Vysoká škola ekonomická v Praze

Fakulta informatiky a statistiky

# Time series analysis of selected pollutants emission in the United Kingdom in years 1990-2016

BACHELOR THESIS

Study programme: Quantitative Methods in Economics

Field of study: Statistics and Econometrics


Author: Egor Shinkariuk

Thesis supervisor: Ing. Karel Helman, Ph.D.


Prague, 05/2019

## Declaration

I hereby declare that I am the sole author of the thesis entitled "Time series analysis of the emissions of the selected pollutants in the United Kingdom in the years 1990-2016”. I duly marked out all quotations. The used literature and sources are stated in the attached list of references.

In Prague on 6 May 2019 ..............................................................

Egor Shinkariuk

## Abstract

This bachelor thesis focuses on the analysis of four pollutants emissions in the UK in different industrial sectors. Data for this thesis were obtained from the British Office for National Statistics.

The objective of the thesis is to describe the development of the harmful pollutants emissions time series in the UK in order to evaluate regulatory controls applied in this country. There are two main regulatory frameworks: the EU National Emission Ceilings Directive (NECD) (2001/81/EC) and the Gothenburg Protocol to the UNECE Convention on Long Range Transboundary Air Pollution (CLRTAP). For the mentioned evaluation, we got data for the period from 1990 to 2016.

The thesis is divided into two parts. The first part is dedicated to the theoretical description of the selected statistical methods and the second part is devoted to an application of those methods on the selected time series.

Tato bakalářská práce je zaměřena na analýzu časových řad emisi čtyř typů znečisťujících látek ve Velké Británii v různých sektorech průmyslu. Data pro tuto práci byla získána na webových stránkách „Office for National Statistics“.

Cílem práce je popsat vývoj časových řad škodlivých látek za účelem hodnocení následujících právních nástrojů ochrany ovzduší ve Velké Británii: „Směrnice o národních emisních stropech“ a „Úmluva o dálkovém znečišťování ovzduší přesahujícím hranice států“. Pro toto zhodnocení máme k dispozici data od roku 1990 do roku 2016.

Práce je rozdělena do dvou částí. První část je věnována teoretickému popisu vybraných statistických metod, druhá část pak jejich aplikaci na vybrané časové řady.



## Keywords

pollutants, emissions, atmosphere, time series.

## Table of Contents

# Table of Figures

# Introduction

The purpose of this bachelor thesis is to describe the development of the emissions time series from harmful pollutants into the atmosphere in the UK by the industry sectors to evaluate regulatory controls applied in this country. From the Office for National Statistics we have the long-term emissions time series which covers the period from 1990 to 2016. This time series has an annual periodicity. The pollutants covered by the thesis are the following:

- Particulate Matter ($PM_{10}$ and $PM_{2.5}$)
- Carbon monoxide (CO)
- Non-methane volatile organic compound emissions (NMVOC)

The reason why those four pollutants were selected is because they are associated with the numerous negative health and environmental effects.  Each of those matters will be defined in the following paragraphs as well as will examples of the negative inputs will be provided.

**Particulate Matter** are small airborne particles which can contain different materials such as dust, soot or components formed within the atmosphere following a chemical reaction. There are natural and anthropogenic sources of the particulate matter pollution.[1] The size fraction is important in the context of health so two types of PMs are defined: PM10 and PM2.5. PM10 are particles which pass through a size-selective inlet with a 50 % efficiency cut-off at 10 μm aerodynamic diameter, as defined in ISO 7708:1995, Clause 6. These particles are small enough to be inhaled into the airways of the lung – described as the 'thoracic convention' in the above ISO standard. PM2.5 are particles which pass through a size-selective inlet with a 50 % efficiency cut-off at 2.5 μm aerodynamic diameter. These particles are small enough to be inhaled very deep into the lung – described as the 'high-risk respirable convention' in the above ISO standard. [2]

Those pollutants are the most harmful to the human health; due to the small size of the particles they can easily penetrate deep into blood streams and lungs unfiltered. According to the research conducted in nine European countries there is no safe level of particulate matter pollutants and "Particulate matter air pollution contributes to lung cancer incidence in Europe".[3] Those findings corresponds with the International Agency for Research on Cancer classifications which designates particular matters as a Group 1 carcinogens. [4]

**Carbone monoxide (CO)** is a colourless and extremely toxic gas without smell or taste, it is slightly lighter than air in normal conditions and it is flammable. This gas released in road vehicle exhausts and emitted by the metallurgical or energy industries.[2]

From perspective of the human health, CO poisoning "may be the cause of more than 50% of fatal poisonings in many industrial countries".[6] CO affects the blood's ability to carry oxygen when inhaled.

**Non-methane volatile organic compound emissions (NMVOC)** are the collection of organic components which are emitted from different sources and display similar behaviour in the atmosphere.[1] Justifications of the monitoring NMVOCs emissions is the contribution to the formation of tropospheric ozone which has a negative impact on human health.

In all cases the pollutants are observed in thousand tonnes per year.

