j > 0, j = 2,3$.

Specific details of model applications will be discussed in Sections 5 and 6.

3. Estimation and Properties

In this section, we consider the maximum likelihood estimation method for estimating the parameters in the DCP models.

3.1. Maximum Likelihood Estimation

Taking the weather factors model as an example, we denote $\Theta_s = \{\theta = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \gamma_0, \gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5, \gamma_6) | 0 \leq \beta_1 \neq \gamma_1 < 1, \beta_2 > 0, \beta_3 > 0, \gamma_2 > 0, \gamma_3 > 0, \beta_i, \gamma_i \in R, i = 0,4,5,6\}$ as the parameter space in the DCP model with weather factors. The conditional probability function of Y_t can be obtained according to Equation (4) as

$$h_t(Y_t | \mathcal{F}_{t-1}) = \mathbb{I}(Y_t = 0)[1 - (1 + u)^{-\zeta_t}] + \mathbb{I}(Y_t > 0)\left[\frac{\zeta_t}{\alpha_t}(1 + u)^{-\zeta_t}\left(1 + \frac{Y_t}{\alpha_t}\right)^{-\zeta_t - 1}\right],$$

where $u > 0, \alpha_t > 0, \zeta_t > 0$.

The corresponding log-likelihood function with respect to the parameter θ is

$$L_n(\theta) = \frac{1}{n} \sum_{t=1}^n \left\{ \mathbb{I}(Y_t = 0) \log[1 - (1 + u)^{-\zeta_t}] + \mathbb{I}(Y_t > 0) \left[\log \zeta_t - \log \alpha_t - \zeta_t \log(1 + u) - (\zeta_t + 1) \log\left(1 + \frac{Y_t}{\alpha_t}\right) \right] \right\}. \quad (14)$$

Next, the process of maximum likelihood estimation is briefly introduced. With reference to [38,50], we adopt two threshold selection methods. The details are given in Section 5. After obtaining a sufficient high threshold u , we can obtain all Y_t from $Y_t = \max(Q_t - u, 0)$. We choose $\log \alpha_1$ as $\frac{\beta_0 - \beta_2/2}{1 - \beta_1}$ and $\log \zeta_1$ as $\frac{\gamma_0 + \gamma_2/2}{1 - \gamma_1}$, which lies in the middle of $G = \left(\frac{\beta_0 - \beta_2}{1 - \beta_1}, \frac{\beta_0}{1 - \beta_1}\right) \times \left(\frac{\gamma_0}{1 - \gamma_1}, \frac{\gamma_0 + \gamma_2}{1 - \gamma_1}\right)$, and obtain all α_t and ζ_t with the DCP model, using $Y_t, \log \alpha_1, \log \zeta_1$ and θ . Finally, we calculate the likelihood function using Equation (14) and obtain the MLE estimator of θ . The details are shown in Section 6.

3.2. Statistical Properties

The dynamic evolution Equations (6) and (7) without weather and air quality factors can be rewritten as

$$\log \alpha_t = \beta_0 + \beta_1 \log \alpha_{t-1} - \beta_2 \exp\{-\beta_3 \alpha_{t-1} \mathbb{I}(P_{t-1} > Z_{t-1}) [(P_{t-1}/Z_{t-1})^{1/\zeta_{t-1}} - 1]\}, \tag{15}$$

$$\log \zeta_t = \gamma_0 + \gamma_1 \log \zeta_{t-1} + \gamma_2 \exp\{-\gamma_3 \alpha_{t-1} \mathbb{I}(P_{t-1} > Z_{t-1}) [(P_{t-1}/Z_{t-1})^{1/\zeta_{t-1}} - 1]\}, \tag{16}$$

where $\{Z_t\}$ is an *i.i.d.* sequence of uniform distribution in (0,1) random variables and P_t is as given in (3).

Next, we propose the stationary and geometrically ergodic process of $\{\alpha_t, \zeta_t\}$ given in (15) and (16).

Theorem 1. *If parameters $\beta_2, \beta_3, \gamma_2, \gamma_3 > 0, \beta_0, \gamma_0 \in \mathbb{R}$ and $0 \leq \beta_1 \neq \gamma_1 < 1$, the latent process $\{\alpha_t, \zeta_t\}$ is defined as stationary and geometrically ergodic.*

Assumption 1. *Assume the parameter space Θ is a compact set of Θ_s . Suppose the observations $\{Y_t\}_{t=1}^n$ are generated from a stationary and ergodic DCP process with the true parameter θ_0 which is in the interior of Θ .*

Denote $L_n(\theta)$ based on θ and an arbitrary initial value $(\tilde{\alpha}_1, \tilde{\zeta}_1)$ as $\tilde{L}_n(\theta)$.

Theorem 2 (Consistency). *Under Assumption 1, there exists a sequence $\{\hat{\theta}_n\}_{n \geq 1}$ of local maximizers of $\tilde{L}_n(\theta)$ such that $\hat{\theta}_n \rightarrow_p \theta_0$ and $\|\hat{\theta}_n - \theta_0\| \leq \tau_n$, where $\tau_n = O_p(n^{-r}), 0 < r < 1/2$.*

Theorem 3 (Asymptotic Normality). *Under the same conditions as in Theorem 2, $\sqrt{n}(\hat{\theta}_n - \theta_0) \rightarrow N(0, M_0^{-1})$, where $\hat{\theta}_n$ is given in Theorem 2 and M_0 is the Fisher information matrix with θ_0 .*

Theorems 2 and 3 show the existence and asymptotic normality, respectively, of the MLE $\hat{\theta}_n$. However, the uniqueness of the MLE must be proved. Proposition 1 gives an answer to this.

Proposition 1 (Asymptotic Uniqueness). *Under the same conditions as in Theorem 2, $P(\hat{\theta}_n$ is the unique global maximizer of $\tilde{L}_n(\theta)$ over $\Theta) \rightarrow 1$, where $\hat{\theta}_n$ is given in Theorem 2.*

The proofs of Theorems 1–3 and Proposition 1 are shown in Appendix A.

4. Long Short-Term Memory Model

The main purpose of LSTM is to solve the problem of long-distance dependency in the training of recurrent neural networks. First, it is necessary to set the cell memory unit, and introduce the forget gate, input gate and output gate into the recurrent neural network (RNN), so that information transmission can be controlled. The state (namely the memory unit) update is also based on these “gates”, which ensure that the LSTM model can save long-distance information. Under the influence of the memory unit, these “gates” will be in a controllable range. LSTM then can save, update and read long-distance information, and gradient explosion or disappearance during training are well solved. In the time-series

data for air pollution, more comprehensive long-distance dependence information can be extracted. From the perspective of the whole model, the main components are as follows: output gate o_t , input gate i_t , memory unit C_t and forget gate f_t . The structure is shown in Figure 1.

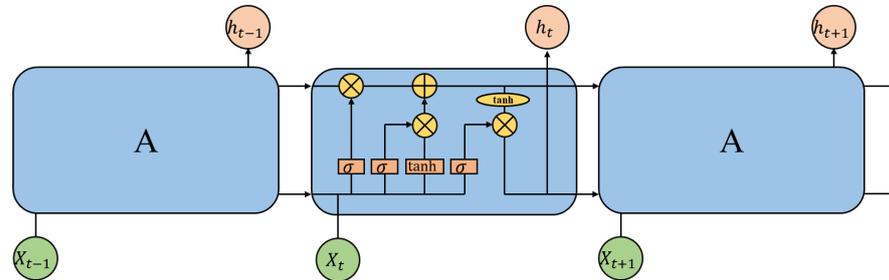


Figure 1. Structure of LSTM.

The first “gate” that the LSTM passes through is the “forget gate”, which discards part of the information in the previous memory unit. This step is realized by a sigmoid function, which uses the t weighted values of the current input and the output of the previous moment to obtain a number in the range of 0–1, which controls the information transfer. The value 1 represents complete retention and 0 represents complete discarding. The details are given in Equation (17):

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f), \tag{17}$$

The input gate controls what information is added to the cell, and the calculation process is shown in Equations (18) and (19):

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \tag{18}$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tanh(W_c \cdot [h_{t-1}, x_t] + b_c), \tag{19}$$

The output gate controls what information is used for the task output at this moment, and the calculation process is shown in (20) and (21):

$$o^t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \tag{20}$$

$$h^t = o^t \cdot \tanh(C_t), \tag{21}$$

In the above equations, W_i , W_f and W_o denote the weight matrices of the corresponding gate, b_i , b_f and b_o denote the corresponding gate bias matrices, σ and \tanh denote the activation functions, o_t denotes the output gate, i_t denotes the input gate, C_t denotes the memory unit, f_t denotes the forget gate, x_t denotes the input at time t and h_t denotes the output at time t .

5. Simulation Study

In this section, the performance of the MLE for the DCP models is investigated using six numerical experiments. To investigate the performance of the MLE, we generate data from the three DCP models given in (5) and (8)–(13), with the parameters shown in Table 1. These sets of parameters are the MLEs obtained from an analysis of real observations in Beijing from 3 January 2015 to 8 August 2020 and from 1 January 2018 to 8 August 2020, where the weather factors are from the China Meteorological Data Service Center and the air quality factors are from the China National Environmental Monitoring Center. In addition, the estimations of β_1 and β_2 are close to 0, especially in the three models from 3 January 2015 to 8 August 2020, which indicates that the scale parameter α_t can be considered a constant to a certain extent (a consideration that will be realized in future research). Due to more attention being given to the tail index ζ_t and the wider applicability of the DCP models, we made no changes to α_t .

Figure 2 displays a line graph of the PM_{2.5} concentration time series in Beijing. As shown in Figure 2, with the improvement in the national environmental governance level and public awareness of environmental protection, the PM_{2.5} concentration generally shows a downward trend. Figure 2 shows that 2018 is a noteworthy year with significant governance effects, suggesting that PM_{2.5} concentrations after this need to be analyzed separately. Hence, real data from 1 January 2018 to 8 August 2020 are also fitted to the three DCP models, in addition to the real data from 3 January 2015 to 8 August 2020. According to the World Air Quality Report 2021 released by IQAir [1], China has seen a 21% overall reduction in annual PM_{2.5} concentrations since 2018, which justifies the separation of the data from 1 January 2018 to 8 August 2020.

Threshold selection is a key issue in extreme value analysis based on the POT method. For these two sets of observations, from 3 January 2015 to 8 August 2020 and 1 January 2018 to 8 August 2020, we select two thresholds determined by Bader et al. (2018) [38] and Davison and Smith (1990) [50]. Bader et al. (2018) [38] proposed an automated threshold selection procedure using a stop rule that controls the false discovery rate in ordered hypothesis testing. The ForwardStop rule provides an automated selection procedure combined with sequential hypothesis testing when the level of desired error control and a set of thresholds are given. Based on the goodness-of-fit of the GPD, Davison and Smith (1990) [50] proposed a threshold selection approach where the threshold is chosen as the lowest value above which the GPD fits the exceedances adequately. In this study, the threshold selection results were 2.4660 using the method of Bader et al. (2018) [38] for PM_{2.5} data from 3 January 2015 to 8 August 2020 and 0.5716 using the method of Davison and Smith (1990) for PM_{2.5} data from 1 January 2018 to 8 August 2020, and approximately 4% and 18% of the corresponding real data exceeded these two thresholds, respectively, ensuring a sufficiently high value.

Tables 2–7 show the averages of the mean values and the standard deviations with different sample sizes from the three DCP models in the above two periods. We also calculated the corresponding root mean squared error (RMSE) and absolute bias (Abias) to measure the estimation effect, as shown in Tables 2–7. We obtained simulated exceedances Y_t with lengths of 1000 and 2000, respectively. The experiments were repeated 500 times for each sample size. As shown in Tables 2–7, the RMSE and Abias values for the parameter estimations using the real data from 1 January 2018 to 8 August 2020 were mostly smaller than those from 3 January 2015 to 8 August 2020, while those for the sample size of 2000 were mostly smaller than those for the sample size of 1000, and the parameter estimation of the tail index ζ_t was better than that of α_t . The values of RMSE and Abias explain the validity of our estimation.

