Advanced Econometrics - Final project

Kamil Matuszelański, Rafał Rysiejko

20/04/2020

Table of Contents

[Abstract](#_Toc42115611)

[Introduction](#_Toc42115612)

[Data](#_Toc42115613)

[Statistical analysis](#_Toc42115614)

[Methods (ARIMA)](#_Toc42115615)

[ARDL Model Part](#_Toc42115616)

[Methodology](#_Toc42115617)

[Analysis](#_Toc42115618)

[Results - ARIMA](#_Toc42115619)

[Conclusions and possible extensions](#_Toc42115620)

[References](#_Toc42115621)

## Abstract

Excess air pollution is one of the environmental issues gaining biggest media and public awareness coverage. In this study, we have tried to assess various factors influencing day-to-day changes in particulate matter measurements in Beijing, China over the course of 4 years. Specifically, we have assessed importance of current weather features, like temperature and pressure. We have also tested the hypothesis about seasonality of pm2.5 measurements. Weather factors were shown to be highly significant in current levels of air pollution. We have also shown that the series exhibits seasonal behaviour, with higher levels of pollution in colder seasons, autumn and winter.

## Introduction

Air pollution is a severe environmental problem that is attracting increasing attention worldwide (Kurt, Oktay, 2010). It is considered to be the major cause of premature death and disease, and the International Agency for Research on Cancer has classified it as carcinogenic (IARC, 2013; Pope, Dockery, 2006). Particulate Matter (PM) is known to have huge impacts on human health, is related to respiratory problems, bronchitis, reduced lung functionality (WHO, 2013; Martuzevicius et al., 2004). Modeling and interpreting air pollution data are recognized as key steps to foster the reduction of the harmful effects of air pollution (Kumar, Peuch, Crawford, Brasseur, 2018). Several studies have been paying attention to the characterization of air pollution time series (C. Belis, et. al, 2013; S. Yatkin, A. Bayram, 2007; R. Salcedo, et. al, 1999). Studies performed in China suggest a strong contribution of meteorological conditions to the development of air pollutants (Liang, Zou, Guo, et. al, 2015). In this study the time series on Beijing PM2.5 air pollution and meteorological records from 2010–2015 are analyzed. Focus of the paper is on the characterization of the considered time series using ARIMA and ARDL class models to provide insights concerning the phenomena and pollution patterns and its relationship with meteorologic factors. The insights obtained from our study might be useful to support the development of more efficient policies aiming at the reduction of emission of pollutant in air. The remainder of the paper is organized as follows. Section 2 explores the air pollution dataset investigated in this paper as well as its explanatory data analysis. Section 3 presents and discusses the results obtained from two undertaken modeling approaches: ARIMA and ARDL. Finally, section 4 summarizes the conclusions and discusses potential future works.

## Data

Dataset used in this study was obtained from UCI Machine Learning Repository[[1]](#footnote-1). The observation period spanned from January 1st, 2010 to December 31st, 2014, and the measurements were taken at the US Embassy in Beijing. In total, there are 43 824 hourly measurements. The dataset also contains weather key weather parameters occurring at that time. Before further analysis, we have decided to aggregate pm2.5 measurements to daily values by taking a simple mean of the values throughout a particular day. This reduced the size of the dataset to 1 826 observations. There were two reasons for doing so. In intra-daily data, a complex seasonal pattern usually occurs, as the seasonalities arise in daily, weekly and annual manner. This poses difficulty to correct model estimation as an ARIMA model, and its seasonal extensions are designed for a smaller and simpler seasonal pattern. According to Hyndman (2018), the only sensible model to use for such data is dynamic harmonic regression. Another reason for daily aggregation was that the quality of pm2.5 measurements was an unknown factor to us. In the dataset, there were numerous observations with an extraordinarily high value of pollution presented. However, there is no way to assess whether these were present due to actual value or faulty measurement. By aggregation, the values were smoothed, and the influence of the outliers mitigated. The dataset contained missing values, but only in the target variable. In total, there were 2067 such observations (before daily aggregation), which accounted for 5% of observation count. We have filled these using forward fill, which means that missing observation was imputed using last non-missing observations.