At our disposal there are time series for 21 industries which are the sources of pollution:

1. Agriculture, forestry and fishing
2. Mining and quarrying
3. Manufacturing
4. Electricity, gas, steam and air conditioning supply
5. Water supply; sewerage, waste management and remediation activities
6. Construction
7. Wholesale and retail trade; repair of motor vehicles and motorcycles
8. Transport and storage
9. Accommodation and food services
10. Information and communication
11. Financial and insurance activities
12. Real estate activities
13. Professional, scientific and technical activities
14. Administrative and support service activities
15. Public administration and defence; compulsory social security
16. Education
17. Human health and social work activities
18. Arts, entertainment and recreation
19. Other service activities
20. Activities of households as employers; undifferentiated goods and services-producing activities of households for own use
21. Consumer expenditure

The goal of this thesis is to get an overview of the development of the emission time series of the pollutants mentioned before using selected statistical methods and compare results between the industry sectors. For the purpose of this thesis all 21 industries will be analysed only on the subject of the Non-methane volatile organic compound pollutant, the other two pollutants will be analysed across the 8 most pollutant industries only. The reason is that more than 92% of the emission is generated by those 8 most pollutant industries. For example, in the year 2016, 98.02% of PM10 pollutant were emitted by the 8 most pollutant industries.

Another goal is to assess several hypotheses:

1. In the last two decades, ecology standards in the transportation area were drastically improved, the number of vehicles with eco-friendly engines is increasing [5] as a result in the “Transport and storage” industry the relative decrease of the CO pollutant emissions is the highest among all the industries.

2. “Real estate activities industry” is the least pollutant in terms of NMVOC emissions across all the industries over the period from the year 1990 to the year 2016. The reason for this assumption is that the European Environment Agency does not have the separate indicators of the NMVOC emissions for this industry [7].
3. There is a long-term trend of decreasing of the all mentioned types of emissions pollutant across all the industries to meet requirements of the EU National Emission Ceilings Directive.

To achieve the goals of this thesis the basic time series characteristics and methods for the estimation of the trend component will be employed. All the results will be used for the comparison of different industries from varieties of perspectives.

The thesis is divided into two main parts. The selected methods for time series analysis are explained in the first part. The second part is devoted to the application of those methods on the data when pursuing achievement of the goals as they are defined above.

For the purpose of the data analysis Microsoft Excel and EViews software were used.

# 1 Basic terms and methodology

## 1.1 Time series

Time series is defined as a sequence of observations of well-defined data items obtained through measurements over time (i.e. unambiguously organised from past to present). Those observations must be spatially and objectively comparable.

## 1.2 Times series classification

By a decisive time period:

- **Stock series** – measuring attribute refers to a specific point in time which is not influenced by the observing period. For instance, a labour force survey. It takes stock if the person was employed in the referenced time period.
- **Flow series** – indicates the cumulative value of activity over a given period. In order to compare those values, the length of the period should be the same. Their values can be added up. The most common example is Gross Domestic Product time series.

By the length of the observation period:

- **High-frequency time series** – the observation periodicity is **shorter** than a week.
- **Short term time series** – the observing period is less than one year e.g. monthly and quarterly time series.
- **Long term time series** – time series with a periodicity of a year or more.

## 1.3 Basic characteristics of time series

For the purpose of this thesis, the long-term flow series data are being used. This chapter will be dedicated to the description of the characteristics related to it.

### 1.3.1 Descriptive characteristics

The most widely used descriptive characteristic of the time series is a simple arithmetic mean. The arithmetic mean of the flow time series can be easily calculated by the following formula

$$\bar{y} = \frac{\sum_{t=1}^{n} y_t}{n}$$

Where $y_t$ are measured values, $t = 1, \ldots, n$ is a given time (time variable) and $n$ is number of observations (i.e. length of a time series).

### 1.3.2 Measures of dynamics

Simple measures of dynamics provide so-called basic characteristics of time series development.

**Absolute increases,** also known as the first differences, express the absolute increase or decrease of the value in time $t$ in comparison with the value in time $t-1$. If the absolute increase is constant, that means there is a linear trend in the time series.

$$\Delta_t = y_t - y_{t-1} \text{ where } t = 2,3,\ldots,n$$

**Mean absolute increase** indicates the mean the inter-period change in the time series during the observation period.

$$\bar{\Delta} = \frac{\sum_{t=2}^{n} \Delta_t}{n-1} = \frac{y_n - y_1}{n-1}$$

**Growth coefficients** express the relative change of the value in time $t$ in comparison with the value in time $t-1$. Can be expressed in percentage if multiplied by 100.

$$k_t = \frac{y_t}{y_{t-1}} \text{ where } t = 2,3,\ldots,n$$

**Mean growth coefficient** indicates the mean of the inter-period relative change in the time series during the observation period.

$$\bar{k} = \sqrt[n-1]{k_2 k_3 \ldots k_n} = \sqrt[n-1]{\frac{y_n}{y_1}}$$

Note that the mean values have a major disadvantage since it is influenced only by the first and the last values of the time series.

### 1.3.3 Measures of dispersion (variation)

**Variance** can be defined as the measure of how far the values of observations are spread out from the mean value. Mathematically it is the expected value of the squared difference of the observed variable value and variable's mean.

$$s_y^2 = \frac{\sum_{t=1}^{n} (y_t - \bar{y})^2}{n}$$

**Standard deviation** has the same meaning as the variance – mean deviation from the arithmetic mean of the observed values but in the same units of measure as the time series data so the value of standard deviation can be interpreted.

$$s_y = \sqrt{s_y^2} = \sqrt{\frac{\sum_{t=1}^{n}(y_t - \bar{y})^2}{n}}$$

**Relative standard deviation** or **coefficient of variation** is a metric used for comparison of the dispersion of the multiple datasets with different units of measurement (with different means). Calculated as a ratio of the standard deviation and mean. Commonly expressed as a percentage when multiplied by 100. As a disadvantage, it may have any value from $\langle -\infty; \infty \rangle$.

$$V_y = \frac{s_y}{\bar{y}}$$

All the measures stated in this chapter have a serious disadvantage. In case if the time series has a trend those characteristics are devalued by it.

