

# Air Quality Risk Measurement Based on CAViaR Model: A Case Study of PM10 in Beijing

Peng Sun, Fuming Lin\*

School of Mathematics and Statistics, Sichuan University of Science and Engineering, Zigong, China

Email: sunpeng1813@163.com, \*215966774@qq.com

**How to cite this paper:** Sun, P. and Lin, F.M. (2023) Air Quality Risk Measurement Based on CAViaR Model: A Case Study of PM10 in Beijing. *Journal of Applied Mathematics and Physics*, 11, 2879-2887. <https://doi.org/10.4236/jamp.2023.1110189>

**Received:** September 9, 2023

**Accepted:** October 17, 2023

**Published:** October 20, 2023

Copyright © 2023 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

Air pollution control has always been a global challenge, and significant progress has been made in recent years in controlling air pollutants. However, in some major cities, air pollutant concentrations still exceed the standards. Some scholars have used linear models or conditional autoregressive iterative models to apply the VaR method to predict pollutant concentrations. However, traditional methods based on quantile regression estimation can lead to inadequate risk estimates. Therefore, we propose a method based on the Conditional Autoregressive Value at Risk (CAViaR) model, which uses the  $k$ th power expectile regression to estimate VaR. This method does not specify the type of the distribution of data, is easier to calculate the asymptotic variance, more sensitive to extreme values. Applying our method to the data of PM10 in Beijing, we investigate the fitting effects in the case of  $k = 1$ ,  $k = 2$ , and  $k = 1.9$  through predictive tests. The results show that the  $k$ th power expectile regression estimates are better than quantile and expectile regression estimates to some extent.

## Keywords

PM10, VaR,  $k$ th Power Expectile, CAViaR Model

## 1. Introduction

Air pollution control has always been a global issue, and there has been significant progress in controlling air pollution in recent years. However, in some major cities, air pollution still exceeds the standards. Most studies focus on the prediction of PM10, but they often struggle with extreme values. Globally, 7 million people die prematurely each year due to air pollution. The main culprits are nitrogen dioxide (NO<sub>2</sub>), ozone (O<sub>3</sub>), and particulate matter with a diameter of 10  $\mu$ m or less (PM10 for short). PM10 is mainly generated from road traffic and

contains a mixture of metals, phosphates, nitrates, sulfates, as well as inorganic and organic carbon (Shahraiyini and Sodoudi [1]). The VaR (Value at Risk) method is a commonly used risk measurement method that can be used to assess the risk level of financial markets. Applying the VaR method to the prediction of atmospheric pollutant concentrations can help governments and relevant departments better understand and manage atmospheric pollution risks. Atmospheric pollution is influenced by numerous factors, including meteorological conditions, emission sources, and geographical location. A single model may not be able to fully capture these complex influencing factors. Therefore, when applying the VaR method to predict atmospheric pollutant concentrations, it is necessary to consider multiple models and factors comprehensively in order to improve the accuracy and reliability of the predictions.

Masseran [2] proposed a suitable model by combining autoregressive moving average (ARMA) with generalized autoregressive conditional heteroskedasticity (ARCH/GARCH) to overcome the problematic volatility effects in PM10 data. Wang [3] introduced a new mixed GARCH (generalized autoregressive conditional heteroskedasticity) method for the prediction models of ARIMA (autoregressive integrated moving average) and SVM (support vector machine). Veleva and Zeleva [4] constructed univariate ARIMA models based on natural logarithm transformation values and statistically evaluated mixed ARIMA-gjr-garch and mixed ARIMA-egarch models. Suleiman [5] proposed a new approach based on machine learning (ML) models to assess the effectiveness of roadside PM10 and PM2.5 reduction schemes.