To enable observation of the performance of our model more directly, Figure 3 depicts the dynamics of the tail index $\hat{\zeta}_t$ estimated by MLE (red line) and the simulated tail index ζ_t (black line) under the experiments for $n = 2000$. We can see that the estimated tail index $\hat{\zeta}_t$ was almost the same as the simulated tail index ζ_t in the three DCP models. In addition, we calculated the correlation between the two series, and the results were 0.9328, 0.9559, 0.9921, 0.9964, 0.9591 and 0.9679, corresponding to Figure 3a–f, respectively, which shows the similarity of the two curves better. However, we cannot judge the simulation effect from the similarity of the two curves only. Figure 3 illustrates the sufficiency of our estimation.

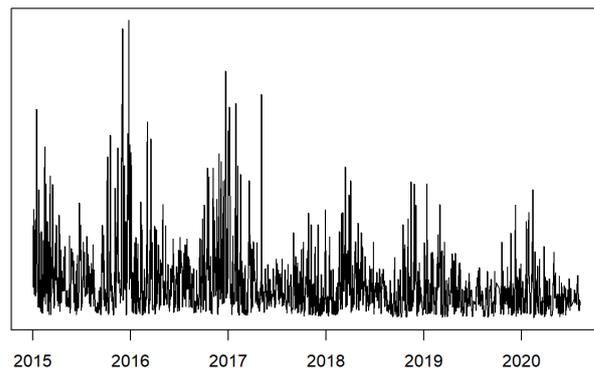


Figure 2. The graph of PM_{2.5} time series in Beijing.

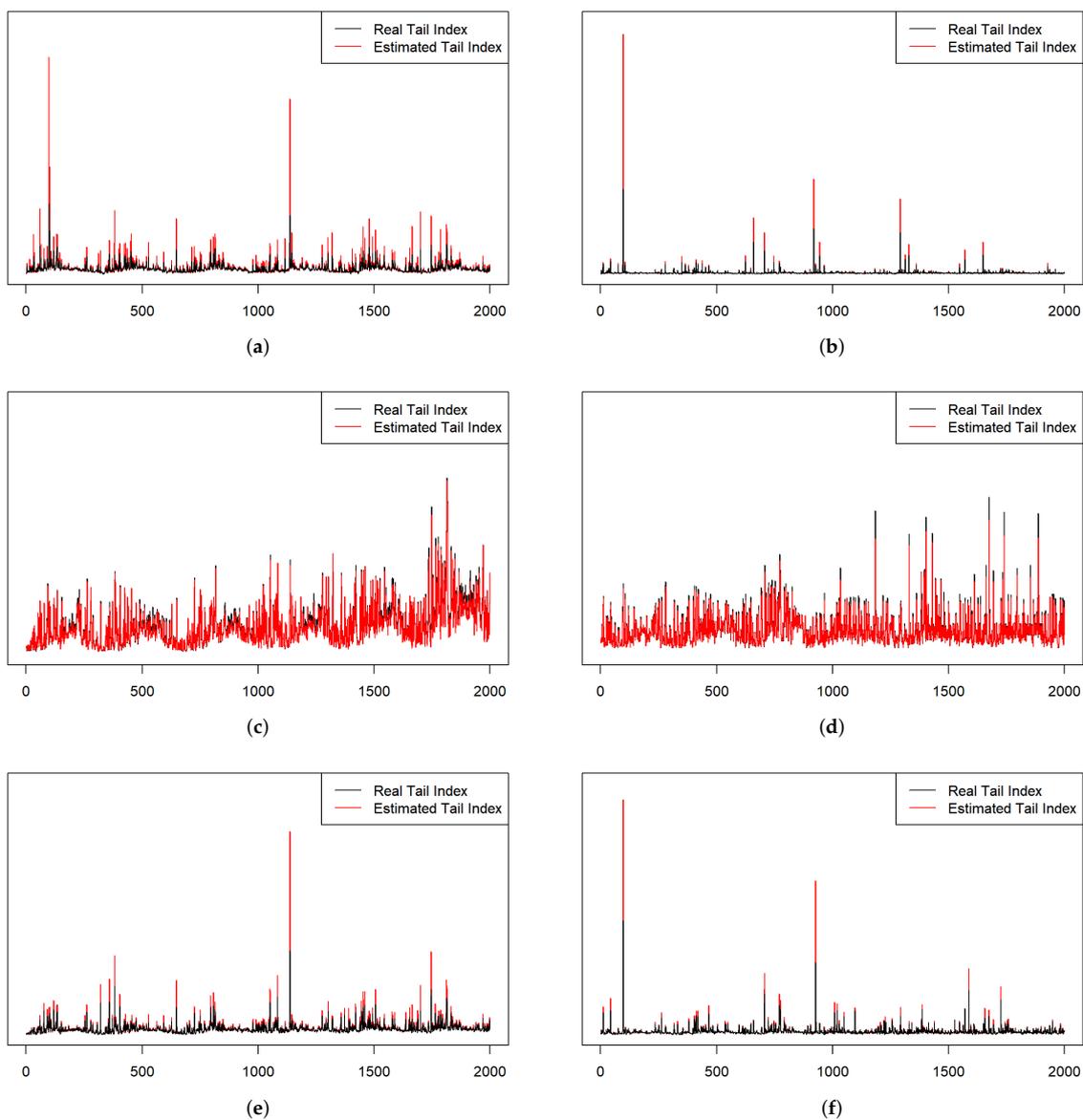


Figure 3. Tail index $\hat{\zeta}_t$ estimated by MLE and simulated tail index ζ_t . (a) The DCP model with weather factors from 2015 to 2020. (b) The DCP model with weather factors from 2018 to 2020. (c) The DCP model with air quality factors from 2015 to 2020. (d) The DCP model with air quality factors from 2018 to 2020. (e) The DCP model with mixed factors from 2015 to 2020. (f) The DCP model with mixed factors from 2018 to 2020.

Table 1. Parameter estimation of the DCP models. Weather1 and weather2 represent the DCP models with weather factors from 3 January 2015 to 8 August 2020 and from 1 January 2018 to 8 August 2020. Air1 and air2, and mixed1 and mixed2 are similar to weather1 and weather2.

	Weather1	Weather2	Air1	Air2	Mixed1	Mixed2
β_0	0.2989	0.0950	−0.0191	0.5932	−0.0186	0.8154
β_1	0.0000	0.9210	0.0033	0.3744	0.0000	0.0000
β_2	0.0000	0.0303	0.0000	0.0602	0.0000	0.0696
β_3	6.4028	33.6097	2.1760	0.0001	2.8285	0.2035
β_4	8.2352	0.1524	0.0042	2.0413	0.3989	2.9834
β_5	0.9159	0.4756	0.3541	0.7755	6.1266	0.0001
β_6	4.4001	1.0734	0.0001	0.0001	0.1056	0.6691
β_7					0.6196	0.0001
γ_0	−1.3618	−0.0116	−0.3471	0.3087	−1.9032	−0.6329
γ_1	0.3739	0.1931	0.3162	0.0182	0.0475	0.0524
γ_2	2.1928	1.2830	1.3070	1.0492	3.2466	2.0631
γ_3	0.3367	0.8814	0.0001	0.5450	0.0001	0.1949
γ_4	0.0666	0.0192	0.1809	0.1848	0.0655	0.1533
γ_5	0.1691	0.3138	0.2600	0.2209	0.1140	0.0785
γ_6	0.0001	0.0001	0.2353	0.2472	0.1141	0.1859
γ_7					0.0001	0.0001

Table 2. Mean, standard deviation, RMSE and Abias of 500 corresponding parameter values estimated by MLE in the DCP with weather factors from 3 January 2015 to 8 August 2020.

Para	True Value	n = 1000				n = 2000			
		Mean	Sd	RMSE	Abias	Mean	Sd	RMSE	Abias
β_0	0.2989	0.7935	1.4991	1.5771	0.6634	0.5055	0.9202	0.9422	0.3709
β_1	0.0000	0.2133	0.2898	0.3596	0.2133	0.1525	0.2499	0.2925	0.1525
β_2	0.0000	0.6980	1.6177	1.7604	0.6980	0.2982	0.9659	1.0100	0.2982
β_3	6.4028	6.8126	16.8403	16.8285	8.1358	8.5570	16.9945	17.1137	8.6371
β_4	8.2352	2.8520	2.4664	5.9203	5.4315	4.1413	2.5876	4.8417	4.1485
β_5	0.9159	0.8824	0.6961	0.6962	0.5726	0.8390	0.6126	0.6168	0.4842
β_6	4.4001	1.6663	1.5977	3.1656	2.8693	2.3185	1.4552	2.5390	2.1497
γ_0	−1.3618	−1.0625	0.8750	0.9239	0.6968	−1.2880	0.9361	0.9380	0.6209
γ_1	0.3739	0.3519	0.1439	0.1455	0.1155	0.3514	0.1119	0.1140	0.0887
γ_2	2.1928	1.9536	0.9417	0.9707	0.7376	2.1537	1.0054	1.0051	0.6710
γ_3	0.3367	1.1742	5.1284	5.1912	1.0035	1.0593	5.0468	5.0933	0.8586
γ_4	0.0666	0.1206	0.1802	0.1880	0.0764	0.0795	0.0662	0.0674	0.0363
γ_5	0.1691	0.3291	0.2729	0.3161	0.1722	0.2380	0.1296	0.1467	0.0889
γ_6	0.0001	0.0259	0.0639	0.0689	0.0258	0.0117	0.0237	0.0263	0.0116

Table 3. Mean, standard deviation, RMSE and Abias of 500 corresponding parameter values estimated by MLE in the DCP with air quality factors from 3 January 2015 to 8 August 2020.

Para	True Value	n = 1000				n = 2000			
		Mean	Sd	RMSE	Abias	Mean	Sd	RMSE	Abias
β_0	−0.0191	0.6258	1.8437	1.9514	0.7166	0.6809	2.0694	2.1826	0.7723
β_1	0.0033	0.2380	0.3223	0.3985	0.2368	0.1968	0.2839	0.3434	0.1959
β_2	0.0000	1.3443	2.2793	2.6442	1.3443	1.0134	2.1222	2.3499	1.0134
β_3	2.1760	13.6808	28.8885	31.0683	13.3849	14.1147	28.1351	30.5374	13.8677
β_4	0.0042	3.7857	6.2986	7.3412	3.7834	4.3086	8.0154	9.0910	4.3064
β_5	0.3541	1.3850	1.9586	2.2116	1.3261	1.2559	1.7233	1.9435	1.1972
β_6	0.0001	1.6231	2.5665	3.0344	1.6230	1.4659	2.3098	2.7337	1.4658

Table 3. Cont.

Para	True Value	n = 1000				n = 2000			
		Mean	Sd	RMSE	Abias	Mean	Sd	RMSE	Abias
γ_0	-0.3471	-0.3156	0.2649	0.2665	0.1966	-0.3339	0.2246	0.2247	0.1594
γ_1	0.3162	0.2823	0.1097	0.1147	0.0875	0.2996	0.0882	0.0896	0.0689
γ_2	1.3070	1.5039	0.7812	0.8049	0.3558	1.3508	0.2809	0.2840	0.1981
γ_3	0.0001	0.1969	1.8029	1.8118	0.1968	0.0543	0.5805	0.5825	0.0542
γ_4	0.1809	0.2698	0.2871	0.3003	0.1461	0.2182	0.2626	0.2650	0.0926
γ_5	0.2600	0.3429	0.2274	0.2418	0.1666	0.2927	0.1313	0.1352	0.0965
γ_6	0.2353	0.3471	0.3084	0.3277	0.2023	0.2860	0.1691	0.1764	0.1214

Table 4. Mean, standard deviation, RMSE and Abias of 500 corresponding parameter values estimated by MLE in the DCP with mixed weather and air quality factors from 3 January 2015 to 8 August 2020.