## 1.4 Classical Time Series Decomposition

In the classic time series analysis approach is assumed that the time series can be decomposed on up to four different components. Components are divided into deterministic and irregular ones.

Deterministic components:

1. **Trend component** ($T_t$) – defined as a main long-term development of the time series.
2. **Cyclical component** ($C_t$) – demonstrates fluctuations around the trend with the periods greater than one year.
3. **Seasonal component** ($S_t$) – represents repetitive movements around the main trend with the periods of one year or less. Can be accrued only in the short-term time series.

**Irregular component** ($\varepsilon_t$) – a residual measure, represents random variations within the time series, for instance, measurement errors. It remains if all the deterministic components are "removed" from a time series.

There are two decomposition models: additive and multiplicative.

- **Additive model** is used if the variation of the time series is constant in time and values of the components are in the same units as the time series.
$$y_t = T_t + C_t + S_t + \varepsilon_t$$
- In case of **multiplicative model** only the value of the trend component is in the same unit as the time series, other components expressed relatively. This model is used in case when the variation of time series changes in time.
$$y_t = T_t C_t S_t \varepsilon_t$$

## 1.5 Trend component analysis

Trend component analysis serves a few purposes. Trend component expresses a long-term tendency of the time series describing the historical movements within the time series and isolating cyclical and seasonal components. In case of an irregular time series development or outliers the adaptive methods such as exponential smoothing or moving average methods should be used. If the progression of the time series corresponds to the function of time (time variable), the mathematical model must be applied.

### 1.5.1 Trend function

Trend function is a special case of the regression function. In this case, a time variable $t$ is used as an explanatory variable. There are the following assumptions: the time series $y_t$ is a series of the arranged values in time $t$ which were gained from the measuring of certain values in the same intervals of time $t$ (e.g. annually, quarterly, monthly), stability of the trend parameters in time is expected and the time series $y_t$ can be described as:

$$y_t = T_t + \varepsilon_{t.}$$

Where $T_t$ is a deterministic component expressing the trend and $\varepsilon_t$ is a random irregular component with the white noise properties:

- zero expected value $E(\varepsilon_t) = 0$
- constant variance $D(\varepsilon_t)$=$\sigma_\varepsilon^2$
- mutually uncorrelated (linear independence)
- come from a normal distribution ($\varepsilon_t \sim N(0, \sigma^2)$)

For the purpose of this thesis only 6 frequently used trend functions will be explained.

**Constant** trend, when the values of the trend do not change in time:

$$T_t = \beta_0 \text{ where } t = 1,2, \dots, n.$$

Estimation of $\beta_0$ is calculated by the method of the least squares as an average of the time series values.

$$\widehat{\beta_0} = \overline{y_t} = \frac{1}{n}\sum_{t=1}^{n} y_t$$

The trend estimation in time $t$ is:

$$\widehat{T_t} = \widehat{y_t} = \bar{y}.$$

**Linear** trend function:

$$T_t = \beta_0 + \beta_1 t \text{ where } t = 1,2, \dots, n.$$

Parameters $\beta_0$ and $\beta_1$ of the linear trend are estimated by the least squares method.

$$\widehat{\beta_0} = \bar{y} - \widehat{\beta_1}\bar{t} \qquad \widehat{\beta_1} = \frac{\sum_{t=1}^{n} t y_t - \bar{t}\sum_{t=1}^{n} y_t}{\sum_{t=1}^{n} t^2 - n(\bar{t})^2} \qquad \bar{t} = \frac{1}{n}\sum_{t=1}^{n} t$$

Estimation of the linear trend is:

$$\widehat{T_t} = \widehat{y_t} = \widehat{\beta_0} + \widehat{\beta_1}t.$$

**Quadratic** trend function:

$$T_t = \beta_0 + \beta_1 t + \beta_2 t^2 \text{ where } t = 1,2,\dots,n.$$

Parameters of the quadratic function $\beta_0, \beta_1$ and $\beta_2$ are estimated by the method of the least squares. Estimation of the quadratic trend is:

$$\widehat{T_t} = \widehat{y_t} = \widehat{\beta_0} + \widehat{\beta_1}t + \widehat{\beta_2}t^2.$$

**Exponential** trend function:

$$T_t = \beta_0 \beta_1^t \text{ where } t = 1,2,\dots,n.$$

This trend function has an important property that the quotient of the two neighbour values $T_t$ and $T_{t-1}$ is always the same number equals to the parameter $\beta_0$.

In order to apply the method of least squares to estimate the parameters $\beta_0,\ \beta_1 > 0$ of the exponential trend function, it should be logarithmically transformed:

$$\ln(T_t) = \ln(\beta_0) + \ln(\beta_1)t \text{ where } t = 1,2,\dots,n.$$

After this, we can conduct the following substitutions:

$$\ln(T_t) = T_t', \ln(\beta_0) = \beta_0', \ln(\beta_1)t = \beta_1' t,$$

which will give us the trend function, which has a similar form as the linear trend function. Estimations and verifications are conducted in this logarithmic form, with the same rules as for the linear trend function, but in the last step a backward calculation based on the substitution rules above must be performed.

$$T_t' = \beta_0' + \beta_1' t$$

**Hyperbolic** trend function:

$$T_t = \beta_0 + \beta_1 \frac{1}{t} \text{ where } t = 1,2,\dots,n.$$

Parameters of the quadratic function $\beta_0, \beta_1$ are estimated by the method of the least squares.