Karimian [6] compared three machine learning methods, namely Multivariate Additive Regression Trees (MART), Deep Feed forward Neural Networks (DFNN), and Long Short-Term Memory (LSTM), to predict PM2.5 concentrations at different time intervals, and found that the LSTM model performed the best. Czernecki [7] evaluated the feasibility of short-term PM<sub>x</sub> prediction using machine learning (ML) and identified the main meteorological covariates. Four machine learning models were tested: AIC-based stepwise regression, two tree-based algorithms (Random Forest and XGBoost), and neural networks. Cai *et al.* [8] used the Land Use Regression model (LUR) to predict fine particulate matter (PM2.5), black carbon (BC), and nitrogen dioxide (NO<sub>2</sub>). Yang [9] proposed a hybrid framework combining PM10 prediction models with evaluation models to predict, evaluate, and warn the impact of PM10 on public health. Manganelli and Engle [10] used Monte Carlo simulations to demonstrate that CAViaR outperforms most indirect VaR (Value at Risk) strategies when dealing with distributions exhibiting heavy tails.

Jiang, Lin, and Zhou [11] propose a method called the *k*th power expectile regression, which to some extent unifies quantile regression and expectile regression, or generalizes the latter two regression methods. They also point out that under certain conditions, the *k*th power expectile regression estimate is a maximum likelihood estimate. Combining Engel's model, we use the *k*th power ex-

pectile regression estimate to predict meteorological data and obtain the forecast value of meteorological data VaR. Based on the predictive performance, the  $k$ th power expectile estimate method outperforms the cases where  $k = 1$  (quantile) and  $k = 2$  (expectile).

## 2. VaR and $k$ th Power Expectile Estimation

The VaR method has been widely used in various fields, and many scholars have applied this estimation method to air quality early warning monitoring. Lidia [12] used the VaR method through the CAViaR model to control air pollution and found that the extended CAViaR model outperformed the standard CAViaR model. In finance, VaR refers to the maximum potential loss that may occur in a certain time period for holding a certain security or asset portfolio at a certain confidence level. Due to its simplicity and ease of understanding, VaR calculation has been favored by investigators, especially those in the field of financial statistics. We further combine the  $k$ th power expectile regression method and the CAViaR model to estimate VaR.

We consider the meteorological time series as  $\{R_t\}_{t=1}^n$ ,  $R_t = (P_t - P_{t-1})/P_{t-1}$ , and  $P_t$  represents the daily average concentration of PM10. The marginal distribution function is  $F_{R_t}$ . Given a  $\alpha (\alpha \in (0,1))$ , the VaR at the significant level  $\alpha$  at time  $t$  is defined as

$$\text{VaR} = \inf \{u : F_{R_t}(u) \geq 1 - \alpha\}$$

In this paper, we consider the PM10 daily growth rate series  $\{r_t\}$ , where  $r_t = -R_t$ . Thus, the lower risk VaR value at the significant level  $\alpha$  for variable  $r_t$  is the  $\alpha$  quantile.

The traditional approach is based on quantiles to calculate VaR, which does not fully consider the overall information in the data distribution and can lead to inadequate risk estimation. In contrast, we use  $k$ th power expectile regression method. The  $k$ th power expectile is defined as the minimization of the loss function below:

$$f_k(\tau) = \arg \min_f E(|\tau - I(r \leq f)| |r - f|^k)$$

It is easy to get its estimate as:

$$\hat{f}_k(\tau) = \arg \min_f T^{-1} \sum_t |\tau - I(r \leq f)| |r - f|^k,$$

where  $\tau \in (0,1)$ ,  $I$  is an indicator function. When  $r \leq f, I = 1$ ; otherwise,  $I = 0$ .

When  $k = 1$ , the  $k$ th power expectile regression estimator can be regarded as the quantile regression estimator, and it has an advantage in computing asymptotic variance. When  $k = 2$ , the  $k$ th power expectile regression estimator can be regarded as the expectile regression estimator, and it has a better requirement for moments than the expectile regression.

Compared to expectiles, the  $k$ th power expectile regression method has weak-

er requirements on moments and easier calculation of variance. In this paper, we use the  $k$ th power expectile regression to estimate the CAViaR model.