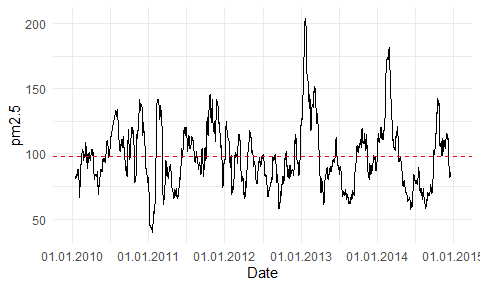
Table 1: Variable description

|  |  |
| --- | --- |
| Variable | Description |
| year | Year of data recording |
| month | Month of data recording |
| day | Day of data recording |
| hours | Hour of data recording |
| pm2.5 | PM2.5 concentration (ug/m^3) |
| DEWP | Dew Point |
| TEMP | Temperature |
| PRES | Pressure |
| cbwd | Combined wind direction |
| lws | Cumulated wind speed (m/s) |
| ls | Cumulated hours of snow |
| lr | Cumulated hours of rain |

Table .. shows basic summary statistics of the features present in the model.

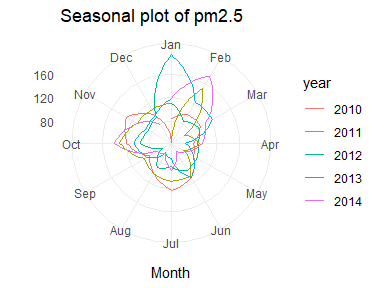
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| variable | min | Q25 | mean | median | Q75 | max |
| pm2.5 | 2.96 | 41.58 | 97.78 | 78.46 | 130.92 | 541.04 |
| DEWP | -33.33 | -10.08 | 1.83 | 2.04 | 15.08 | 26.21 |
| TEMP | -14.46 | 1.54 | 12.46 | 13.92 | 23.17 | 32.88 |
| PRES | 994.04 | 1007.92 | 1016.45 | 1016.21 | 1024.54 | 1043.46 |
| Iws | 1.41 | 5.90 | 23.89 | 10.95 | 22.23 | 463.19 |
| Is | 0.00 | 0.00 | 0.05 | 0.00 | 0.00 | 14.17 |
| Ir | 0.00 | 0.00 | 0.20 | 0.00 | 0.00 | 17.58 |

On the plot …, pm2.5 measurements are presented. For this particular visualization, we have applied smoothing using moving average of order 30. Because of the high day-to-day volatility, a standard daily plot would be not very clear. As can be seen, biggest spikes in pollution measurements are present at the beginning of years 2013 and 2014. For other periods, the measurement oscillates around mean (97.8, shown as red line).



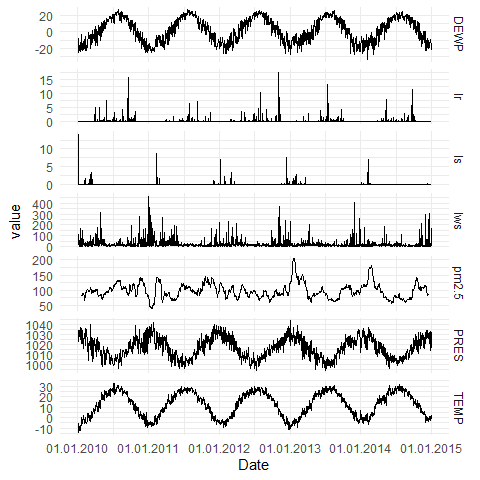
Plot of pm2.5 with MA(30) smoothing

The annual differences of the distribution of the pm2.5 pollution can be also seen on the seasonal polar plot.



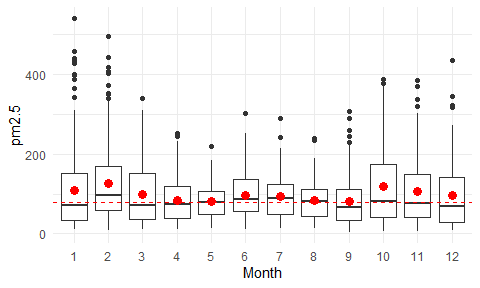
seasonal plot pm2.5 variable

On the plot …, all variables from the dataset are presented one after another. Important observation is that temperature, pressure and dew point are probably highly cointegrated.