**Logarithmic** trend function:

Parameters of the quadratic function $\beta_0, \beta_1$ and $\beta_2$ are estimated by the method of the least squares.

$$T_t = \beta_0 + \beta_1 \ln(\mathrm{t}) \text{ where } t = 1,2, \ldots, n.$$

### 1.5.2 Trend function selection

Three most frequently used approaches to selecting a trend function will be briefly described in this section.

**Graph analysis** - based on a visualisation of time series. At the beginning of the analysis, preliminary selection of the trend function can be done with the help of graphical representation of the time series or the graphical analysis of the absolute increases and the growth coefficients. For instance, if the first differences fluctuate around zero, constant trend function should be used. For the visualisation of time series line chart is the most suitable one. Based on this chart it is possible to visually estimate the approximate time series progression, i.e. what kind of the trend function could be used. Selection of the trend function based on the graph analysis is subjective.

**Interpolation criteria** – simply speaking, those criteria help to evaluate the appropriateness of the chosen trend function. Most commonly used method called **Mean Squared Error** (MSE) is based on measuring the average squared difference between the original and estimated values. Those differences are called residuals:

$$\widehat{\varepsilon_t} = (y_t - \widehat{y_t}),$$

$$MSE = \frac{1}{n} \sum_{t=1}^{n} \widehat{\varepsilon_t}^2.$$

The lower value of the Mean Squared Error, the better-fitting trend function. In the EViews Root Mean Squared Error statistic calculated:

$$RMSE = \sqrt{MSE}.$$

**Coefficient of determination** is another quality indicator of the model. It provides a measure of how good observed values are replicated by the model, ranges from 0 to 1. Than closes the coefficient of determination to 1, than better model captures the trend of the time series.

$$R^2 = 1 - \frac{\sum_{t=1}^{n} (y_t - \hat{y}_t)^2}{\sum_{t=1}^{n} (y_t - \bar{y})^2}$$

Since the coefficient of determination depends on the number of explanatory variables of the model (number of parameters), it growth with the increase of numbers of parameters, it is unsuitable for comparison of models with a different number of parameters. In this case, **adjusted coefficient of determination** should be used.

$$R_{adj}^2 = 1 - (1 - R^2)\frac{n-1}{n-p-1}, \text{where } p \text{ is a number of parameters}$$

**F-test of Overall Significance** helps to identify if at least one of the parameters of the chosen trend function is not equal to zero. In other words, if the intercept-only model is not equal to the selected model**.**

$$H_0: \beta_0 = c; \beta_1 = \beta_2 = \cdots \beta_{p-1} = 0$$

$$H_1: non\ H_0$$

For this test, the following test statistic should be used:

$$F = \frac{\frac{S_t}{p-1}}{\frac{S_r}{n-p}}, F \sim F(p-1; n-p).$$

where $p$ is a number of parameters, $n$ – number of observations, $S_t$ is a total sum of squares and $S_r$ is a residual sum of squares.

$$S_t = \sum_{t=1}^{n} (y_t - \bar{y}_t)^2$$

$$S_r = \sum_{t=1}^{n} (y_t - \hat{y}_t)^2$$

In case if the test statistic is greater than the critical value of the F-distribution for selected significance level $\alpha$, the null hypothesis is rejected in favour of the alternative hypothesis.

Note that for each test in this thesis the 5% Level Of Significance ($\alpha$) assumed.

**Partial t-tests**: testing the model's parameters individually; if each parameter differs from zero. Rejecting the null hypothesis means that the parameter significantly differs from zero, practical interpretation being that and the corresponding explanatory variable belongs to the model.

$$H_0: \beta_i = 0 \text{ where } i = 0,1,\dots,p-1$$

$$H_1: non\ H_0$$

The test statistic is the following:

$$T = \frac{\widehat{\beta_i}}{s_{\widehat{\beta_i}}}.$$

With the rejection region:

$$W_\alpha = \{t; |t| \geq t_{1-\alpha/2}\}.$$

### 1.5.3 Residuals diagnostics

One of the assumed properties of the irregular component is a mutual independence of the values in time $t$ and time $t - k$ (where $k$ is so-called lag). Below there will be described a way to test this assumption.

**Durbin–Watson statistic** used in order to detect the presence of autocorrelation in the irregular component at lag 1. The null hypothesis of the Durbin–Watson test ($H_0$) states that the irregular component is not autocorrelated against the alternative ($H_1$) that it follows a first order autoregressive process (this means that the current value is based on the previous value).

$$H_0: \rho_1 = 0$$

$$H_1: non\ H_0$$

The test statistic can be calculated with this formula:

$$DW = \frac{\sum_{t=2}^{n}(\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1})^2}{\sum_{t=1}^{n} \hat{\varepsilon}_t} \approx 2(1 - \hat{\rho}_1) \in \langle 0,4 \rangle$$

The result of the test can be approximately interpreted as follows:

- if the value of $DW$ test statistic is near 2, the null hypothesis is not rejected.
- If the value of $DW$ test statistic is near 0 or 4, the null hypothesis is rejected in favour of the alternative.

### 1.5.4 Smoothing methods

In order to remove the effect of random variations from the observations of time series smoothing methods should be used (also known as adaptive methods). Assumptions of the smoothing methods:

- stability of the trend parameters in time is not expected.
- estimated values of the trend component are adapting to the changes in the time series progression.