### 3. Model and Its Estimation

Existing research has shown that PM10 growth rate exhibits characteristics such as volatility clustering and asymmetric volatility, very similar to financial data. Conditional autoregressive models have been widely used to study such phenomena. Building upon the CAViaR model studied by Engle and Manganelli [10], Taylor [13] and Kuan, Yeh, and Hsu [14] estimated the traditional conditional autoregressive model using expectile regression, resulting in a class of CARE-Expectile models. Lidia [12], on the other hand, estimated the CAViaR model using quantile regression. Given these findings, we consider estimating the CAViaR models using  $k$ th power expectile regression. One of CAViaR models is symmetric absolute value CARE model is

$$f(\beta_\tau) = \beta_1 + \beta_2 f_{t-1}(\tau) + \beta_3 |r_{t-1}|.$$

Among them,  $f(\beta_\tau)$  is the  $k$ th power expectile,  $\beta_i$  is the parameter to be estimated, and  $r_t$  is the daily growth rate of PM10. We will input CARE model into the loss function  $f_k(\tau)$  to obtain an estimate of parameter  $\beta_i$ . The specific algorithm steps are as follows: first, generate one hundred thousand sets of parameter vectors that follow a uniform distribution within the range of 0 to 1 (or  $-1$  to 1), and use these parameter vectors to train the function to obtain ten sets of parameter vectors that minimize the loss function. These ten sets of vectors are then used as initial values for the simulated annealing algorithm to iteratively find the parameter vector that minimizes the loss function. The final estimated values of the model are the parameter vectors that minimize the loss function among these ten sets. We then input the estimated optimal parameter vector into the lagged VaR calculation model to obtain the percentile. Sometimes, for the sake of clarity, we need to transform the  $k$ th power expectile into quantile. When  $k = 2$ , existing studies have used the correspondence between expectiles and quantiles under the assumption of a normal distribution to measure risk, which can lead to significant specification errors. In complex financial markets, assuming the distribution of data artificially can result in overestimation or underestimation of risk. In this study, we modify a method proposed by Efron [15] to obtain the quantile model based on the  $k$ th power model in a linear model framework. Efron [15] calculated the ratio of the number of observations below the expectile regression function to the total number of observations. Precisely we extend this to non-parametric method to the CAViaR model in this study. First, given a significant level  $\alpha$  and a chosen interval  $I \in (0.5, 1)$ . Choose  $\tau_i \in I$  and estimate the CAViaR model. When calculating VaR in the CAViaR model, we obtain the model estimate

$$f(\hat{\beta}_\tau) = \hat{\beta}_1 + \hat{\beta}_2 f_{t-1}(\tau) + \hat{\beta}_3 |r_{t-1}|.$$

By taking 100 lag orders of the CAViaR model, we obtain an approximate li-

near model

$$f_t(\tau_i) \approx \sum_{n=1}^{100} \hat{\beta}_2^{n-1} (\hat{\beta}_1 + \hat{\beta}_3 |r_{t-n}|) = h(r_{t-1}, \dots, r_{t-100}). \tag{3.1}$$

Let  $\chi_t(\tau_i) = h(r_{t-1}, \dots, r_{t-100})$ , if  $T$  is the sample size, we have the following set of expressions

$$\begin{aligned} \chi_{101}(\tau_i) &= h(r_{100}, \dots, r_1) \\ &\vdots \\ \chi_T(\tau_i) &= h(r_{T-1}, \dots, r_{T-100}) \\ \chi_{T+1}(\tau_i) &= h(r_T, \dots, r_{T-99}). \end{aligned}$$

By substituting the PM10 data into the above equations, we obtain the sequence  $\{K_{j+1}\}_{j=100}^T$ , where

$$K_{j+1} = \begin{cases} 1, & r_{j+1} < \chi_{j+1}(\tau_i) \\ 0, & r_{j+1} \geq \chi_{j+1}(\tau_i) \end{cases}$$

Let  $T_1$  denote the number of non-zero elements in the sequence  $\{K_j\}_{j=101}^T$ .

We obtain  $\alpha = T_1 / (T - 100)$ , which is essentially a function of  $\alpha$ . According to the definition of quantiles, Equation (3.1) is exactly the quantile regression equation with weights  $\alpha$ .