Time series plots of all variables in the dataset

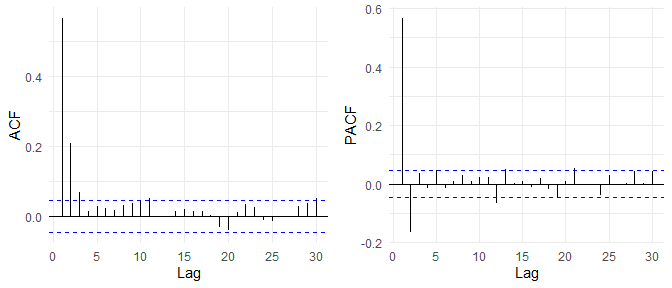
The box plot … shows concentration of pm2.5 particles grouped by month. Red dots indicate the mean in particular month. In all months values of mean are larger than the median, indicating that the distributions are right-skewed. The month with highest both mean and median values is February, followed by June and July. While looking at the right tail of the distribution, winter months (from October until March) account for more extreme measurements. This is an indicator that the series may exhibit seasonal behaviour.



Boxplot of pm2.5 variable by month

### Statistical analysis

ACF and PACF measures are presented on plot … . ACF becomes insignificant at lag 4, with small crossings of insignificance border at lag 10 and 11. PACF vanishes at lag 2, but with significant lags 13, 14 and 21. Both of the functions exhibit vanishing of exponential type. This does not fit to any of the theoretical patterns helpful for determining correct order of AR and MA parts.



ACF and PACF of pm2.5 variable

We have assessed stationarity of pm2.5 concentration variable using Augmented Dickey-Fuller test. Visual assessment of the variable indicates that the series does not look like random walk with drift or trend. Thus, correct type of the test is “none”. Breusch-Godfrey test helped to assess correct number of needed augmentations. We have decided to check for auto-correlation up to 5 lags. Results of the test are presented in the table …. . p-values from BG test indicate that the correct number of augmentations is 8, as this is a first number of augmentations for which all p-values are higher than 0.05. In this version, p-value of Augmented Dickey Fuller test is lower than 0.01 (R function was unable to determine the exact number). There is a reason to reject null hypothesis of unit root present in the test set. That means that the dependent variable is stationary.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| augmentations | p\_adf | p\_bg.p.value.1 | p\_bg.p.value.2 | p\_bg.p.value.3 | p\_bg.p.value.4 | p\_bg.p.value.5 |
| 0 | <0.01 | 0.0095 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 1 | <0.01 | 0.0115 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 2 | <0.01 | 0.0065 | 0.0001 | 0.0000 | 0.0000 | 0.0000 |
| 3 | <0.01 | 0.0320 | 0.0006 | 0.0000 | 0.0000 | 0.0000 |
| 4 | <0.01 | 0.1287 | 0.0328 | 0.0011 | 0.0000 | 0.0000 |
| 5 | <0.01 | 0.2514 | 0.0810 | 0.0117 | 0.0003 | 0.0000 |
| 6 | <0.01 | 0.2777 | 0.2035 | 0.0539 | 0.0076 | 0.0004 |
| 7 | <0.01 | 0.3720 | 0.2818 | 0.1676 | 0.0898 | 0.0202 |
| 8 | <0.01 | 0.4594 | 0.3845 | 0.3497 | 0.2504 | 0.0971 |

## Methods (ARIMA)

In our study we have compared 2 approaches to time series modeling, that is autoregressive moving average (AR(I)MA) and Autoregressive Distributed Lags.

At first, we have assessed the need of making the dependent variable stationary.

As for ARIMA modeling, we have applied general-to-specific procedure concerning appropriate lags selection. The most important criterion was if obtained residuals present autocorrelation patterns. This was assessed using both Ljung-Box test and visual inspection of (Partial) Autocorrelation Function. In Ljung-Box test, 10 lags were used.

Second concern was significance of all variables included in the model. This was assessed using z test for coefficients. Lastly, if two models were proper by previous criteria, better model was defined using goodness-of-fit measures, specifically Akaike Information Criterion.

One of the hypotheses we wanted to check was if air pollution in Beijing exhibits seasonal behaviour. We have assessed this by adding additional regressors to obtained ARIMA model. Specifically, we have included 11 dummy variables, each of them representing one month, with January being base level. On these variables we have applied step wise feature elimination. Starting from full model, we have removed dummies one by one, starting from the one with highest p-value.

Apart from dummy regressors for months, we have also tested quarterly seasonality. This time, 3 dummies were included in the model.