**Moving averages method** is a technique based on replacing the time series values by the means calculated from those values. Trend component is estimated separately in the short sections of a time series by calculating an average in each of these short sections (windows). Then moving forward with the section through the series by excluding the first value and adding the following one. There are three basic types of moving averages:

- **Simple moving average** is a simple arithmetic mean of a few consecutive time series values, length of the section is an odd number

$$\overline{y_t} = \frac{y_{t-d} + \cdots + y_t + \cdots + y_{t+d}}{m},$$

where $m$ is a length of the section and $d = \frac{(m-1)}{2}$ (number of values before and after time $t$)

- **Weighted moving average** – weighted arithmetic mean of several consecutive values, sum of the weights equals 1. Can be used for both odd and even section lengths.

- **Centred moving average** is used in case if the length of the section is an even number.

$$\overline{y_t} = \frac{y_{t-d} + 2y_{t-d+1} + \cdots + 2y_t + \cdots + 2y_{t+d-1} + y_{t+d}}{2m}$$

The higher value of $m$, the stronger smoothing. If a time series contains a seasonal component, $m$ should be equal to the length of seasonality. The main disadvantages of this technique are the loosing values at the beginning and the end of a time series and the sensitivity of the extreme values.

# 2 Time series analysis

The aim of this bachelor thesis is to describe the development of the emissions time series from harmful pollutants into the atmosphere in the UK by the industry sectors to evaluate regulatory controls applied in this country. The data were published by the Office for National Statistics.  At our disposal there are time series for a period 1990 - 2016. The next step is to analyse the emissions by the industries. Same as the overall time series, the time series by the industries is from the year 1990 to 2016. By the Office for National Statistics in total 21 industries were defined (excluding the natural world emissions).

This thesis is limited to the analysis of the emissions of the four pollutants: Particulate Matter (PM10 and PM2.5), Carbon monoxide (CO), Non-methane volatile organic compound emissions (NMVOC) for the reasons stated in the introduction (significant influence on human health)

## 2.1 Non-methane volatile organic compound emissions

### 2.1.1 Non-methane volatile organic compound emission in the United Kingdom

In this section, a time series of the total emissions of a non-methane volatile organic compound in the UK will be analysed. In the figure below the progression of the overall non-methane volatile organic compound emissions in years 1990-2016 is displayed.

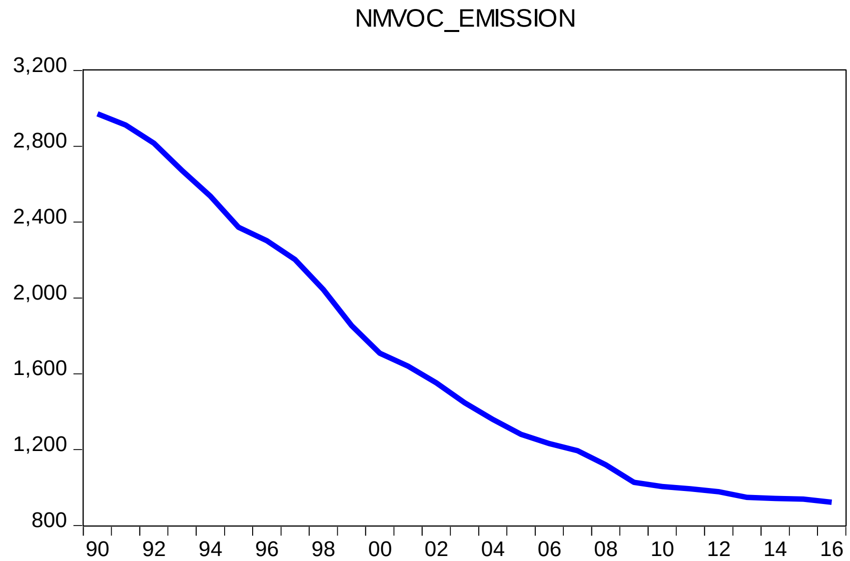


**Figure 1.** Total non-methane volatile organic compound emissions progression in the UK

**Table 1.** Basic characteristics of the NMVOC emissions in the UK

| Arithmetic mean | Mean absolute increase | Mean growth coefficient | Coefficient of variation | Ratio 2016/1990 |
|---|---|---|---|---|
| 1665.813 | -78.818 | 0.956 | 0.411 | 0.310 |

Table 1 contains the basic characteristics of the total non-methane volatile organic compound emissions in the United Kingdom in the years 1990 – 2016. For this period of time the United Kingdom has emitted 1665.813 thousand tonnes of the NMVOC pollutants at average per year. The mean growth coefficient of 0.956 can be interpreted as an average year-to-year decrease of the NMVOC pollutant emissions equals to 4.4%. It is worth to mention that the NMVOC emissions in 2016 have only 31% of the volume of emissions in the year 1990.

The next step is the estimation of the selected trend functions' parameters. Those estimations will be conducted in the statistical software EViews 10 using the Least Squares method. Results are provided in the table below.

**Table 2.** Estimated parameters for various trend functions

| Trend function | $b_0$ | $b_1$ | $b_2$ |
|---|---|---|---|
| Constant | **1665.813** | | |
| **Linear** | 2853.699 | -84.849 | |
| Quadratic | **3278.261** | **-172.6894** | **3.137157** |
| **Exponential** | 3133.450 | 0.950 | |
| Hyperbolic | **1295.822** | **2567.103** | |
| **Logarithmic** | 3575.893 | -798.856 | |

In order to select the most suitable trend function for our data further calculations should be conducted. This purpose will serve the F-test of Overall Significance and partial t-tests.