Para	True Value	n = 1000				n = 2000			
		Mean	Sd	RMSE	Abias	Mean	Sd	RMSE	Abias
β_0	-0.0186	0.4915	1.5853	1.6639	0.5978	0.3881	1.2049	1.2705	0.4428
β_1	0.0000	0.2219	0.3009	0.3736	0.2219	0.2228	0.3111	0.3824	0.2228
β_2	0.0000	0.5849	1.6337	1.7337	0.5849	0.3853	1.2311	1.2888	0.3853
β_3	2.8285	7.7282	19.5914	20.1758	7.7747	8.2657	19.3077	20.0401	7.9301
β_4	0.3989	1.5797	2.8494	3.0817	1.6143	1.7926	2.9913	3.2974	1.7668
β_5	6.1266	2.9585	4.8771	5.8116	5.1223	3.1062	4.3778	5.3150	4.7418
β_6	0.1056	1.0466	1.1051	1.4507	0.9986	0.9204	0.9381	1.2418	0.8764
β_7	0.6196	1.3400	2.0752	2.1947	1.3241	1.4397	2.2763	2.4174	1.4125
γ_0	-1.9032	-1.7337	0.9557	0.9697	0.7988	-1.8052	0.8761	0.8807	0.7080
γ_1	0.0475	0.0730	0.0899	0.0933	0.0705	0.0666	0.0803	0.0824	0.0628
γ_2	3.2466	3.0196	0.9698	0.9950	0.8249	3.0970	0.8919	0.9035	0.7355
γ_3	0.0001	0.0243	0.3376	0.3381	0.0242	0.0007	0.0033	0.0034	0.0006
γ_4	0.0655	0.0712	0.0517	0.0520	0.0319	0.0688	0.0326	0.0328	0.0236
γ_5	0.1140	0.1621	0.1388	0.1467	0.0754	0.1370	0.0646	0.0685	0.0486
γ_6	0.1141	0.1626	0.1192	0.1286	0.0717	0.1345	0.0587	0.0621	0.0428
γ_7	0.0001	0.0144	0.0311	0.0342	0.0143	0.0088	0.0136	0.0161	0.0087

Table 5. Mean, standard deviation, RMSE and Abias of 500 corresponding parameter values estimated by MLE in the DCP with weather factors from 1 January 2018 to 8 August 2020.

Para	True Value	n = 1000				n = 2000			
		Mean	Sd	RMSE	Abias	Mean	Sd	RMSE	Abias
β_0	0.0950	0.7499	1.5926	1.7205	0.6839	0.3280	0.8420	0.8728	0.2556
β_1	0.9210	0.6431	0.3598	0.4543	0.3029	0.8072	0.2381	0.2637	0.1355
β_2	0.0303	0.4727	1.5367	1.5976	0.4729	0.1854	0.7942	0.8084	0.1764
β_3	33.6097	24.5359	38.6197	39.6337	36.2003	29.1089	40.3405	40.5506	36.7034
β_4	0.1524	1.4197	1.9596	2.3321	1.3589	0.4509	0.6220	0.6893	0.3917
β_5	0.4756	1.0975	1.2445	1.3901	0.9800	0.5554	0.5925	0.5972	0.4135
β_6	1.0734	1.5782	1.6604	1.7339	1.2105	1.1625	0.8826	0.8862	0.6713
γ_0	-0.0116	0.1267	0.2986	0.3289	0.2612	0.0886	0.2156	0.2376	0.1884
γ_1	0.1931	0.1890	0.0823	0.0823	0.0669	0.1917	0.0574	0.0573	0.0451
γ_2	1.2830	1.1413	0.3072	0.3380	0.2702	1.1741	0.2223	0.2473	0.2020
γ_3	0.8814	1.0276	0.7411	0.7547	0.4188	0.9004	0.3328	0.3330	0.2592
γ_4	0.0192	0.0256	0.0366	0.0371	0.0223	0.0222	0.0204	0.0206	0.0163
γ_5	0.3138	0.4060	0.1500	0.1759	0.1187	0.3642	0.0810	0.0953	0.0718
γ_6	0.0001	0.0185	0.0297	0.0349	0.0184	0.0107	0.0165	0.0196	0.0106

Table 6. Mean, standard deviation, RMSE and Abias of 500 corresponding parameter values estimated by MLE in the DCP with air quality factors from 1 January 2018 to 8 August 2020.

Para	True Value	n = 1000				n = 2000			
		Mean	Sd	RMSE	Abias	Mean	Sd	RMSE	Abias
β_0	0.5932	0.9593	1.1273	1.1842	0.4807	0.6085	0.0871	0.0884	0.0667
β_1	0.3744	0.2434	0.2857	0.3140	0.2832	0.3525	0.0805	0.0834	0.0613
β_2	0.0602	0.3507	1.1224	1.1583	0.3455	0.0638	0.0350	0.0351	0.0271
β_3	0.0001	10.7321	25.8047	27.9236	10.7320	1.4567	9.0550	9.1625	1.4566
β_4	2.0413	4.6154	7.3362	7.7678	4.3355	2.1536	0.3725	0.3887	0.2452
β_5	0.7755	1.4947	2.1804	2.2939	1.3944	0.8279	0.5246	0.5267	0.2043
β_6	0.0001	1.0952	2.2492	2.4996	1.0951	0.0561	0.1118	0.1249	0.0560
γ_0	0.3087	0.3379	0.1903	0.1923	0.1488	0.3385	0.1461	0.1490	0.1185
γ_1	0.0182	0.0546	0.0650	0.0744	0.0516	0.0392	0.0456	0.0502	0.0358
γ_2	1.0492	0.9473	0.2153	0.2380	0.1910	0.9757	0.1639	0.1795	0.1448
γ_3	0.5450	0.4920	0.2818	0.2864	0.2076	0.4996	0.2112	0.2158	0.1548
γ_4	0.1848	0.2208	0.1069	0.1128	0.0833	0.2031	0.0690	0.0714	0.0420
γ_5	0.2209	0.2539	0.0943	0.0998	0.0752	0.2445	0.0598	0.0642	0.0441
γ_6	0.2472	0.2810	0.0979	0.1035	0.0771	0.2714	0.0581	0.0629	0.0479

Table 7. Mean, standard deviation, RMSE and Abias of 500 corresponding parameter values estimated by MLE in the DCP with mixed weather and air quality factors from 1 January 2018 to 8 August 2020.

Para	True Value	n = 1000				n = 2000			
		Mean	Sd	RMSE	Abias	Mean	Sd	RMSE	Abias
β_0	0.8154	0.8577	0.9534	0.9534	0.3941	0.6715	0.9247	0.9349	0.2631
β_1	0.0000	0.2211	0.2985	0.3712	0.2211	0.0501	0.1502	0.1582	0.0501
β_2	0.0696	0.3003	0.9290	0.9563	0.3031	0.0770	0.0731	0.0734	0.0559
β_3	0.2035	12.2266	26.3745	28.9616	12.1660	6.1382	14.5687	15.7176	6.0234
β_4	2.9834	2.8016	3.4427	3.4440	2.8849	2.5224	1.1180	1.2083	0.7971
β_5	0.0001	1.7871	3.2419	3.6989	1.7870	0.9101	1.7622	1.9817	0.9100
β_6	0.6691	1.0305	1.0784	1.1364	0.8332	0.9123	0.9829	1.0116	0.5585
β_7	0.0001	1.5230	2.4326	2.8679	1.5229	0.6633	1.4938	1.6330	0.6632
γ_0	-0.6329	-0.4751	0.3628	0.3953	0.3139	-0.9250	1.1942	1.2282	0.6621
γ_1	0.0524	0.0620	0.0667	0.0673	0.0538	0.0630	0.0559	0.0569	0.0429
γ_2	2.0631	1.8830	0.3908	0.4299	0.3385	2.3010	1.1443	1.1677	0.6600
γ_3	0.1949	0.1904	0.0851	0.0852	0.0628	0.2905	0.9609	0.9647	0.1879
γ_4	0.1533	0.1799	0.0607	0.0662	0.0471	0.1689	0.0915	0.0927	0.0539
γ_5	0.0785	0.0982	0.0416	0.0460	0.0344	0.0877	0.0740	0.0745	0.0319
γ_6	0.1859	0.2243	0.0631	0.0738	0.0533	0.2061	0.1122	0.1139	0.0612
γ_7	0.0001	0.0123	0.0210	0.0243	0.0122	0.0079	0.0143	0.0163	0.0078

6. Real Data Applications

In this section, to verify the performance of the DCP models, we consider the weather and air quality time series in Beijing from 3 January 2015 to 8 August 2020 and from 1 January 2018 to 8 August 2020 obtained from the China Meteorological Data Service Center and China National Environmental Monitoring Center, as the sample for our experimental analysis. Based on these two periods observations, we employed the three DCP models (one with weather factors, one with air quality factors and the last with mixed weather and air quality factors) given in Equations (5) and (8)–(13) to fit the smog data and used the MLE method described in Section 3 to estimate the parameters. In both cases, the DCP models showed their superiority in reflecting the time dependence of the pollutant concentration, providing a potential warning signal for smog prevention and control. First, we made a fat-tailed diagnosis of the observations using an exponential QQ plot. Figure 4 shows that the real data for PM_{2.5} in Beijing from 3 January 2015 to 8 August 2020 and from 1 January 2018 to 8 August 2020 are fat tailed.

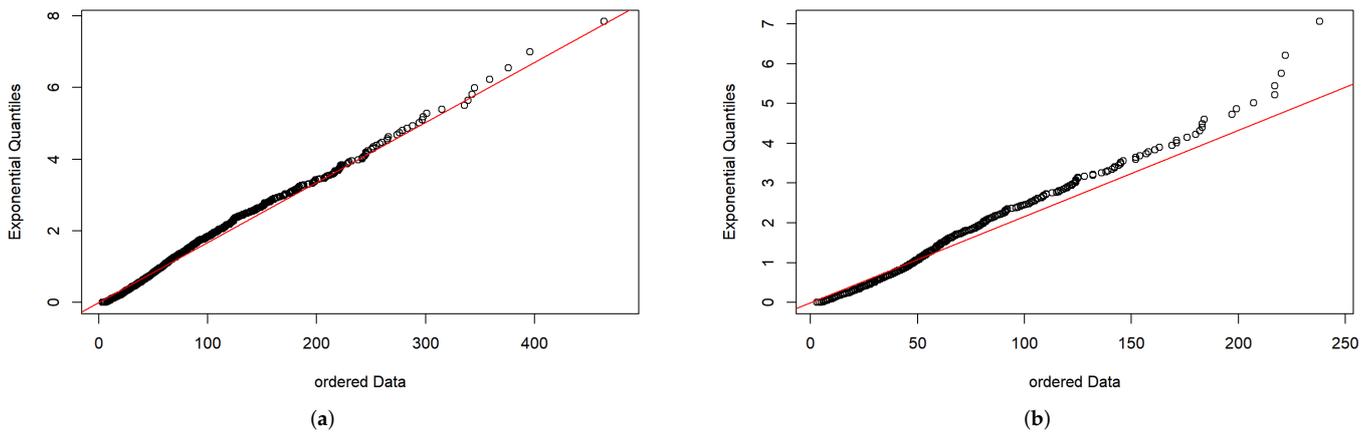


Figure 4. Exponential QQ plot of real PM_{2.5} data in Beijing: (a) from 3 January 2015 to 8 August 2020; (b) from 1 January 2018 to 8 August 2020.

From the results in Table 1, we can see that the tail index ζ_t is more affected by Y_{t-1} , which is consistent with Figure 5. The estimated tail index of the DCP can reflect the severity of smog to some extent and may even play an early warning role for smog disruption. The graph of the estimated tail index ζ_t and positive exceedances Y_t from the three DCP models is given in Figure 5, which shows that there is a strong negative correlation between ζ_t and Y_t , and the tail index volatility is more intuitive. It is interesting to note that Figure 5a,c,e and Figure 5b,d,f have very similar variation tendencies, and it can clearly be seen that the tail index ζ_t starts to decline in the middle of each year, and at the end of the year the tail index becomes lower, which can be regarded as an effective indicator for measuring the level of smog.