An alternative way to account for seasonal effect in ARIMA model is seasonal ARIMA model. In this extension, the model consists of 2 components. One is the same as in standard model, that is a set of parameters for AR and MA parts. Second part also contains the AR and MA parts, however these are estimated on variable lagged by the period of seasonality. There were three reasons for choosing external regressors approach instead of SARIMA. First, using seasonal dummies are possible to interpret, in the same way as in ARDL approach. Second, in seasonal ARIMA one has to decide about exact order of seasonality. We have expected the dependent variable to exhibit yearly seasonality. In the easiest case, lag of order 365 would be apparent (??). However, as shown later, lag of order 365 was not significant from visual inspection of ACF and PACF plots. Adding seasonal variables is more “forgiving”, as information passed (??) to the model is much more general and less error prone.

Third reason for not using SARIMA model is that for estimation of period *t* lag of period *t-365* would be used. This means that the whole first year of the measurements present in the dataset would be excluded from model predictions. Using seasonal dummies allows for estimation from day 1 (without accounting for lags needed in estimation of standard AR part).

This facilitates testing of our main hypothesis, that is assessing seasonality of the air pollution.

## ARDL Model Part

### Methodology

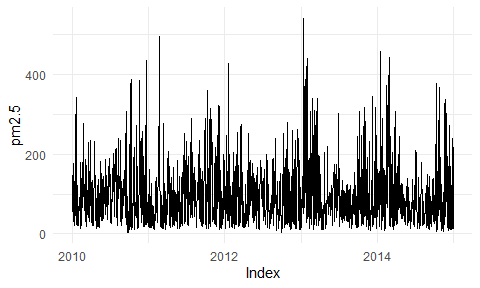
An autoregressive distributed lag (ARDL) model is an ordinary least square (OLS) based model which is applicable for both non-stationary time series as well as for times series with mixed order of integration. This model takes sufficient numbers of lags to capture the data generating process in a general-to-specific modeling framework (Paseran,Hasem,Youngcheol). ARDL model addresses the issue of collinearity by introducing the lag of dependent variable into the model with other independent variables and their lags.Equation of ARDL(m,n) can be described as follows:

Where *m* and *n* are the number of years for lag is the disturbance term , are coefficients for short run and are coefficients for long run relationship.

Econometric modelling using ARDL class of models needs to meet a set of assumptions. Absence of auto correlation is the very first requirement of ARDL. The model requires that the error terms should have no autocorrelation with each other. The variance-covariance matrix should also display spherical, meaning there is no heteroscedasticity in the data. As mentioned above, analyzed data should have stationarity either on *I(0)* or *I(1)* level or on both. Last but not least residuals should follow normal distribution.

## Analysis

As the most of analyzed variables are of the meteorological origin, it is worth to perform explanatory analysis of our variables of interests to better understand kinds of processes that generate them. As the variable *ls* (Cumulated hours of snow) and *lr* (cumulated hours of rain) are basically the same process, authors decided to merge them for the purposes of ARDL modelling. Newly created variable *Isr* is a sum of both rain and snow hours.



Annual dstribution of Isr

Following, the authors analyzed other meteorological factors, which later will be used in the modelling. Using the Augmented Dickey-Fuller test, stationarity of variables *TEMP*, *DEWP*,*PRES*,*Iws*,*Isr* was assessed. Breush-Godfrey test were deployed to determine the correct number of required augmentations. The number of augmentations was selected for a given number for which all p-values are higher than 5% in the BG-test. As a next step the result for ADF-test for this particular number of augmentations was assessed. The results are presented in table X.

With a significance with a *p-value* lower than assumed 5% threshold, we can reject the null hypothesis about variable [list] being integrated of order 1 or more. We can assume that alternate hypothesis, saying that variables are integrated of order 0. **!Check this !**

As a first step authors constructed a simple model with only variable of interest *pm2.5* and its lags. The number of lag selection was based on the goodness-of-fit measure AIC. Obtained values are presented in table X.