**Table 3.** P-Values of the partial t-test and the F-test

| Trend function | Partial t-test of $b_1$ | Partial t-test of $b_2$ | F-test |
|---|---|---|---|
| Constant | | | |
| **Linear** | 0.000 | | 0.000 |
| Quadratic | 0.000 | 0.000 | 0.000 |
| **Exponential** | 0.000 | | 0.000 |
| Hyperbolic | 0.000 | | 0.000 |
| **Logarithmic** | 0.000 | | 0.000 |

As you can see each P-Value of each test is lower than the selected level of significance, since those P-Values are very close to zero in EViews those values displayed as 0.000000. Thus, we can assume that the parameters of each trend are statistically significant.

Our next step is to compare each trend function and select the most suitable one. For this comparation the statistical methods from the theoretical section of this thesis. Specifically, Coefficient of Determination (normal and adjusted) and Root Mean Squared Error.

**Table 4.** Criteria statistics for selecting suitable trend

| Trend function | $R^2$ | $R^2$ adj. | DW-statistic | RMSE |
|---|---|---|---|---|
| Constant | 0 | 0 | 0.018482 | 683.938 |
| **Linear** | 0.933703 | 0.931051 | 0.087004 | 176.1019 |
| Quadratic | 0.995396 | 0.995012 | 0.605366 | 46.40763 |
| **Exponential** | 0.976479 | 0.975539 | 0.17191 | 84.15926 |
| Hyperbolic | 0.546673 | 0.52854 | 0.300664 | 460.4928 |
| **Logarithmic** | 0.929744 | 0.926933 | 0.383711 | 181.2841 |

Based on those indicators – quadratic trend function is the most suitable for our data, it has the highest value of the Coefficient of Determination and it has significantly the lowest RMSE statistic.

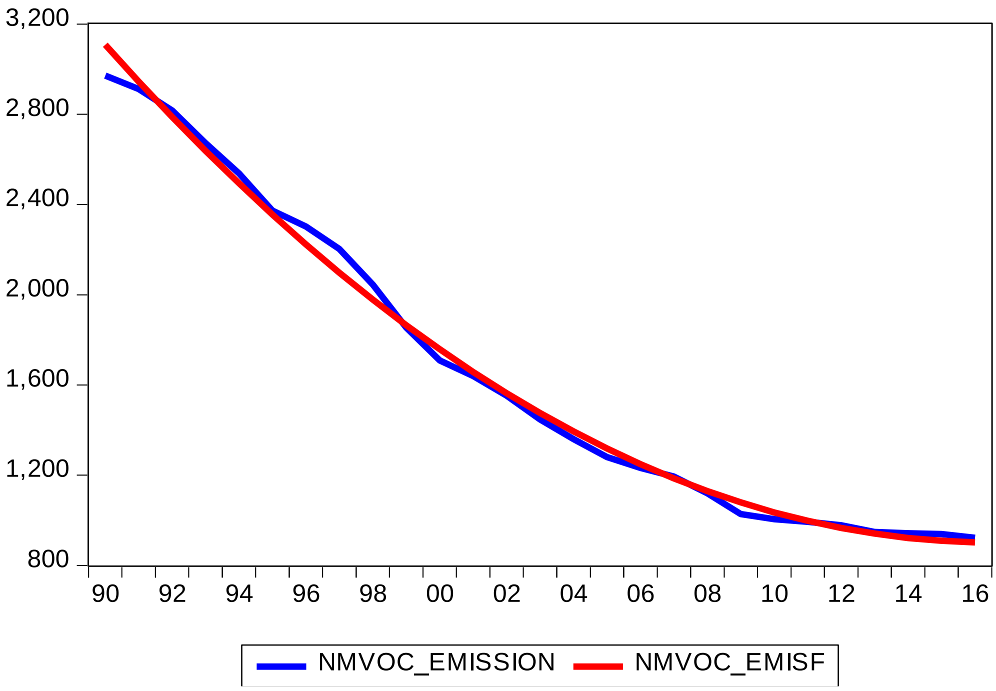


**Figure 2.** Graph of the actual data and quadratic trend function

### 2.1.2 Non-methane volatile organic compound emission by the industries

As was mentioned before, at our disposal there are also the data of NMVOC emissions classified by the 21 industry sectors. Once again, our analysis starts with the calculation of basic statistics.

**Table 5.** Basic characteristics of the NMVOC emissions by the industry (assorted from the highest arithmetic mean to the lowest)

| Industry | Arithmetic mean | Mean absolute increase | Mean growth coefficient | Coefficient of variation | Ratio 2016/1990 |
|---|---|---|---|---|---|
| Consumer expenditure | 452.389 | -24.520 | 0.949 | 0.489 | 0.253 |
| Manufacturing | 433.080 | -20.934 | 0.957 | 0.432 | 0.323 |
| Mining and quarrying | 232.803 | -17.987 | 0.922 | 0.625 | 0.122 |
| Agriculture, forestry and fishing | 214.536 | -1.256 | 0.994 | 0.051 | 0.861 |
| Wholesale and retail trade; repair of motor vehicles and motorcycles | 94.141 | -5.811 | 0.937 | 0.561 | 0.183 |
| Construction | 75.085 | -2.293 | 0.973 | 0.258 | 0.494 |
| Transport and storage | 48.633 | -1.810 | 0.961 | 0.347 | 0.357 |
| Electricity, gas, steam and air conditioning supply | 45.272 | -0.939 | 0.975 | 0.221 | 0.515 |
| Water supply; sewerage, waste management and remediation activities | 10.415 | -0.383 | 0.957 | 0.351 | 0.322 |
| Administrative and support service activities | 7.776 | -0.605 | 0.892 | 0.707 | 0.052 |
| Public administration and defence; compulsory social security | 7.267 | -0.648 | 0.920 | 0.768 | 0.115 |
| Other service activities | 4.777 | -0.460 | 0.901 | 0.869 | 0.066 |
| Professional, scientific and technical activities | 4.320 | -0.419 | 0.875 | 0.906 | 0.031 |
| Information and communication | 3.332 | -0.344 | 0.867 | 0.955 | 0.025 |
| Arts, entertainment and recreation | 2.020 | -0.084 | 0.957 | 0.375 | 0.320 |
| Activities of households as employers | 2.014 | -0.110 | 0.908 | 0.522 | 0.082 |
| Accommodation and food services | 1.749 | -0.120 | 0.928 | 0.624 | 0.143 |
| Human health and social work activities | 1.373 | -0.095 | 0.925 | 0.643 | 0.132 |
| Real estate activities | 0.925 | -0.103 | 0.860 | 1.018 | 0.020 |
| Education | 0.656 | -0.051 | 0.914 | 0.718 | 0.095 |
| Financial and insurance activities | 0.022 | -0.002 | 0.902 | 0.711 | 0.068 |