In this paper, we used backtesting to evaluate the effectiveness of the proposed method. Specifically, we used the negative order of the daily growth rate of PM10 concentration to calculate the upper risk of the growth rate, and equivalently calculated the lower risk of the growth rate. When the estimated value of the sequence exceeds the VaR, we refer to it as a ‘‘VaR violation’’. Define  $\{J_{t+1}\}_{t=101}^T$  as follows:

$$J_{t+1} = \begin{cases} 1, & r_{t+1} \geq VaR_{t+1} \\ 0, & r_{t+1} < VaR_{t+1} \end{cases}$$

This article evaluates the forecasting performance of the VaR method using two testing indicators: failure rate and relative error rate.

The failure rate is defined as  $\hat{\delta} = (N_1 / N) \cdot 100\%$ ,  $N_1$  represents the number of occurrences of value 1 in sequence  $\{J_{t+1}\}_{t=101}^T$ . The relative error rate is defined as  $RE = |(\hat{\delta} - \mu) / \mu|$ , and the smaller the value of  $RE$ , the better the risk measurement effect of the corresponding VaR method, where  $\mu$  is the significance level.

### 4. Empirical Analysis

The data for this study is from the atmospheric pollutant concentration data of the Beijing Environmental Monitoring Station. The daily average concentration of PM10 in the Wanshouxigong area was chosen from the data, and median interpolation was performed on the data with zero daily average concentration to obtain the daily growth rate of PM10 in the Wanshouxigong area. The sample

interval is 2000 data from September 2017 to March 2023. The descriptive statistics of the data are shown in **Table 1**, and the daily growth rate trend of PM10 is shown in **Figure 1**. Since we study the upper risk, we take the daily growth rate of Mansu Nishimiya as negative for research, and there is no need to take negative VaR. According to the descriptive statistical results, the daily growth rate of pm10 is similar to financial data, with fluctuations and asymmetry. According to the statistical results of kurtosis and skewness in **Table 1**, the skewness of the daily growth rate of Wanshou Xigong is negative and the kurtosis is greater than 3, which indicates that the PM10 has left-skew and peak characteristics and the sequence is asymmetric.

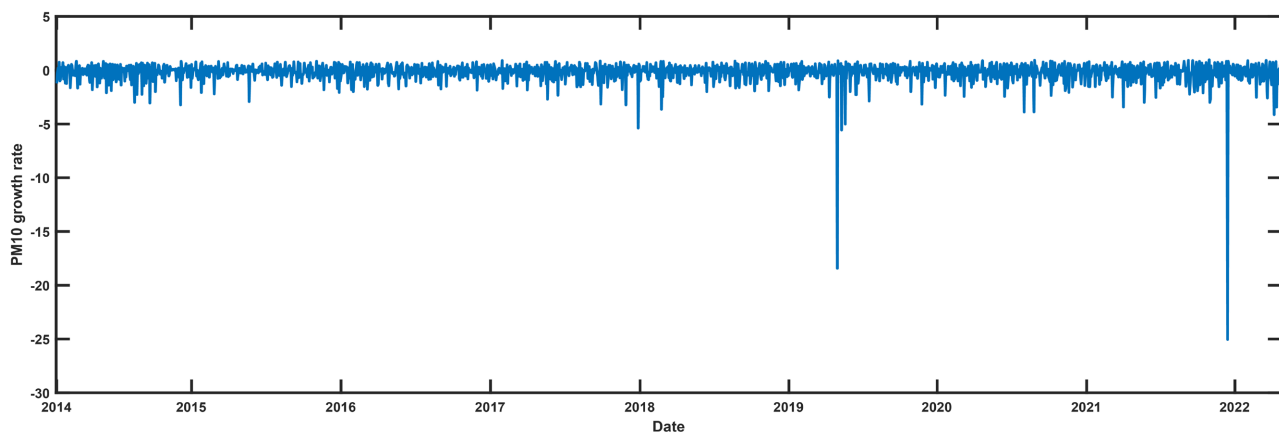
**Figure 2** and **Figure 3** have the same violation days and similar fitting effect, **Figure 4** has slightly less violation days than **Figure 2** and **Figure 3**, and better fitting effect, indicating that our estimation method has more advantages than quantile regression estimation and expectile regression estimation when the distribution is unknown.