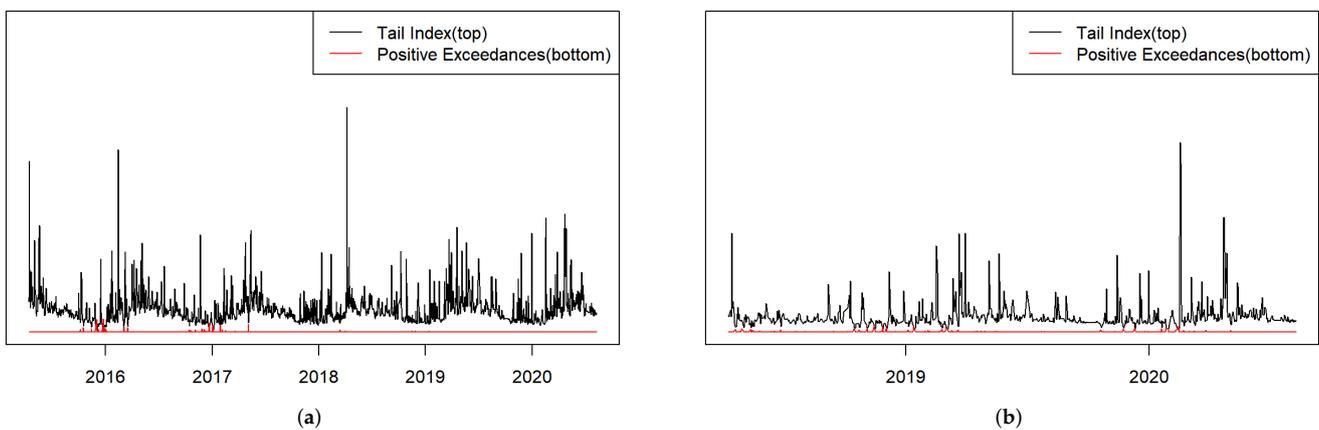


Figure 5. Cont.

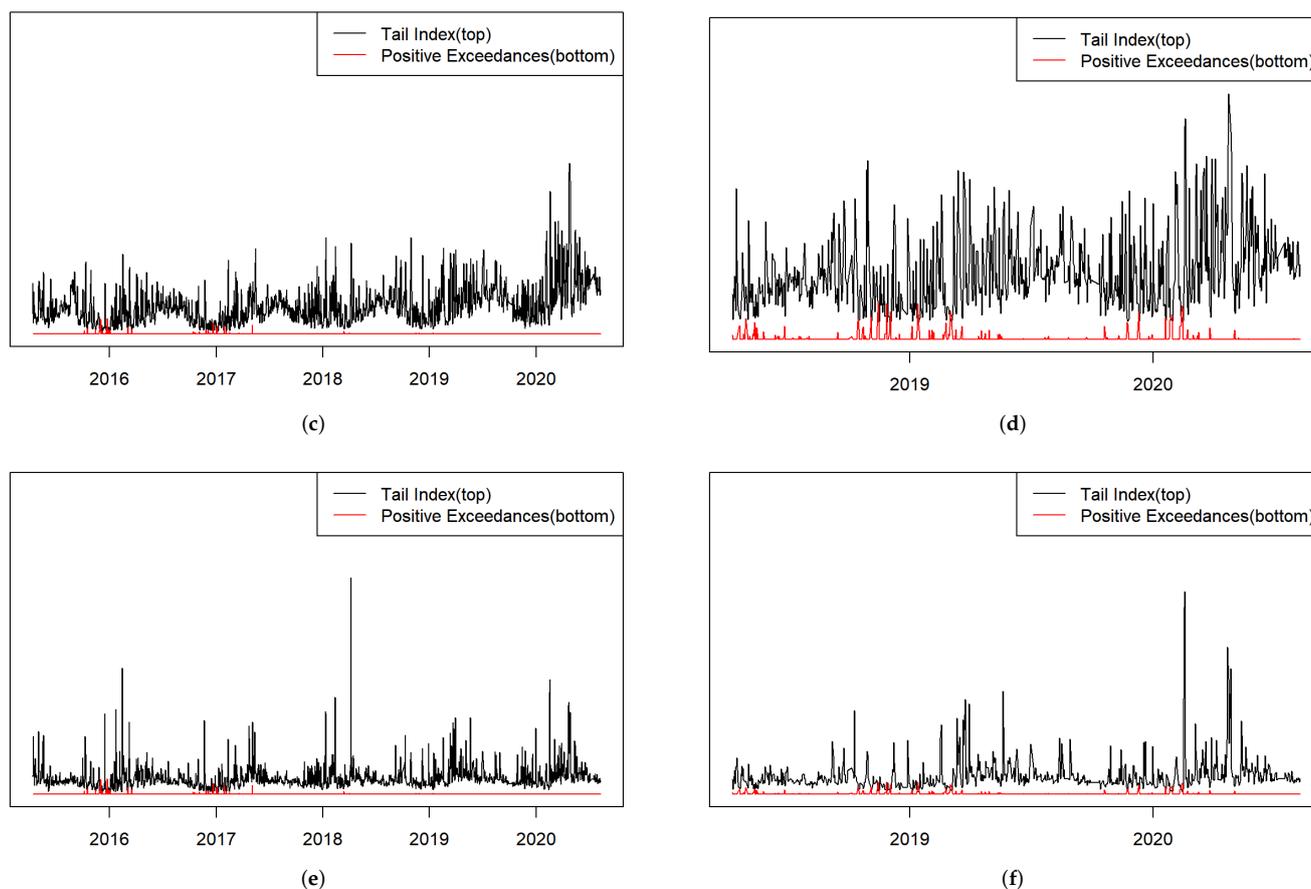


Figure 5. Estimated tail index $\hat{\zeta}_t$ from the three DCP models and positive exceedances Y_t . (a) The DCP model with weather factors from 3 January 2015 to 8 August 2020 with $u = 2.4660$. (b) The DCP model with weather factors from 1 January 2018 to 8 August 2020 with $u = 0.5716$. (c) The DCP model with air quality factors from from 3 January 2015 to 8 August 2020 with $u = 2.4660$. (d) The DCP model with air quality factors from 1 January 2018 to 8 August 2020 with $u = 0.5716$. (e) The DCP model with mixed factors from from 3 January 2015 to 8 August 2020 with $u = 2.4660$. (f) The DCP model with mixed factors from 1 January 2018 to 8 August 2020 with $u = 0.5716$.

Using the estimated parameters given in Table 1, we generated a sequence of fitted \hat{Y}_t values based on Equation (5) and plotted the line graphs of the fitted \hat{Y}_t values and real exceedances Y_t during the period from 1 January 2018 to 8 August 2020, as shown in Figure 6. It can be seen that the true values Y_t and the estimated values \hat{Y}_t from the three models were almost consistent in trend, and the three models were more sensitive to the estimation of Y_t , but the model with mixed weather factors and air quality factors showed values closest to the true values, which also verifies the superiority of the mixed model over the other two models and is consistent with the conclusion mentioned in Section 5. A comparison between the fitted \hat{Y}_t and real Y_t exceedances during the period from 3 January 2015 to 8 August 2020 was also performed, and similar results were obtained. Due to limited space, only Figure 6 is shown.

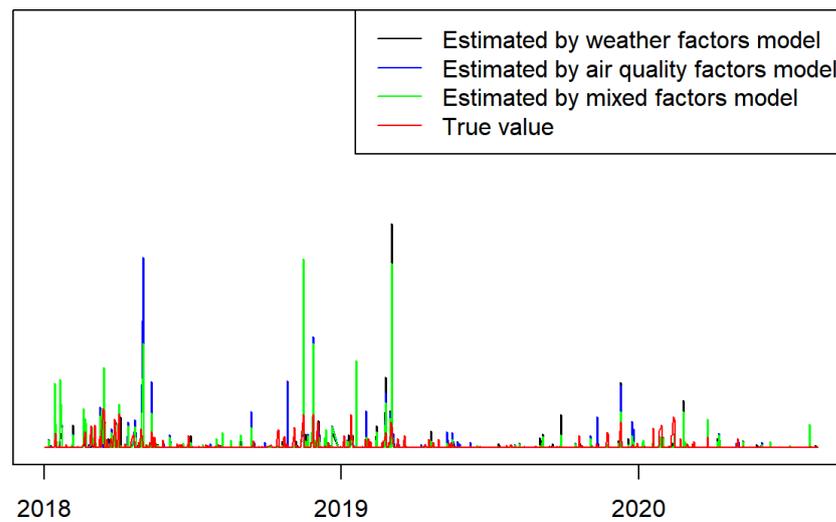


Figure 6. The line graphs of fitted \hat{Y}_t values and real exceedances Y_t

Next, we compared the estimated variances from the DCP models and GARCH, as shown in Figure 7. Similarly to [27], we calculated the conditional variance

$$\text{Var}(Y_t | \mathcal{F}_{t-1}) = \frac{\alpha_t^2}{(1+u)^{\zeta_t} \zeta_t - 1} \left[\frac{2}{\zeta_t - 2} - \frac{1}{(1+u)^{\zeta_t} (\zeta_t - 1)} \right].$$

Figure 7 shows that the standard deviations given by the DCP models and GARCH had similar trends, indicating that the DCP models could accurately reflect the volatility in a sense. Compared to the estimated volatility of GARCH, the DCP models are more sensitive in smog instances, thus potentially playing a better role in early warning. This is clearest in Figure 7e, where the fluctuation is largest.

We computed AIC and BIC from the DCP and dynamic conditional Weibull (DCW) model given in [35]. The results are presented in Table 8. As shown in Table 8, the DCP model is more suitable than the DCW model, based on AIC and BIC criteria.

Finally, we used our proposed three models (5) and (8)–(13) to predict the daily $PM_{2.5}$ values from 9 August 2020 to 31 December 2021. We present only the results for the mixed models in (5), (12) and (13) here, with a training sample from 1 January 2018 to 8 August 2020, since similar results were obtained from the three models. The tail index ζ_t given in (13) and $PM_{2.5}$ given in (5) were predicted by using the real weather and air quality factors and the parameter estimation results given in Table 1. In order to analyze the fluctuating tendency and correlations of ζ_t and $PM_{2.5}$, the prediction results are presented together in Figure 8. From Figure 8, we can see that there is a strong negative correlation between ζ_t and $PM_{2.5}$, which enables the tail index to be used as a warning signal for air pollution. Furthermore, compared with the real smog values, the predictability of the future variation of $PM_{2.5}$ performs relatively well, as the real and predicted values are relative close and have a similar tendency.

Table 8. Comparison of the DCP and DCW models based on AIC and BIC criteria.

		AIC			BIC		
		Weather	Air	Mixed	Weather	Air	Mixed
3 January 2015–8 August 2020	DCP	588.9960	535.8680	531.0784	666.7786	613.6507	619.9729
	DCW	3327.6850	3907.5960	4908.1610	3411.0240	3990.9350	5002.6110
1 January 2018–8 August 2020	DCP	1035.3950	1019.2530	1001.3240	1035.3950	1086.0280	1077.6380
	DCW	2717.4300	3219.0270	1700.4180	2788.9740	3290.5710	1700.4180