Based on the arithmetic mean, the top 3 of the most polluting industries are consumer expenditure, manufacturing and mining, and quarrying, which is not surprising. At the same time, those industries show a significant decrease in the emissions of NMVOC as in the year 2016 they demonstrated only 25.3%, 32.3% and 12.2% of the emission of the reference values in the year 1990 respectively. The least Non-methane volatile organic compound pollutant industries are real estate activities, education and Financial and insurance activities, with the average for the whole period less than one thousand tonnes. Nevertheless, the emission in those industries drastically decreased during the observed period and in the year 2016 each of those them emitted less than 10% of the emissions of the year 1990.

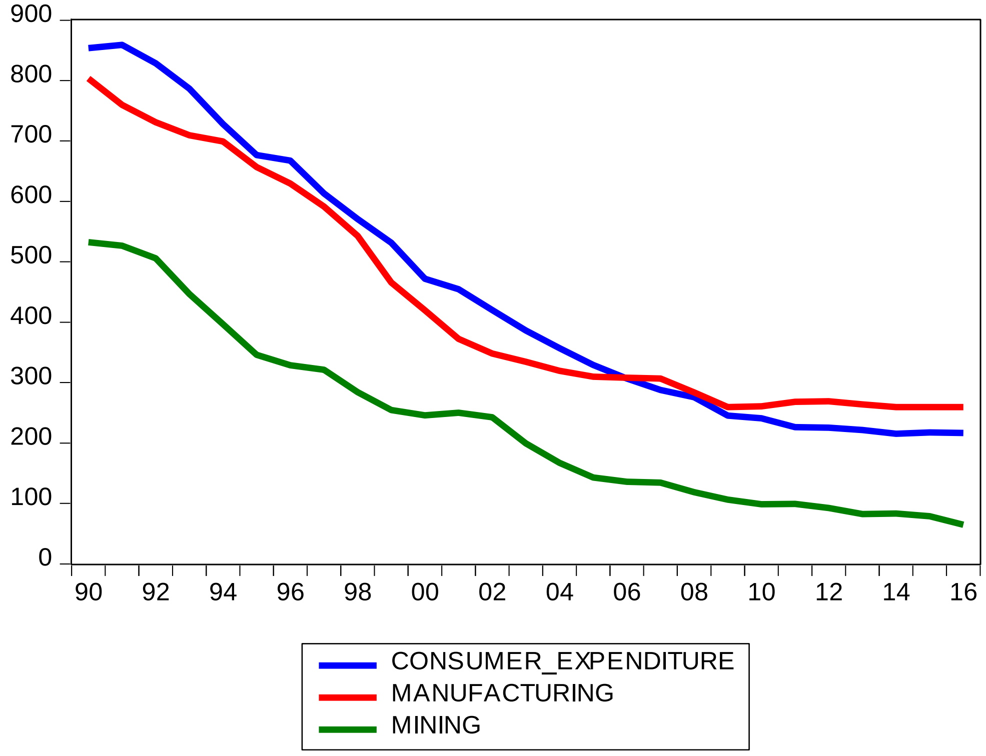


**Figure 3.** Progression of the 3 industries with the highest average of the NMVOC emission

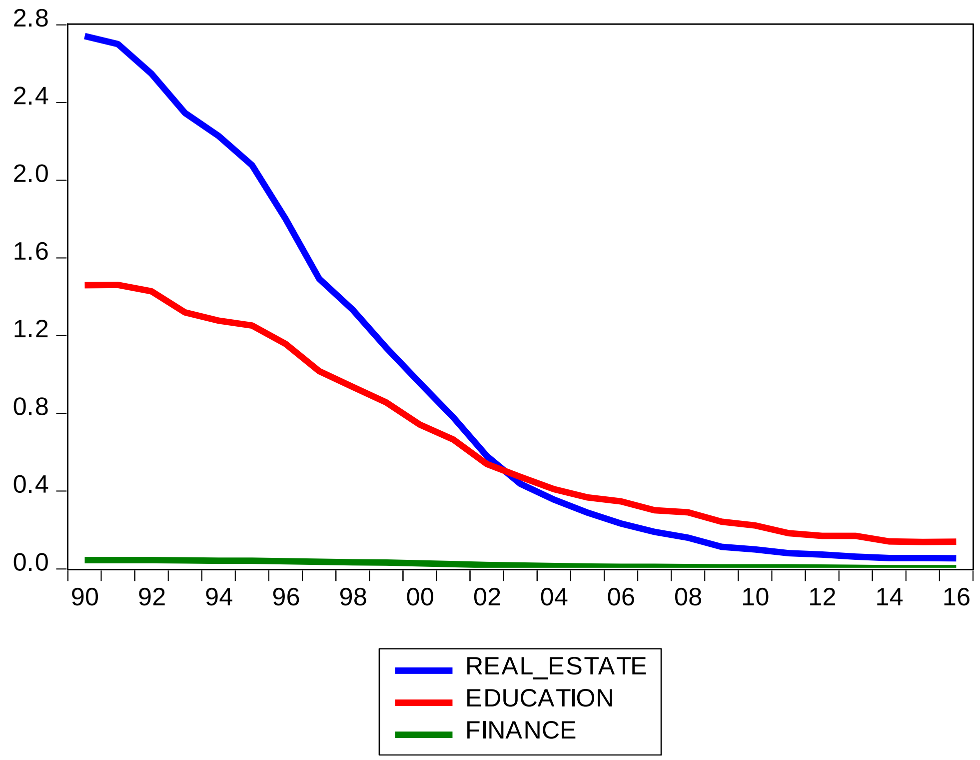


**Figure 4.** Progression of the NMVOC emission for the least emitting industries

The next step is to identify the most suitable trend for each industry and provide the relative metrics and statistics for each of them. Once more, the estimation will be tested by the partial t-tests and F-test of Overall Significance. Basically, the same approach as in the previous chapter will be used.

**Table 6.** Estimation of the parameters for selected trend function and P-Values for the t-test and F-tests

| Industry | Selected trend function | $b_0$ | $b_1$ | $b_2$ | Partial t-test of $b_1$ | Partial t-test of $b_2$ | F-test |
|---|---|---|---|---|---|---|---|
| Accommodation and food services | Quadratic | 4.207 | -0.250 | 0.004 | 0 | 0 | 0 |
| Activities of households as employers | Linear | 3.846 | -0.131 | - | 0 | - | 0 |
| Administrative and support service activities | Quadratic | 19.191 | -1.040 | 0.012 | 0 | 0 | 0 |
| Agriculture, forestry and fishing | Quadratic | 237.734 | -2.200 | 0.030 | 0 | 0 | 0 |
| Arts, entertainment and recreation | Quadratic | 3.522 | -0.128 | 0.001 | 0 | 0 | 0 |
| Construction | Quadratic | 121.333 | -5.075 | 0.097 | 0 | 0 | 0 |
| Consumer expenditure | Quadratic | 969.613 | -54.943 | 0.982 | 0 | 0 | 0 |
| Education | Quadratic | 1.733 | -0.112 | 0.002 | 0 | 0 | 0 |
| Electricity, gas, steam and air conditioning supply | Quadratic | 45.130 | 1.824 | -0.099 | 0 | 0 | 0 |