### 5. Conclusion Remark

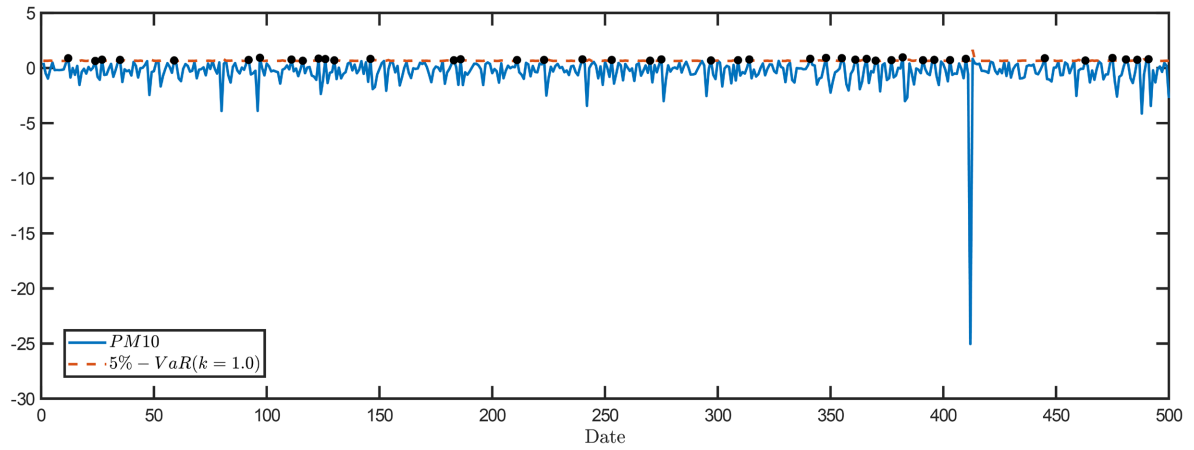
It can be seen from the fitting effect that, based on the CAViaR model, the effect of VaR estimation is similar when  $k = 1$  (quantile) and  $k = 2$  (expectile), while the fitting effect of the  $k$ th power expectile regression estimation of VaR when  $k = 1.9$  is better than the other two methods. The proposed method has some merits at least in the following aspects: forecasting effect, its excellent properties such as not making any assumptions about the distribution of data, easy calculation of

**Table 1.** The description statistics of PM10 daily growth rate of Wanshou West Palace, Beijing.

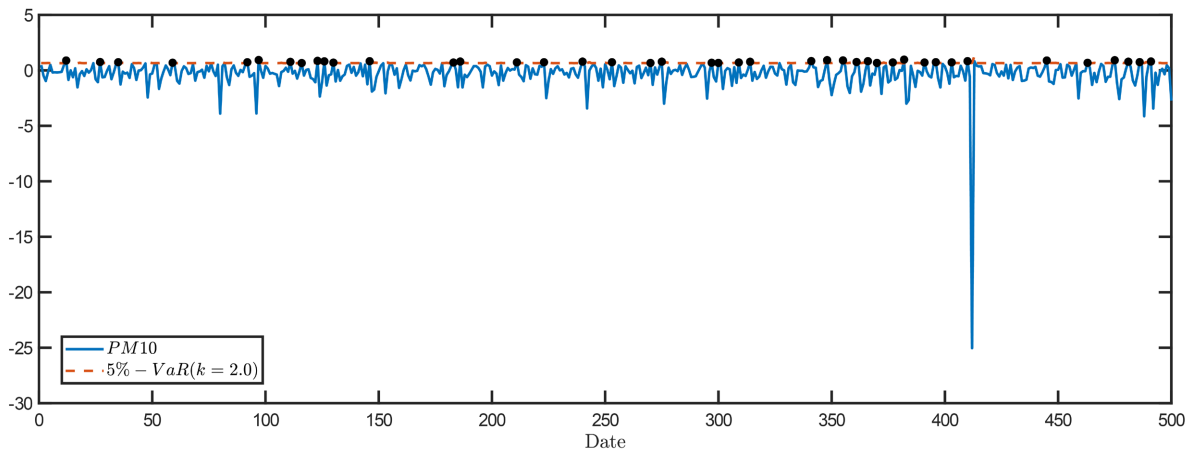
Statistical magnitude	PM10	Statistical magnitude	PM10
Mean	0.21	Max	0.97
SD	0.99	Range	26.01
Median	0.07	Skew	11.69
Min	25.04	Kurtosis	255.54