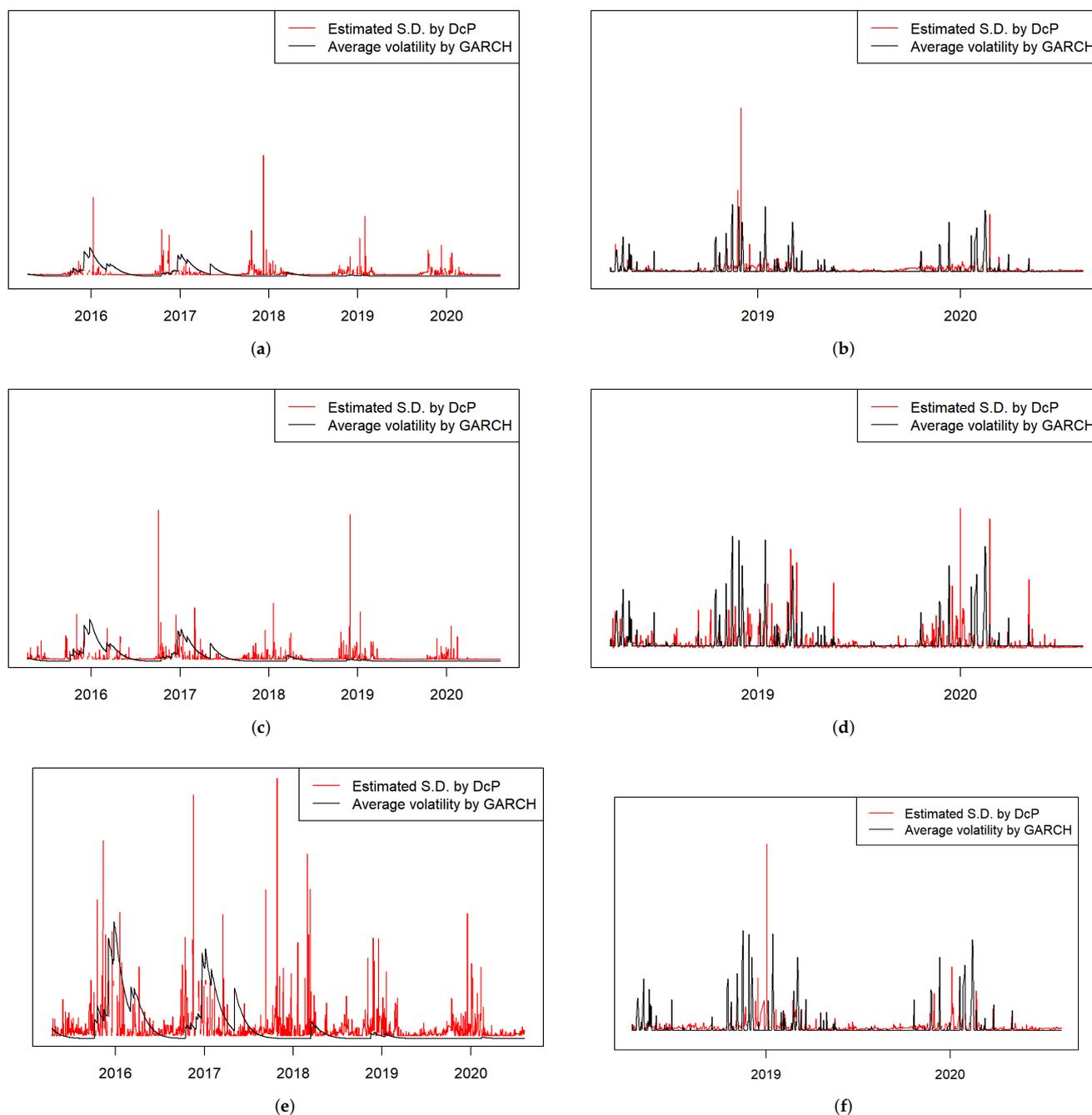


Figure 7. Estimated standard deviation for DCP vs. GARCH. (a) The DCP model with weather factors from 3 January 2015 to 8 August 2020. (b) The DCP model with weather factors from 1 January 2018 to 8 August 2020. (c) The DCP model with air quality factors from 3 January 2015 to 8 August 2020. (d) The DCP model with air quality factors from 1 January 2018 to 8 August 2020. (e) The DCP model with mixed factors from 3 January 2015 to 8 August 2020. (f) The DCP model with mixed factors from 1 January 2018 to 8 August 2020.

In addition, we used weather and air quality factors (including SO_2 , NO_2 , CO , O_3) with the LSTM technique to predict daily $PM_{2.5}$ values. The air pollution data from 1 January 2015 to 25 November 2019 were used as training data, those from 26 November 2019 to 8 August 2020 were used as verification data and those from 9 August 2020 to 31 December 2021 were used as test data. The LSTM network was trained by using the weather, air quality factors and $PM_{2.5}$ time series in Beijing to construct a training set. Then,

various weather and air quality factors were input into the test set to predict the $PM_{2.5}$ from 9 August 2020 to 31 December 2021 over a long period. As shown in Figure 9, the trend of the prediction for $PM_{2.5}$ was accurate, especially when the true values of $PM_{2.5}$ were less than 100. To better evaluate the experimental results quantitatively, the RMSE and coefficient of determination (R^2) were calculated, and the results were 20.14 and 0.65, respectively.

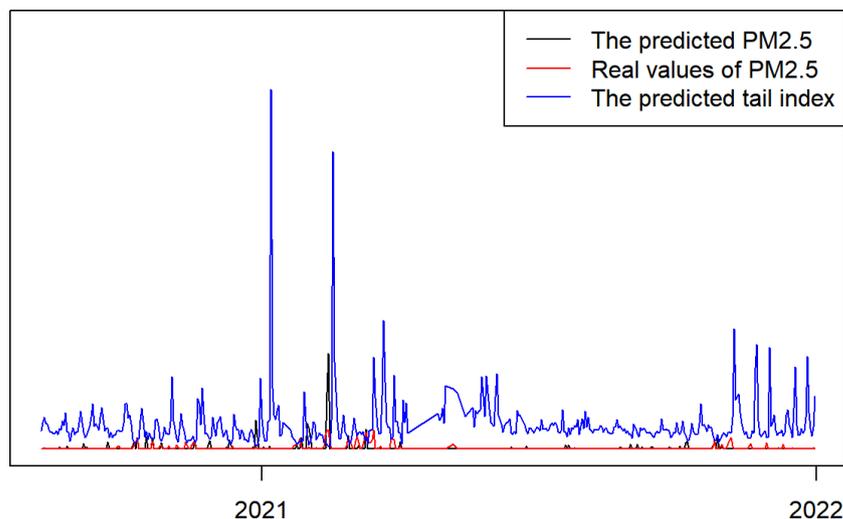


Figure 8. The line graphs of predicted \hat{Y}_t values and real exceedances Y_t .

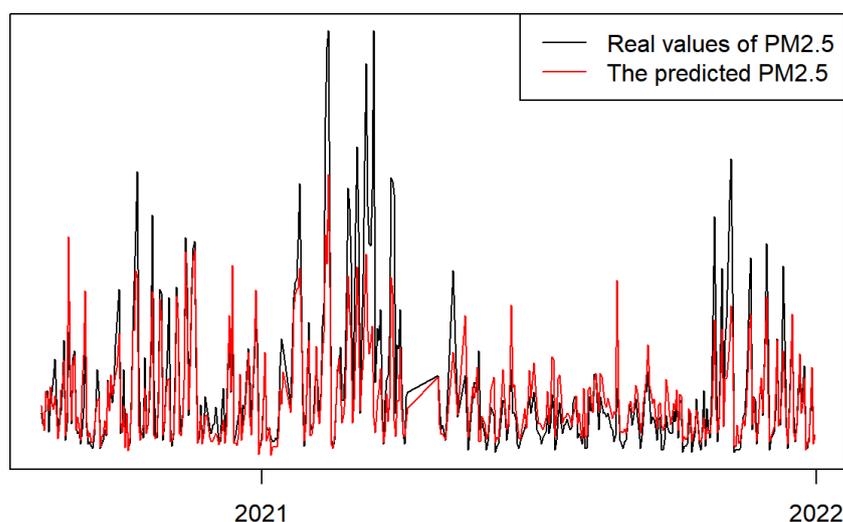


Figure 9. The long-term prediction of $PM_{2.5}$ values in Beijing.

7. Conclusions

In this paper, we investigated the prediction of pollutant concentrations using statistical inference methods and deep learning techniques. On the one hand, we proposed three models combined with the autoregressive structure under the POT framework. After obtaining two sufficiently high thresholds selected using the methods of by Bader et al. (2018) [38] and Davison and Smith (1990) [50], the DCP models provided a direct dynamic modeling of exceedances in the $PM_{2.5}$ time series, such that the scale parameter and tail index of the conditional generalized Pareto distribution changed over time. Weather and air quality factors were added to the DCP models for better performance and higher efficiency. The maximum likelihood estimation method was introduced to estimate the parameters in the DCP models, and its asymptotic properties were investigated. Simulation studies were carried out to demonstrate the validity and sufficiency of the estimation, revealing that the

parameter estimation of the DCP models was not sufficiently accurate but the tail index dynamics could be well approximated in the DCP models. Real data applications were used to present the superiority of the DCP models, showing that they could shed new light on the prevention and control of smog. On the other hand, based on the factors used in the mixed DCP model, we used LSTM to study the prediction of pollutant concentrations, and achieved satisfactory results. This paper aimed to improve the prediction ability for the concentration of pollutants, and valuable results were achieved. Given the requirements of the air pollution control target for further promoting ecological and environmental protection in the next five years, the proposed approaches and results in our paper are useful. To some extent, they could provide a theoretical basis and effective tools for improving the national air quality forecasting system, thus benefiting public health.

Nevertheless, there are still some points to be considered. In the DCP model with the autoregressive structure, it is meaningful to add weather and air quality factors, enriching the model and making it consistent, stable and sensitive. However, the relationship between the factors has not been scrutinized carefully, resulting in a lack of attention to its impacts on the model. In addition, it is possible to obtain better results when we compare other estimation methods. Therefore, prediction, as an important direction in our study of air pollutant concentrations, still has a long way to go. Combining artificial intelligence and machine learning, the prediction accuracy will certainly be improved by using a forecast combination, synthesizing the methods used to obtain the estimated results. Finally, with the help of combination forecasting, the advantages of individual forecasts are retained, and effective information is fully utilized to comprehensively forecast air pollution. In this work, we strive to make valuable advances in the intersection of statistics and machine learning and to provide effective theoretical and technical support for national continuous improvement of the modernization level of ecological environmental governance.

Author Contributions: Data curation, X.Z. and C.H.; funding acquisition, X.Z.; methodology, X.Z. and W.C.; project administration, X.Z.; software, X.Z., C.H. and Q.J.; supervision, W.C.; writing—original draft, X.Z., C.H. and Q.J.; writing—review and editing, Q.D. and Y.H. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (grant number 11801019).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The weather data and the air quality data were obtained from the China Meteorological Data Service Center and the China National Environmental Monitoring Center.

Acknowledgments: The authors would like to thank the referees and editors for their very helpful and constructive comments, which have significantly improved the quality of this paper.

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

Appendix A

The procedure for the proof follows Shen (2020) [27] and Zhao et al. (2018) [31]. The main difference from Zhao et al. (2018) [31] is the distribution of observations $\{Y_t\}$, and the main difference from Shen (2020) [27] is that we consider the autoregressive structure on both the scale and tail index series.

Proof of Theorem 1. The proof of the theorem is similar to Shen et al. (2020) [27] and Zhao et al. (2018) [31], so this proof is omitted. \square

Before proving Theorems 2 and 3, we first provide some lemmas.

Lemma A1 (Identifiability). *If $Y_t(\theta) = Y_t(\theta_0)$ almost sure (a.s.) for all t , we have $\theta = \theta_0$, where $\{Y_t\}$ is given in (5).*

Proof. Denote $\alpha_t = \alpha_t(\theta), \zeta_t = \zeta_t(\theta)$ and $\alpha_t^0 = \alpha_t(\theta_0), \zeta_t^0 = \zeta_t(\theta_0), P_t^0 = (1 + u)^{-\zeta_t^0}$. From $Y_t(\theta) = Y_t(\theta_0)$ a.s., we arrive at

$$\alpha_t \mathbb{I}(P_t > Z_t) [(P_t / Z_t)^{1/\zeta_t} - 1] = \alpha_t^0 \mathbb{I}(P_t^0 > Z_t) [(P_t^0 / Z_t)^{1/\zeta_t^0} - 1] \quad a.s.,$$

so

$$\begin{aligned} \alpha_t [(P_t / Z_t)^{1/\zeta_t} - 1] &= \alpha_t^0 [(P_t^0 / Z_t)^{1/\zeta_t^0} - 1] \quad a.s., \\ \mathbb{I}(P_t > Z_t) &= \mathbb{I}(P_t^0 > Z_t) \quad a.s.. \end{aligned}$$

After straightforward manipulations,

$$\alpha_t (P_t / Z_t)^{1/\zeta_t} - \alpha_t^0 (P_t^0 / Z_t)^{1/\zeta_t^0} = \alpha_t - \alpha_t^0 \quad a.s.,$$

Denote $\mathfrak{S}_t = \sigma(Y_t, Y_{t-1}, \dots)$, then $Z_t \perp \mathfrak{S}_{t-1}$ and $\alpha_t, \alpha_t^0, \zeta_t, \zeta_t^0, P_t, P_t^0 \in \mathfrak{S}_{t-1}$. Therefore, the above equation holds if and only if $\alpha_t = \alpha_t^0$ and $\zeta_t = \zeta_t^0$ a.s. From the autoregressive equations of $\log \alpha_t$ and $\log \zeta_t, \log \alpha_t = \log \alpha_t^0$ and $\log \zeta_t = \log \zeta_t^0$ a.s. can be rewritten as

$$\begin{aligned} \beta_0^0 + \beta_1^0 \log \alpha_{t-1} - \beta_2^0 \exp\{-\beta_3^0 Y_{t-1}\} &= \beta_0 + \beta_1 \log \alpha_{t-1} - \beta_2 \exp\{-\beta_3 Y_{t-1}\}, \\ \gamma_0^0 + \gamma_1^0 \log \zeta_{t-1} + \gamma_2^0 \exp\{-\gamma_3^0 Y_{t-1}\} &= \gamma_0 + \gamma_1 \log \zeta_{t-1} + \gamma_2 \exp\{-\gamma_3 Y_{t-1}\}. \end{aligned}$$