As we can observe the lowest value of AIC is obtained for 3 lags. Summary of model with 3 lags of a dependent variable *pm2.5* is presented in table X

[Variable interpretation]

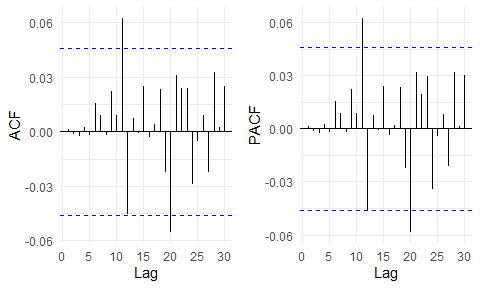
Out of variables included in the model, first and second lags are statistically significant at an assumed 5% threshold. Starting with the short term multiplier (STM), whenever the first lag of *pm2.5* increases by one unit, the amount of *pm2.5* increases by XXX. For the second order lag, whenever it increases by one unit,the amount of *pm2.5* increases by XXX. As for the Long Term Multiplier (LTM), whenever an unitary shock happens to the lagged *pm2.5*, then the overall impact on the *pm2.5* would be equal to XXX.

#### Model diagnostics

## Results - ARIMA

#### Standard

We have selected ARIMA(3, 0, 3) (or ARMA(3, 3)) as the starting point. We have inferred about the number of initially included lags from ACF and PACF plot *p1* of the series. The model includes constant, as mean of the air pollution is non-zero, and the series is already stationary without the need for differentiating. On the plot *p2*, ACF and PACF of the residuals are shown. Lags 11 and 20 are slightly crossing the significance threshold, using both functions.



ACF and PACF of ARIMA(3, 0, 3) residuals

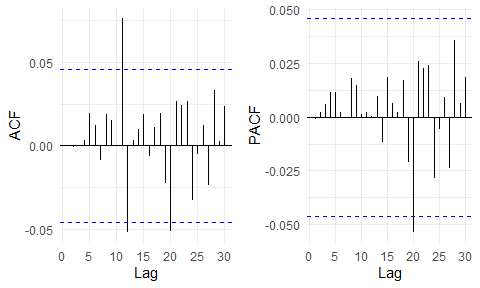
In table *t1*, basic summary statistics of the tested ARIMA models are shown. We have also included results of Ljung-Box test for autocorrelation of residuals. For model ARIMA(3,0,3), with p-value = 0.6343, one can conclude that number of lags included is correct.

Basing on the results of first model, we have tried fitting the same model, but including lags 11 and 20, one at a time. Results are presented in table, *t1*, in rows 2 and 3, respectively. In both models, there is autocorrelation of residuals present. This means that initial model cannot be improved by adding significant lags.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| order | AIC | BIC | RMSE | Ljung.Box.statistic | Ljung.Box.p.value |
| ARIMA(3, 0, 3) | 20274.88 | 20318.95 | 62.25928 | 1.711928 | 0.6342851 |
| … with added lag 11 | 20265.22 | 20320.31 | 62.02496 | 14.308861 | 0.0025135 |
| … with added lag 20 | 20277.18 | 20332.28 | 62.22899 | 51.928341 | 0.0000000 |
| ARIMA(0,0,3) | 20273.10 | 20300.64 | 62.33161 | 2.218946 | 0.8985012 |
| … with added lag 11 | 20264.36 | 20297.42 | 62.14741 | 9.710101 | 0.0211982 |

We have also tried another way of selecting initial model than visual inspection of autocorrelation plots. R package *forecast* provides function for automatic order selection of ARIMA model based on predefined criteria. We have run it using AIC criterium. Obtained model is ARIMA(0, 0, 3) (or MA(3)). As previously, summary statistics of the model are shown in table *t1*. From assessment of Ljung-Box p-value (0.8985), one can conclude about lack of autocorrelation of residuals.

ACF and PACF functions for this model are shown on plot *p3* . This time, only lag standing out as potentially significant was lag 11. We have also tried to fit the model with this one included. Results, as previously, are presented in table *t1*, in the last row. P-value for autocorrelation test was 0.0212, which means that autocorrelation of residuals is present. Thus, we have not analyzed this model further.