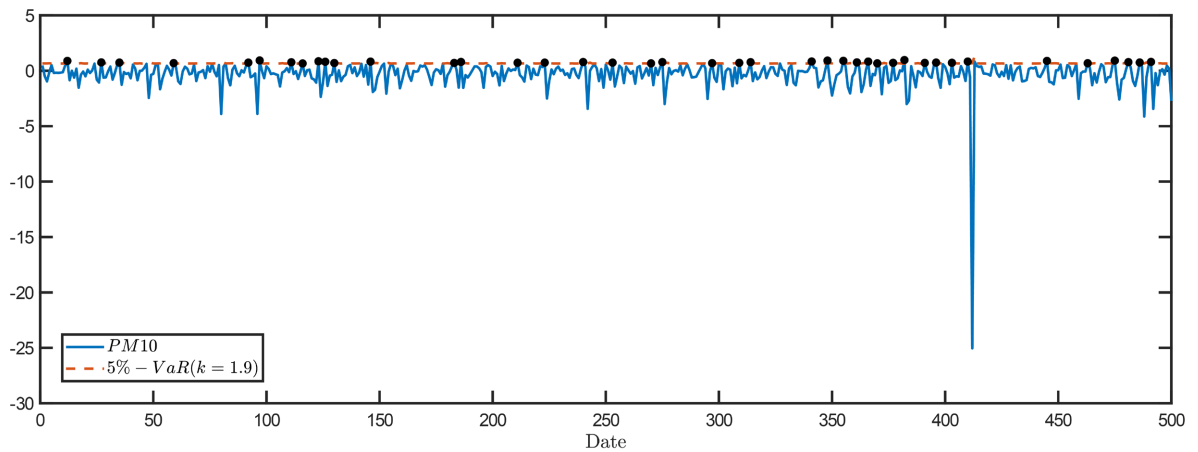
**Figure 1.** PM10 daily growth rate of Wanshou West Palace, Beijing.



**Figure 2.** 5%-VaR prediction of PM10 based on quantile regression ( $\hat{\sigma} = 0.084, RE = 0.68$ ).



**Figure 3.** 5%-VaR prediction of PM10 based on expectile regression ( $\hat{\sigma} = 0.084, RE = 0.68$ ).



**Figure 4.** 5%-VaR prediction of PM10 based on the kth power expectile regression ( $\hat{\sigma} = 0.082, RE = 0.64$ ).

asymptotic variance, and excellent sensitivity to tail events. Therefore, this method provides a good idea for air pollution control.

Since the PM10 monitoring concentration data is affected by meteorological

conditions and human factors, the data is incomplete and highly volatile. It is suggested to use the hourly growth rate of air pollutant concentration as sample data to reduce the influence of meteorological conditions and human factors on the data.

### Acknowledgements

We appreciate for Editors' and Reviewers' warm work earnestly. Their comments and suggestions are very meaningful to our research work.

### Funding

This work is partly supported by Graduate Textbook Construction Project of Sichuan University of Science and Engineering (Grant No. KA202011).

### Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

### References

- [1] Shahraiyni, H.T. and Sodoudi, S. (2016) Statistical Modeling Approaches for PM10 Prediction in Urban Areas: A Review of 21st Century Studies. *Atmosphere*, **7**, 15. <https://doi.org/10.3390/atmos7020015>
- [2] Masseran, N. (2017) Modeling Fluctuation of PM10 Data with Existence of Volatility Effect. *Environmental Engineering Science*, **34**, 816-827. <https://doi.org/10.1089/ees.2016.0448>
- [3] Wang, P., Zhang, H., Qin, Z. and Zhang, G. (2017) A Novel Hybrid-Garch Model Based on ARIMA and SVM for PM2.5 Concentrations Forecasting. *Atmospheric Pollution Research*, **8**, 850-860. <https://doi.org/10.1016/j.apr.2017.01.003>
- [4] Veleva, E. and Zeleva, I. (2018) GARCH Models for Particulate Matter PM10 Air Pollutant in the City of Ruse, Bulgaria. *AIP Conference Proceedings*, **2025**, 040016 <https://doi.org/10.1063/1.5064900>
- [5] Suleiman, A., Tight, M.R. and Quinn, A.D. (2019) Applying Machine Learning Methods in Managing Urban Contaminations of Traffic-Related Particulate Matter (PM10 and PM2.5). *Atmospheric Pollution Research*, **10**, 134-144. <https://doi.org/10.1016/j.apr.2018.07.001>
- [6] Karimian, H., Li, Q., Wu, C., Qi, Y., Mo, Y., Chen, G., Zhang, X. and Sachdeva, S. (2019) Evaluation of Different Machine Learning Approaches to Forecasting PM2.5 Mass Concentrations. *Aerosol and Air Quality Research*, **19**, 1400-1410.
- [7] Czernecki, B., Marosz, M. and Jędruszkiewicz, J. (2021) Assessment of Machine Learning Algorithms in Short-Term Forecasting of PM10 and PM2.5 Concentrations in Selected Polish Agglomerations. *Aerosol and Air Quality Research*, **21**, 200586-200586.
- [8] Cai, J., Ge, Y., Li, H., Yang, C., Liu, C., Meng, X., Wang, W., Niu, C., Kan, L., Shikowski, T., Yan, B., Chillrud, S.N., Kan, H. and Jin, L. (2020) Application of Land Use Regression to Assess Exposure and Identify Potential Sources in PM2.5, BC, NO<sub>2</sub> Concentrations. *Atmospheric Environment*, **223**, 117267. <https://doi.org/10.1016/j.atmosenv.2020.117267>



- 
- [9] Yang, W., Tang, G., Hao, Y. and Wang, J. (2021) A Novel Framework for Forecasting, Evaluation and Early-Warning for the Influence of PM10 on Public Health. *Atmosphere*, **12**, 1020. <https://doi.org/10.3390/atmos12081020>
- [10] Manganeli, S. and Engle, R.F. (2004) A Comparison of Value-at-Risk Models in Finance. In: Szegő, G. (Ed.), *Risk Measures for the 21st Century*, Wiley, Chichester, UK, 123-144.
- [11] Jiang, Y.Y., Lin, F.M. and Zhou, Y. (2021) The kth Power Expectile Regression. *Annals of the Institute of Statistical Mathematics*, **73**, 83-113. <https://doi.org/10.1007/s10463-019-00738-y>
- [12] Sanchis-Marco, L., Montero, J.-M. and Fernandez-Aviles, G. (2022) An Extended CAViaR Model for Early-Warning of Exceedances of the Air Pollution Standards. The Case of PM10 in the City of Madrid. *Atmospheric Pollution Research*, **13**, 101355. <https://doi.org/10.1016/j.apr.2022.101355>
- [13] Taylor, J.W. (2008) Estimation Value at Risk and Expected Short-Fall Using Expectiles. *Journal of Financial Econometrics*, **6**, 231-252. <https://doi.org/10.1093/jjfinec/nbn001>
- [14] Kuan, C., Yeh, J. and Hsu, Y. (2009) Assessing Value at Risk with CARE, the Conditional Autoregressive Expectile Models. *Journal of Econometrics*, **150**, 261-270. <https://doi.org/10.1016/j.jeconom.2008.12.002>
- [15] Efron, B. (1991) Regression Percentiles Using Asymmetric Squared Error Loss. *Statistica Sinica*, **1**, 93-125.