After rearrangement, the above two equations can be expressed as

$$\begin{aligned} \beta_0^0 - \beta_0 + (\beta_1^0 - \beta_1) \log \alpha_{t-1} &= \beta_2^0 \exp\{-\beta_3^0 Y_{t-1}\} - \beta_2 \exp\{-\beta_3 Y_{t-1}\}, \\ \gamma_0^0 - \gamma_0 + (\gamma_1^0 - \gamma_1) \log \zeta_{t-1} &= \gamma_2 \exp\{-\gamma_3 Y_{t-1}\} - \gamma_2^0 \exp\{-\gamma_3^0 Y_{t-1}\}. \end{aligned}$$

Since $\alpha_{t-1} \in \mathfrak{S}_{t-2}, \zeta_{t-1} \in \mathfrak{S}_{t-2}$ and $Y_{t-1} \notin \mathfrak{S}_{t-2}$, then $\beta_i = \beta_i^0$ and $\gamma_i = \gamma_i^0$ must hold for $i = 0, 1, 2, 3$. \square

In the following, we denote $(\alpha_t(\theta), \zeta_t(\theta))$ (or (α_t, ζ_t) for simplicity) as the time-dependent scale parameter and tail index based on θ and the true initial (α_1^0, ζ_1^0) , and denote $(\alpha_t(\theta_0), \zeta_t(\theta_0))$ (or (α_t^0, ζ_t^0) for simplicity) as the unobserved true hidden process based on the true θ_0 and the true initial (α_1^0, ζ_1^0) , and denote the t -th iterate series $(\tilde{\alpha}_t(\theta), \tilde{\zeta}_t(\theta))$ (or $(\tilde{\alpha}_t, \tilde{\zeta}_t)$ for simplicity) as the scale parameter and tail index series based on θ and an arbitrary initial $(\tilde{\alpha}_1, \tilde{\zeta}_1)$, and denote (α_L, α_U) and (ζ_L, ζ_U) as the uniform bound of α_t (or $\tilde{\alpha}_t$) and ζ_t (or $\tilde{\zeta}_t$) for all $\theta \in \Theta$ due to the compactness of Θ and boundedness of $-\beta_2 \exp(-\beta_3 Y_{t-1})$ and $\gamma_2 \exp(-\gamma_3 Y_{t-1})$.

Given (α_t, ζ_t) , the conditional log-likelihood function of Y_t is expressed as

$$l_t(\theta) = \mathbb{I}(Y_t = 0) \log[1 - (1 + u)^{-\zeta_t}] + \mathbb{I}(Y_t > 0) \left[\log \zeta_t - \log \alpha_t - \zeta_t \log(1 + u) - (\zeta_t + 1) \log \left(1 + \frac{Y_t}{\alpha_t} \right) \right].$$

Due to the conditional independence, the log-likelihood function is then given by

$$L_n(\theta) = \frac{1}{n} \sum_{t=1}^n l_t(\theta).$$

We denote $l_t(\theta)$ and $L_n(\theta)$ based on θ and an arbitrary initial value $(\tilde{\alpha}_1, \tilde{\zeta}_1)$ as $\tilde{L}_n(\theta)$ and $\tilde{l}_t(\theta)$.

Lemma A2. Under the same conditions as in Theorem 2, $E_{\theta_0}(\frac{\partial}{\partial \theta} l_t(\theta_0)) = 0$ and $M_0 = \text{Var}_{\theta_0}(\frac{\partial}{\partial \theta} l_t(\theta_0)) = -E_{\theta_0}(\frac{\partial^2}{\partial \theta \partial \theta^T} l_t(\theta_0))$ and M_0 is finite and positive definite.

Proof. The proof of the lemma is similar to Zhao et al. (2018) [31], so the proof is omitted. \square

Lemma A3. Under the same conditions as in Theorem 2, if $\|\Phi - \Phi_0\| < \tau_n$ and $\tau_n \searrow 0$, we have

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

uniformly over $\|\Phi - \Phi_0\| < \tau_n$, where $\Phi = (\beta_0, \beta_1, \beta_2, \beta_3)$ and $\Phi_0 = (\beta_0^0, \beta_1^0, \beta_2^0, \beta_3^0)$.

Proof. The proof of the lemma is similar to Zhao et al. (2018) [31], so the proof is omitted. \square

Lemma A4. Under the same conditions as in Theorem 2, if $\|\Psi - \Psi_0\| < \tau_n$ and $\tau_n \searrow 0$, we have

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

uniformly over $\|\Psi - \Psi_0\| < \tau_n$, where $\Psi = (\gamma_0, \gamma_1, \gamma_2, \gamma_3)$ and $\Psi_0 = (\gamma_0^0, \gamma_1^0, \gamma_2^0, \gamma_3^0)$.

Proof. The proof of the lemma is similar to Zhao et al. (2018) [31], so the proof is omitted. \square

Lemma A5. Under the same conditions as in Theorem 2, $\frac{\partial^2}{\partial \theta_i \partial \theta_j} L_n(\theta_n) \rightarrow_p -m_{\theta_i \theta_j}(\theta_0)$, uniformly over $\|\theta_n - \theta_0\| < \tau_n$, where $\tau_n \sim n^{-r}, r > 0, m_{\theta_i \theta_j}(\theta_0) = -E_{\theta_0}(\frac{\partial^2}{\partial \theta_i \partial \theta_j} I_1(\theta_0))$.

Proof. We only prove the case for $\frac{\partial^2}{\partial \beta_0^2} L_n(\theta_n) \rightarrow_p -m_{\beta_0 \beta_0}(\theta_0)$, as the proof for other cases is similar. From the law of large numbers, we know that $\frac{\partial^2}{\partial \beta_0^2} L_n(\theta_0) \rightarrow_p -m_{\beta_0 \beta_0}(\theta_0)$. Then, we need to prove that $\frac{\partial^2}{\partial \beta_0^2} L_n(\theta_n) - \frac{\partial^2}{\partial \beta_0^2} L_n(\theta_0) \rightarrow_p 0$ uniformly over $\|\theta_n - \theta_0\| < \tau_n$, where $\tau_n \sim n^{-r}, r > 0$.

By the repeatedly using autoregressive formula, $\log \alpha_t$ can be expressed as

$$\log \alpha_t = \beta_0 \sum_{k=1}^{t-1} \beta_1^{k-1} - \beta_2 \sum_{k=1}^{t-1} \beta_1^{k-1} \exp(-\beta_3 Y_{t-k}) + \beta_1^{t-1} \log \alpha_1^0.$$

We have

$$\frac{\partial \alpha_t}{\partial \beta_0} = \alpha_t \sum_{k=1}^{t-1} \beta_1^{k-1}, \quad \frac{\partial^2 \alpha_t}{\partial \beta_0^2} = \alpha_t \left(\sum_{k=1}^{t-1} \beta_1^{k-1} \right)^2,$$

$$\frac{\partial L_n(\theta)}{\partial \beta_0} = \frac{1}{n} \sum_{t=1}^n \mathbb{I}(Y_t > 0) \sum_{k=1}^{t-1} \beta_1^{k-1} \left[\frac{(\zeta_t + 1) Y_t}{\alpha_t + Y_t} - 1 \right], \quad \frac{\partial^2 L_n(\theta)}{\partial \beta_0^2} = -\frac{1}{n} \sum_{t=1}^n \mathbb{I}(Y_t > 0) \left(\sum_{k=1}^{t-1} \beta_1^{k-1} \right)^2 \frac{(\zeta_t + 1) \alpha_t Y_t}{(\alpha_t + Y_t)^2}.$$

Then

$$\begin{aligned} & \left| \frac{\partial^2}{\partial \beta_0^2} L_n(\theta_n) - \frac{\partial^2}{\partial \beta_0^2} L_n(\theta_0) \right| \\ &= \left| \frac{1}{n} \sum_{t=1}^n \mathbb{I}(Y_t > 0) Y_t \left[\left(\sum_{k=1}^{t-1} \beta_1^{k-1} \right)^2 \frac{(\zeta_t + 1) \alpha_t}{(\alpha_t + Y_t)^2} - \left(\sum_{k=1}^{t-1} (\beta_1^0)^{k-1} \right)^2 \frac{(\zeta_t^0 + 1) \alpha_t^0}{(\alpha_t^0 + Y_t)^2} \right] \right| \\ &\leq \frac{1}{n} \sum_{t=1}^n \mathbb{I}(Y_t > 0) Y_t \left| \left(\sum_{k=1}^{t-1} \beta_1^{k-1} \right)^2 \frac{(\zeta_t + 1) \alpha_t}{(\alpha_t + Y_t)^2} - \left(\sum_{k=1}^{t-1} \beta_1^{k-1} \right)^2 \frac{(\zeta_t^0 + 1) \alpha_t^0}{(\alpha_t^0 + Y_t)^2} \right| \\ &\quad + \frac{1}{n} \sum_{t=1}^n \mathbb{I}(Y_t > 0) Y_t \left| \left(\sum_{k=1}^{t-1} \beta_1^{k-1} \right)^2 \frac{(\zeta_t^0 + 1) \alpha_t^0}{(\alpha_t^0 + Y_t)^2} - \left(\sum_{k=1}^{t-1} (\beta_1^0)^{k-1} \right)^2 \frac{(\zeta_t^0 + 1) \alpha_t^0}{(\alpha_t^0 + Y_t)^2} \right| \\ &=: I + II, \end{aligned}$$

where

$$\begin{aligned}
 I &= \frac{1}{n} \sum_{t=1}^n \mathbb{I}(Y_t > 0) Y_t \left(\sum_{k=1}^{t-1} \beta_1^{k-1} \right)^2 \left| \frac{(\zeta_t + 1)\alpha_t}{(\alpha_t + Y_t)^2} - \frac{(\zeta_t^0 + 1)\alpha_t^0}{(\alpha_t^0 + Y_t)^2} \right| \\
 &= \frac{1}{n} \sum_{t=1}^n \mathbb{I}(Y_t > 0) Y_t \left(\sum_{k=1}^{t-1} \beta_1^{k-1} \right)^2 \frac{|\alpha_t(\zeta_t + 1)(\alpha_t^0 + Y_t)^2 - \alpha_t^0(\zeta_t^0 + 1)(\alpha_t + Y_t)^2|}{(\alpha_t + Y_t)^2(\alpha_t^0 + Y_t)^2} \\
 &\leq \frac{1}{n} \sum_{t=1}^n \frac{\mathbb{I}(Y_t > 0) Y_t \left[|\alpha_t^0(\alpha_t \alpha_t^0 + 2\alpha_t Y_t + Y_t^2)| |\zeta_t - \zeta_t^0| + |\alpha_t \alpha_t^0(\zeta_t^0 + 1) - Y_t^2(\zeta_t + 1)| |\alpha_t - \alpha_t^0| \right]}{(1 - \beta_1)^2 (\alpha_t + Y_t)^2 (\alpha_t^0 + Y_t)^2} \\
 &\leq \frac{1}{n(1 - \beta_1)^2} \sum_{t=1}^n \mathbb{I}(Y_t > 0) \frac{\left[|\alpha_t^0(\alpha_t \alpha_t^0 + 2\alpha_t Y_t + Y_t^2)| |\zeta_t - \zeta_t^0| + |\alpha_t \alpha_t^0(\zeta_t^0 + 1) - Y_t^2(\zeta_t + 1)| |\alpha_t - \alpha_t^0| \right]}{\alpha_t (\alpha_t^0)^2} \\
 &\leq \frac{1}{n(1 - \beta_1)^2} \sum_{t=1}^n \mathbb{I}(Y_t > 0) \left[\left(\frac{Y_t^2}{\alpha_t \alpha_t^0} + \frac{2Y_t}{\alpha_t^0} + 1 \right) |\zeta_t - \zeta_t^0| + \left| \frac{\zeta_t^0 + 1}{\alpha_t^0} - \frac{Y_t^2(\zeta_t + 1)}{\alpha_t (\alpha_t^0)^2} \right| |\alpha_t - \alpha_t^0| \right],
 \end{aligned}$$