ACF and PACF of ARIMA(0, 0, 3) residuals

From the table *t1*, one can also compare models ARMA(3,3) and MA(3) in terms of goodness of fit. Model ARMA(0,0,3) is better according to both AIC (20273.10 vs. 20274.88) and BIC (20300.64 vs. 20318.95) criteria. Residual mean squared error is very similar, slightly in favour of less parsimonious model. In table *t2*, we have included parameter estimates of both models. As can be seen, for every parameter p-value is even less than 0.005, meaning that all parameters in both models are significant.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| term | estimate\_303 | p.value\_303 | estimate\_003 | p.value\_003 |
| ar1 | -0.6131 | 0.0036 |  |  |
| ar2 | -0.4391 | 0.0001 |  |  |
| ar3 | 0.2589 | 0.0016 |  |  |
| ma1 | 1.2841 | 0 | 0.6719 | 0 |
| ma2 | 1.107 | 0 | 0.2557 | 0 |
| ma3 | 0.2887 | 0 | 0.0949 | 0 |
| intercept | 97.7125 | 0 | 97.7888 | 0 |

Summarizing, the best ARIMA model without additional dummy variables is ARIMA(0,0,3) (??). This was indicated by AIC and BIC scores. In the model all variables are significant, and autocorrelation of residuals is nonexistent. For further analysis, we have decided to build upon this model also because of the reason that it contains less parameters than ARMA(3,3), giving even better fit.

#### Seasonal

Next step in testing ARIMA approach was adding simple external regressors. We have started with 11 dummies indicating consecutive months, with January as base level.

In this model, most of the variables were insignificant. Autocorrelation of residuals based on Ljung-Box test was also present (p-value 0.0004). To improve the measures, we started removing dummies, one by one, starting from the one with the highest p-value. Finally, dummies indicating months February and November both were significant. Autocorrelation was non-existent (p-value 0.9549). Results for model summary are presented in table *t3*. In terms of goodness-of-fit measures, model with monthly dummies was better than standard MA(3) by AIC, and worse by BIC criteria. RMSE for extended model improved by small amount.

Apart from including external monthly dummies, we have also tried incorporating even more general seasonality indicator. We have included one dummy, to indicate if the observation was taken from April to September. Summary statistics are also presented in table *t3*. Autocorrelation of residuals was not present (p-value 0.6788). All 3 goodness-of-fit measures improved from both the MA(3) model, and model with monthly dummies.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| order | AIC | BIC | RMSE | Ljung.Box.statistic | Ljung.Box.p.value |
| ARIMA(0,0,3) | 20273.10 | 20300.64 | 62.3316 | 2.2189 | 0.8985 |
| … with dummies for 2, 11 months | 20262.66 | 20301.23 | 62.0858 | 1.8595 | 0.7616 |
| … with dummy for summer season | 20257.78 | 20290.84 | 62.0367 | 1.5290 | 0.9097 |

In table *t4*, we have included parameter estimates of standard MA model, the one with monthly dummies and the one with season dummy. In all 3 models, all parameters are significant (p-value<0.01). Estimated parameters in monthly model were 30.2 and 25.9 for months February and November, respectively. These parameters can be interpreted, in a way that the value of pm2.5 concentration in air is expected to be higher than in other months of the year by 30.2, and 25.9, respectively. While in seasonal model, one can conclude that on average in warmer half of the year the level of air pollution is lower by 23.95 /m^2.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| term | estimate | p.value | estimate\_months | p.value\_months | estimate\_seasons | p.value\_seasons |
| ma1 | 0.6719 | 0 | 0.6659 | 0 | 0.6662 | 0 |
| ma2 | 0.2557 | 0 | 0.2461 | 0 | 0.2466 | 0 |
| ma3 | 0.0949 | 0 | 0.0883 | 0.0001 | 0.0885 | 0.0001 |
| intercept | 97.7888 | 0 | 93.2806 | 0 | 109.8073 | 0 |
| mon2 |  |  | 30.2691 | 0.0031 |  |  |
| mon10 |  |  | 25.9038 | 0.0085 |  |  |
| if\_summer |  |  |  |  | -23.9477 | 0 |

## Conclusions and possible extensions

In our work, we have not included forecasts for out-of-sample data. In such task, problem with obtaining independent variables arises. In ARDL model we have used weather data without lags. In forecasting, such data would not be available. One way would be to include forecasting of these variables with ARIMA model. This approach would perform poorly, as it does not take into account underlying complex and dynamic behaviours of the atmosphere. Current state-of-the art weather forecasting models will be correct in 90% of the cases, for 5-day forecast. One of the possible extensions of our work would be to include such forecasts from external service, to improve performance of prediction for longer periods of time.