It is known that $\{\alpha_t, \zeta_t\}$ is bounded, so $E(Y_t) = \alpha_t / [(\zeta_t - 1)(1 + \gamma)^{\zeta_t}] < \infty, E(Y_t^2) = 2\alpha_t^2 / [(2 - \zeta_t)(1 - \zeta_t)(1 + \gamma)^{\zeta_t}] < \infty$ and $\frac{Y_t^2}{\alpha_t \alpha_t^0} + \frac{2Y_t}{\alpha_t^0} + 1, \frac{\zeta_t^0 + 1}{\alpha_t^0} - \frac{Y_t^2(\zeta_t + 1)}{\alpha_t (\alpha_t^0)^2}$ are bounded too. Therefore, by Lemmas A3(a) and A4(a),

$$I \sim \frac{1}{n} \sum_{t=1}^n (|\zeta_t - \zeta_t^0| + |\alpha_t - \alpha_t^0|) \sim O_p(\tau_n) \rightarrow 0.$$

$$\begin{aligned}
 II &= \frac{1}{n} \sum_{t=1}^n \mathbb{I}(Y_t > 0) \frac{|\zeta_t^0 + 1| \alpha_t^0 |Y_t|}{(\alpha_t^0 + Y_t)^2} \left| \left(\sum_{k=1}^{t-1} \beta_1^{k-1} \right)^2 - \left(\sum_{k=1}^{t-1} (\beta_1^0)^{k-1} \right)^2 \right| \\
 &\leq \frac{2\tau_n}{(1 - C_b)^3} \frac{1}{n} \sum_{t=1}^n \mathbb{I}(Y_t > 0) \frac{|\zeta_t^0 + 1| \alpha_t^0 |Y_t|}{(\alpha_t^0 + Y_t)^2} \leq \frac{2\tau_n}{(1 - C_b)^3} \frac{1}{n} \sum_{t=1}^n \mathbb{I}(Y_t > 0) \left| \frac{\zeta_t^0 + 1}{\alpha_t^0} \right| Y_t \\
 &\leq \frac{2M\tau_n}{(1 - C_b)^3} \frac{1}{n} \sum_{t=1}^n \mathbb{I}(Y_t > 0) Y_t = O_p(\tau_n) \rightarrow 0.
 \end{aligned}$$

where the first inequality comes from the fact that $\sum_{k=1}^t \beta_1^{k-1} < 1/(1 - \beta_1) \leq 1/(1 - C_b)$ and

$$\left| \left(\sum_{k=1}^{t-1} \beta_1^{k-1} \right)^2 - \left(\sum_{k=1}^{t-1} (\beta_1^0)^{k-1} \right)^2 \right| \leq \left| \left(\frac{1}{1 - \beta_1} \right)^2 - \left(\frac{1}{1 - \beta_1^0} \right)^2 \right| \leq \frac{2\tau_n}{(1 - C_b)^3} = O_p(\tau_n),$$

where C_b is a constant and $0 < C_b < 1$. The last inequality of II shows the boundedness of $\{\alpha_t^0, \zeta_t^0\}$, so there exists $|(\zeta_t^0 + 1)/\alpha_t^0| \leq M$, and $E(Y_t) = \alpha_t / [(\zeta_t - 1)(1 + \gamma)^{\zeta_t}] < \infty$. \square

Lemma A6. Under the same conditions as in Theorem 2, two positive constants C and $C_b < 1$ exist such that for all $\theta \in \Theta$ and $t \geq 1$,

$$\begin{aligned}
 (a) \quad &|\alpha_t - \tilde{\alpha}_t| \leq CC_b^{t-1}, (b) \quad \left| \frac{\partial \alpha_t}{\partial \Phi} - \frac{\partial \tilde{\alpha}_t}{\partial \Phi} \right| \leq CtC_b^{t-1}, (c) \quad \left| \frac{\partial^2 \alpha_t}{\partial \Phi_i \partial \Phi_j} - \frac{\partial^2 \tilde{\alpha}_t}{\partial \Phi_i \partial \Phi_j} \right| \leq Ct^2 C_b^{t-1}, \\
 (d) \quad &|\zeta_t - \tilde{\zeta}_t| \leq CC_b^{t-1}, (e) \quad \left| \frac{\partial \zeta_t}{\partial \Psi} - \frac{\partial \tilde{\zeta}_t}{\partial \Psi} \right| \leq CtC_b^{t-1} (f) \quad \left| \frac{\partial^2 \zeta_t}{\partial \Psi_i \partial \Psi_j} - \frac{\partial^2 \tilde{\zeta}_t}{\partial \Psi_i \partial \Psi_j} \right| \leq Ct^2 C_b^{t-1}.
 \end{aligned}$$

Proof. The proof is omitted because it follows from direct calculation. \square

Lemma A7. Under the same conditions as in Theorem A2,

- (a) $\frac{\partial^2}{\partial\theta_i\partial\theta_j}\tilde{L}_n(\theta) \rightarrow_p -m_{\theta_i\theta_j}(\theta_0)$, uniformly over $\|\theta - \theta_0\| < \tau_n$, where $\tau_n \sim n^{-r}$, $r > 0$,
- (b) $(\tau_n^*)^{-1}\left(\frac{\partial}{\partial\theta}\tilde{L}_n(\theta_0) - \frac{\partial}{\partial\theta}L_n(\theta_0)\right) \rightarrow_p 0$ if $\tau_n^*n \rightarrow \infty$.

Proof. (a) By the result of Lemma A5, we have $\frac{\partial^2}{\partial\theta_i\partial\theta_j}L_n(\theta) - \frac{\partial^2}{\partial\theta_i\partial\theta_j}L_n(\theta_0) \rightarrow_p 0$ uniformly over $\|\theta - \theta_0\| < \tau_n$, so we need to prove that $\frac{\partial^2}{\partial\theta_i\partial\theta_j}\tilde{L}_n(\theta) - \frac{\partial^2}{\partial\theta_i\partial\theta_j}L_n(\theta) \rightarrow_p 0$ uniformly under the claimed region. Based on Lemma A6, we can use the same method as in the proof of Lemma A5, then it is omitted.

(b) Here we only prove for $\frac{\partial}{\partial\beta_0}\tilde{L}_n(\theta_0)$, as other proofs are similar.

$$\begin{aligned} & \left| \frac{1}{\tau_n^*} \left(\frac{\partial}{\partial\beta_0}\tilde{L}_n(\theta_0) - \frac{\partial}{\partial\beta_0}L_n(\theta_0) \right) \right| = \left| \frac{1}{n\tau_n^*} \sum_{t=1}^n \mathbb{I}(Y_t > 0) \left(\sum_{k=1}^{t-1} \beta_1^{k-1} \right) \left(\frac{\tilde{\xi}_t Y_t - \tilde{\alpha}_t}{\tilde{\alpha}_t + Y_t} - \frac{\xi_t Y_t - \alpha_t}{\alpha_t + Y_t} \right) \right| \\ & \leq \frac{1}{n\tau_n^*} \sum_{t=1}^n \left(\sum_{k=1}^{t-1} \beta_1^{k-1} \right) \mathbb{I}(Y_t > 0) \left| \frac{\tilde{\xi}_t Y_t - \tilde{\alpha}_t}{\tilde{\alpha}_t + Y_t} - \frac{\xi_t Y_t - \alpha_t}{\alpha_t + Y_t} \right| \\ & \leq \frac{1}{n\tau_n^*(1 - \beta_1)} \sum_{t=1}^n \mathbb{I}(Y_t > 0) \left| \frac{\tilde{\xi}_t Y_t - \tilde{\alpha}_t}{\tilde{\alpha}_t + Y_t} - \frac{\xi_t Y_t - \alpha_t}{\alpha_t + Y_t} \right| \\ & = \frac{1}{n\tau_n^*(1 - \beta_1)} \sum_{t=1}^n \mathbb{I}(Y_t > 0) \left| \frac{\tilde{\xi}_t Y_t - \tilde{\alpha}_t}{\tilde{\alpha}_t + Y_t} - \frac{\tilde{\xi}_t Y_t - \tilde{\alpha}_t}{\alpha_t + Y_t} + \frac{\tilde{\xi}_t Y_t - \tilde{\alpha}_t}{\alpha_t + Y_t} - \frac{\xi_t Y_t - \alpha_t}{\alpha_t + Y_t} \right| \\ & = \frac{1}{n\tau_n^*(1 - \beta_1)} \sum_{t=1}^n \mathbb{I}(Y_t > 0) \left| \frac{-(\tilde{\xi}_t Y_t - \tilde{\alpha}_t)(\tilde{\alpha}_t - \alpha_t)}{(\tilde{\alpha}_t + Y_t)(\alpha_t + Y_t)} + \frac{Y_t(\tilde{\xi}_t - \xi_t)}{\alpha_t + Y_t} - \frac{\tilde{\alpha}_t - \alpha_t}{\alpha_t + Y_t} \right| \\ & \leq \frac{C}{n\tau_n^*(1 - \beta_1)} \sum_{t=1}^n \mathbb{I}(Y_t > 0) \left[C_b^{t-1} \left(\frac{|\tilde{\xi}_t Y_t - \tilde{\alpha}_t|}{\alpha_L^2} + \frac{1}{\alpha_L} + 1 \right) \right] \\ & \leq \frac{C}{n\tau_n^*(1 - \beta_1)} \sum_{t=1}^n \mathbb{I}(Y_t > 0) \left[C_b^{t-1} \left(\frac{\tilde{\xi}_t Y_t}{\alpha_L^2} + \frac{\alpha_U}{\alpha_L^2} + \frac{1}{\alpha_L} + 1 \right) \right], \end{aligned}$$

where the second-to-last inequality comes from Lemma A6 (a) and (d). Next, we need to prove the boundedness of $\sum_{t=1}^n C_b^{t-1} \tilde{\xi}_t Y_t$.

$$\begin{aligned} & \mathbb{E} \left(\sum_{t=1}^n C_b^{t-1} \tilde{\xi}_t Y_t \right) \leq \mathbb{E} \left(\sum_{t=1}^{\infty} C_b^{t-1} \tilde{\xi}_t Y_t \right) = \sum_{t=1}^{\infty} C_b^{t-1} \mathbb{E}(\tilde{\xi}_t Y_t) \\ & = \sum_{t=1}^{\infty} C_b^{t-1} \mathbb{E}[\mathbb{E}(\tilde{\xi}_t Y_t | \tilde{\xi}_t)] = \sum_{t=1}^{\infty} C_b^{t-1} \mathbb{E} \left[\frac{\alpha_t \tilde{\xi}_t}{(\tilde{\xi}_t - 1)(1 + u)^{\tilde{\xi}_t}} \right] \leq \sum_{t=1}^{\infty} C_b^{t-1} \mathbb{E} \left(\frac{\xi_t u \alpha_t}{\xi_t - 1} \right) < \infty. \end{aligned}$$

Therefore, when $n\tau_n^* \rightarrow \infty$,

$$\frac{1}{\tau_n^*} \left[\frac{\partial}{\partial\beta_0}\tilde{L}_n(\theta_0) - \frac{\partial}{\partial\beta_0}L_n(\theta_0) \right] \rightarrow_p 0.$$

□

Lemma A8. Under the same conditions as 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 at θ_0 .

Proof. The proof of the lemma is similar Zhao et al. (2018) [31], so it is omitted. □

Proof of Theorem 2. Theorem 2 can be proved by Lemmas A6–A8, and the details can be seen in Shen et al. (2020) [27]. The proof is similar to Shen et al. (2020) [27], so it is omitted. The main difference is that we denote $f_n(t, \mathbf{y}) = \tau_n^{-2} \tilde{L}_n(\beta_0 + \tau_n t, \Phi^0 + \tau_n \mathbf{y})$, $\Phi^0 = (\beta_1^0, \beta_2^0, \beta_3^0, \gamma_0^0, \gamma_1^0, \gamma_2^0, \gamma_3^0)$, where $t \in R$, $\mathbf{y} \in R^7$. \square

Proof of Theorem 3. Theorem 3 can be proved by Lemmas A7 and A8, and the detail can be seen in Zhao et al. (2018) [31]. The proof is similar to Zhao et al. (2018) [31], so it is omitted. \square

References

- 2021 World Air Quality Report. 2022. Available online: <https://www.iqair.com/world-air-quality-report> (accessed on 9 April 2022).
- Sun, H.; Yang, X.; Leng, Z. Research on the spatial effects of haze pollution on public health: Spatial–temporal evidence from the Yangtze River Delta urban agglomerations, China. *Environ. Sci. Pollut. Res.* **2022**, 1–20. doi: 10.1007/s11356-022-19017-0. [[CrossRef](#)] [[PubMed](#)]
- Shen, W.T.; Yu, X.; Zhong, S.B.; Ge, H.R. Population health effects of air pollution: Fresh evidence from China health and retirement longitudinal survey. *Front. Public Health* **2021**, *9*, 779552. [[CrossRef](#)] [[PubMed](#)]
- Maji, K.J.; Dikshit, A.K.; Arora, M.; Deshpande, A. Estimating premature mortality attributable to PM2.5 exposure and benefit of air pollution control policies in China for 2020. *Sci. Total Environ.* **2018**, *612*, 683–693. [[CrossRef](#)] [[PubMed](#)]
- 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. Assoc.* **2018**, *68*, 265–287. [[CrossRef](#)] [[PubMed](#)]
- Raaschou-Nielsen, O.; Andersen, Z.J.; Beelen, R.; Samoli, E.; Stafoggia, M.; Weinmayr, G.; Hoffmann, B.; Fischer, P.; Nieuwenhuijsen, M.J.; Brunekreef, B.; et al. Air pollution and lung cancer incidence in 17 European cohorts: Prospective analyses from the European Study of Cohorts for Air Pollution Effects (ESCAPE). *Lancet Oncol.* **2013**, *14*, 813–822. [[CrossRef](#)]
- Gan, W.Q.; Davies, H.W.; Koehoorn, M.; Brauer, M. Association of long-term exposure to community noise and traffic-related air pollution with coronary heart disease mortality. *Am. J. Epidemiol.* **2012**, *175*, 898–906. [[CrossRef](#)] [[PubMed](#)]
- Brunekreef, B.; Beelen, R.; Hoek, G.; Schouten, L.; Bausch-Goldbohm, S.; Fischer, P.; Armstrong, B.; Hughes, E.; Jerrett, M.; van den Brandt, P. Effects of long-term exposure to traffic-related air pollution on respiratory and cardiovascular mortality in the Netherlands: The NLCS-AIR study. *Res. Rep.* **2009**, *139*, 5–71.
- Kim, S.Y.; Sheppard, L.; Kim, H. Health effects of long-term air pollution: Influence of exposure prediction methods. *Epidemiology* **2009**, *20*, 442–450. [[CrossRef](#)]
- Cheng, Z.; Li, L.; Liu, J. Identifying the spatial effects and driving factors of urban PM2.5 pollution in China. *Ecol. Indic.* **2017**, *82*, 61–75. [[CrossRef](#)]
- Gan, T.; Yang, H.; Liang, W.; Liao, X. Do economic development and population agglomeration inevitably aggravate haze pollution in China? New evidence from spatial econometric analysis. *Environ. Sci. Pollut. Res.* **2021**, *28*, 5063–5079. [[CrossRef](#)]
- Ma, R.; Wang, C.; Jin, Y.; Zhou, X. Estimating the effects of economic agglomeration on haze pollution in Yangtze River Delta China using an econometric analysis. *Sustainability* **2019**, *11*, 1893. [[CrossRef](#)]
- Xie, Q.; Xu, X.; Liu, X. Is there an EKC between economic growth and smog pollution in China? New evidence from semiparametric spatial autoregressive models. *J. Clean. Prod.* **2019**, *220*, 873–883. [[CrossRef](#)]
- Zhang, X.; Xu, X.; Ding, Y.; Liu, Y.; Zhang, H.; Wang, Y.; Zhong, J. The impact of meteorological changes from 2013 to 2017 on PM2.5 mass reduction in key regions in China. *Sci. China Earth Sci.* **2019**, *62*, 1885–1902. [[CrossRef](#)]
- Fontes, T.; Li, P.; Barros, N.; Zhao, P. Trends of PM2.5 concentrations in China: A long term approach. *J. Environ. Manag.* **2017**, *196*, 719–732. [[CrossRef](#)]
- Pui, D.Y.H.; Chen, S.C.; Zuo, Z. PM2.5 in China: Measurements, sources, visibility and health effects, and mitigation. *Particuology* **2014**, *13*, 1–26. [[CrossRef](#)]
- Liang, X.; Zou, T.; Guo, B.; Li, S.; Zhang, H.; Zhang, S.; Huang, H.; Chen, S.X. Assessing Beijing’s PM2.5 pollution: Severity, weather impact, APEC and winter heating. *Proc. R. Soc. A* **2015**, *471*, 20150257. [[CrossRef](#)]
- Zhang, S.; Guo, B.; Dong, A.; He, J.; Xu, Z.; Chen, S.X. Cautionary tales on air-quality improvement in Beijing. *Proc. R. Soc. A* **2017**, *473*, 20170457. [[CrossRef](#)]
- Chen, L.; Guo, B.; Huang, J.; He, J.; Wang, H.; Zhang, S.; Chen, S.X. Assessing air-quality in Beijing–Tianjin–Hebei Region: The method and mixed tales of PM2.5 and O3. *Atmos. Environ.* **2018**, *193*, 290–301. [[CrossRef](#)]
- Wu, H.; Zheng, X.; Zhu, J.; Lin, W.; Zheng, H.; Chen, X.; Wang, W.; Wang, Z.; Chen, S.X. Improving PM2.5 forecasts in China using an initial error transport mode. *Environ. Sci. Technol.* **2020**, *54*, 10493–10501. [[CrossRef](#)]
- Wan, Y.; Xu, M.; Huang, H.; Xi Chen, S. A spatio-temporal model for the analysis and prediction of fine particulate matter concentration in Beijing. *Environmetrics* **2020**, *32*, e2648. [[CrossRef](#)]
- Zhu, Y.; Liang, Y.; Chen, S. Assessing local emission for air pollution via data experiments. *Atmos. Environ.* **2021**, *252*, 118323. [[CrossRef](#)]
- Wang, J.; Cohan, D.S.; Xu, H. Spatiotemporal ozone pollution LUR models: Suitable statistical algorithms and time scales for a megacity scale. *Atmos. Environ.* **2020**, *237*, 117671. [[CrossRef](#)]

24. Pickands, J. Statistical inference using extreme order statistics. *Ann. Stat.* **1975**, *3*, 119–131.
25. Tencaliec, P.; Favre, A.C.; Naveau, P.; Prieur, C.; Nicolet, G. Flexible semiparametric generalized Pareto modeling of the entire range of rainfall amount. *Environmetrics* **2020**, *31*, e2582. [[CrossRef](#)]
26. Gharib, A.; Davies, E.G.R.; Goss, G.G.; Faramarzi, M. Assessment of the combined effects of threshold selection and parameter estimation of generalized Pareto distribution with applications to flood frequency analysis. *Water* **2017**, *9*, 692. [[CrossRef](#)]
27. Shen, Z.Y.; Chen, Y.; Shi, R.X. Modeling tail index with autoregressive conditional Pareto model. *J. Bus. Econ. Stat.* **2022**, *40*, 458–466. [[CrossRef](#)]
28. Chen, Y.; Yu, W. Setting the margins of Hang Seng Index Futures on different positions using an APARCH-GPD Model based on extreme value theory. *Phys. A Stat. Mech. Its Appl.* **2020**, *544*, 123207. [[CrossRef](#)]
29. Park, E.; Brorsen, B.W.; Harri, A. Using Bayesian Kriging for spatial smoothing in crop insurance rating. *Am. J. Agric. Econ.* **2019**, *101*, 330–351. [[CrossRef](#)]
30. Liu, X.H.; Zhang, X.; Xue, J.Y. Fraud risk measurement of basic medical insurance for urban and rural residents in China. *Econ. Comput. Econ. Cybern. Stud. Res.* **2019**, *53*, 277–296. [[CrossRef](#)]
31. Zhao, Z.; Zhang, Z.; Chen, R. Modeling maxima with autoregressive conditional Fréchet model. *J. Econom.* **2018**, *207*, 325–351. [[CrossRef](#)]
32. Chavez-Demoulin, V.; Embrechts, P.; Sardy, S. Extreme-quantile tracking for financial time series. *J. Econom.* **2014**, *181*, 44–52. [[CrossRef](#)]
33. Kelly, B.; Jiang, H. Tail risk and asset prices. *Rev. Financ. Stud.* **2014**, *27*, 2841–2871. [[CrossRef](#)]
34. Massacci, D. Tail risk dynamics in stock returns: Links to the macroeconomy and global markets connectedness. *Manag. Sci.* **2017**, *63*, 3072–3089. [[CrossRef](#)]
35. Deng, L.; Yu, M.X.; Zhang, Z.J. Statistical learning of the worst regional smog extremes with dynamic conditional modeling. *Atmosphere* **2020**, *11*, 665. [[CrossRef](#)]
36. Choulakian, V.; Stephens, M.A. Goodness-of-fit tests for the generalized Pareto distribution. *Technometrics* **2001**, *43*, 478–484. [[CrossRef](#)]
37. Bermudez, P.Z.; Turkman, M.A.A.; Turkman, K.F. A predictive approach to tail probability estimation. *Extremes* **2001**, *4*, 295–314. [[CrossRef](#)]
38. Bader, B.; Yan, J.; Zhang, X.B. Automated threshold selection for extreme value analysis via ordered goodness-of-fit tests with adjustment for false discovery rate. *Ann. Appl. Stat.* **2018**, *12*, 310–329. [[CrossRef](#)]
39. Yang, X.; Zhang, J.; Ren, W.X. Threshold selection for extreme value estimation of vehicle load effect on bridges. *Int. J. Distrib. Sens. Netw.* **2018**, *14*, 1550147718757698. [[CrossRef](#)]
40. Schneider, L.F.; Krajina, A.; Krivobokova, T. Threshold selection in univariate extreme value analysis. *Extremes* **2021**, *24*, 881–913. [[CrossRef](#)]
41. Boznar, M.; Lesjak, M.; Mlakar, P. A neural network-based method for short-term predictions of ambient SO₂ concentrations in highly polluted industrial areas of complex terrain. *Atmos. Environ. Part B Urban Atmos.* **1993**, *27*, 221–230. [[CrossRef](#)]
42. Neagu, C.D.; Avouris, N.; Kalapanidas, E.; Palade, V. Neural and neuro-fuzzy integration in a knowledge-based system for air quality prediction. *Appl. Intell.* **2002**, *17*, 141–169. [[CrossRef](#)]
43. Esfandani, M.A.; Nematzadeh, H. Predicting air pollution in tehran: Genetic algorithm and back propagation neural network. *J. Data Min.* **2016**, *4*, 49–54.
44. Amarpuri, L.; Yadav, N.; Kumar, G.; Agrawal, S. Prediction of CO₂ emissions using deep learning hybrid approach: A case study in Indian context. In Proceedings of the 2019 Twelfth International Conference on Contemporary Computing (IC3), Noida, India, 8–10 August 2019.
45. Menéndez García, L.A.; Sánchez Lasheras, F.; García Nieto, P.J.; Álvarez de Prado, L.; Bernardo Sánchez, A. Predicting Benzene concentration using machine learning and time series algorithms. *Mathematics* **2020**, *8*, 2205. [[CrossRef](#)]
46. Sánchez-Pérez, J.F.; Mena-Requena, M.R.; Cánovas, M. Mathematical modeling and simulation of a gas emission source using the network simulation method. *Mathematics* **2020**, *8*, 1996. [[CrossRef](#)]
47. Sayeed, A.; Lops, Y.; Choi, Y.; Jung, J.; Salman, A.K. Bias correcting and extending the PM forecast by CMAQ up to 7 days using deep convolutional neural networks. *Atmos. Environ.* **2021**, *253*, 118376. [[CrossRef](#)]
48. Kang, S.; Song, J. Parameter and quantile estimation for the generalized Pareto distribution in peaks over threshold framework. *J. Korean Stat. Soc.* **2017**, *46*, 487–501. [[CrossRef](#)]
49. Balkema, A.A.; De Haan, L. Residual life time at great age. *Ann. Probab.* **1974**, *2*, 792–804. [[CrossRef](#)]
50. Davison, A.C.; Smith, R.L. Models for exceedances over high thresholds. *J. R. Stat. Soc. Ser. B (Methodol.)* **1990**, *52*, 393–425. [[CrossRef](#)]