Other extension of our current model, specifically addressing the assessment of hypothesis of seasonality, would be including some type of series decomposition methods, such as Seasonal and Trend decomposition using Loess (STL). This method enables to decompose time series into 3 components - long-term trend, seasonal part and remainder. In this way, the remainder part would be “seasonally adjusted”. In such case assessment of the short-term weather fluctuations versus seasonal, general trends influence on level of air pollution would be possible.

As we stated before, cointegration of weather variables is apparent from visual inspection of the plots. One of the ways to simplify the model and still include the information from all variables would be to use some dimension reduction technique, e.g. Principal Component Analysis on the weather variables.

## References

Hyndman, R.J., & Athanasopoulos, G. (2018) Forecasting: principles and practice, 2nd edition, OTexts: Melbourne, Australia. OTexts.com/fpp2. Accessed on 23.05.2020.

Kurt, A.; Oktay, A.B. Forecasting air pollutant indicator levels with geographic models 3 days in advance using neural networks. Expert Syst. Appl. 2010, 37, 7986–7992.

R. Peled, Air pollution exposure: Who is at high risk?, Atmospheric Environment 45 (10) (2011) 1781–1785. <doi:10.1016/J.ATMOSENV.2011>.

IARC, Outdoor air pollution a leading environmental cause of cancer deaths, Tech. rep., International Agency for Research on Cancer, WHO (2013).

C. A. Pope III, D. W. Dockery, Health effects of fine particulate air pollu- tion: lines that connect, Journal of the air & waste management association 56 (6) (2006) 709–742.

W. WHO, Health aspects of air pollution with particulate matter, ozone and nitrogen dioxide, World Health Organization Working Group Bonn, Germany, 13-15.

D. Martuzevicius, S. A. Grinshpun, T. Reponen, R. L. G´orny, R. Shukla, J. Lockey, S. Hu, R. McDonald, P. Biswas, L. Kliucininkas, et al., Spatial and temporal variations of pm2. 5 concentration and composition through- out an urban area with high freeway densitythe greater cincinnati study, Atmospheric Environment 38 (8) (2004) 1091–1105.

R. Kumar, V.-H. Peuch, J. H. Crawford, G. Brasseur, Five steps to improve air-quality forecasts (2018).

C. Belis, F. Karagulian, B. R. Larsen, P. Hopke, Critical review and meta- analysis of ambient particulate matter source apportionment using receptor models in europe, Atmospheric Environment 69 (2013) 94–108.

S. Yatkin, A. Bayram, Elemental composition and sources of particulate matter in the ambient air of a metropolitan city, Atmospheric Research 85 (1) (2007) 126 – 139. <doi:https://doi.org/10.1016/j.atmosres.2006.12.002>. URL <http://www.sciencedirect.com/science/article/pii/> S0169809506002882

R. Salcedo, M. A. Ferraz, C. Alves, F. Martins, Time-series analysis of air pollution data, Atmospheric Environment 33 (15) (1999) 2361–2372.

Liang, X., Zou, T., Guo, B., Li, S., Zhang, H., Zhang, S., Huang, H., & Chen, S. X. (2015). Assessing Beijing’s PM2.5 pollution: Severity, weather impact, APEC and winter heating. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 471(2182). <https://doi.org/10.1098/rspa.2015.0257>

| term | estimate | std.error | statistic | p.value |
| --- | --- | --- | --- | --- |
| ar1 | -0.90377522 | NaN | NaN | NaN |
| ar2 | -0.92004046 | NaN | NaN | NaN |
| ma1 | 0.79515565 | NaN | NaN | NaN |
| ma2 | 0.86966874 | NaN | NaN | NaN |
| ma3 | -0.07080841 | 0.06580462 | -1.0760402 | 0.28190927 |
| ma4 | 0.13114233 | 0.06775885 | 1.9354273 | 0.05293788 |
| ma5 | 0.08832036 | 0.07089832 | 1.2457327 | 0.21286253 |
| ma6 | 0.03819121 | 0.06574193 | 0.5809263 | 0.56129015 |
| ma7 | -0.02493630 | 0.06498064 | -0.3837497 | 0.70116397 |
| ma8 | -0.06049744 | 0.05784811 | -1.0457981 | 0.29565426 |
| ma9 | -0.04315530 | 0.02934785 | -1.4704759 | 0.14143291 |

1. <https://archive.ics.uci.edu/ml/datasets/Beijing+PM2.5+Data> [↑](#footnote-ref-1)