| Financial and insurance activities | Quadratic | 0.053 | -0.003 | 0.000 | 0 | 0 | 0 |
| Human health and social work activities | Quadratic | 3.330 | -0.195 | 0.003 | 0 | 0 | 0 |
| Information and communication | Quadratic | 11.112 | -0.875 | 0.017 | 0 | 0 | 0 |
| Manufacturing | Quadratic | 899.216 | -53.944 | 1.126 | 0 | 0 | 0 |
| Mining and quarrying | Exponential | 604.552 | 0.921 | - | 0 | - | 0 |
| Other service activities | Quadratic | 15.095 | -1.182 | 0.024 | 0 | 0 | 0 |
| Professional, scientific and technical activities | Quadratic | 13.662 | -1.023 | 0.019 | 0 | 0 | 0 |
| Public administration and defence; compulsory social security | Exponential | 2.950 | -0.087 | - | 0 | - | 0 |
| Real estate activities | Quadratic | 3.296 | -0.276 | 0.006 | 0 | 0 | 0 |
| Transport and storage | Quadratic | 79.757 | -2.371 | 0.008 | 0 | 0 | 0 |
| Water supply; sewerage, waste management and remediation activities | Quadratic | 15.745 | -0.228 | -0.008 | 0 | 0 | 0 |
| Wholesale and retail trade; repair of motor vehicles and motorcycles | Quadratic | 210.518 | -11.559 | 0.177 | 0 | 0 | 0 |

P-Values of every partial t-test and F-tests of overall significance a much lower than the chosen Level of Significance ($\alpha$) so it makes sense to use those trend functions with those parameters.

For most of the industries the quadratic function has been used. "Activities of households as employers" industry had better-adjusted Coefficient of Determination with the linear trend against the alternatives. Both industries modelled with an exponential trend ("Public administration and c" and "Transport and storage") had very similar values of the adjusted Coefficient of Determination with the quadratic trend, but the exponential trend displayed the lower Root Mean Squared Error statistic.

**Table 7.** Key statistics of the chosen trend functions

| **Industry** | **$R^2$** | **$R^2$ adj.** | **DW-statistic** | **RMSE** |
|---|---|---|---|---|
| Accommodation and food services | 0.987 | 0.986 | 0.452 | 0.123569 |
| Activities of households as employers | **0.940889** | **0.941** | **0.111** | **0.255478** |
| Administrative and support service activities | 0.990 | 0.989 | 0.302 | 0.55726 |
| Agriculture, forestry and fishing | **0.965103** | **0.962** | **1.104** | **2.053089** |
| Arts, entertainment and recreation | 0.990 | 0.989 | 0.378 | 0.076007 |
| Construction | **0.983681** | **0.982** | **0.738** | **2.470591** |
| Consumer expenditure | 0.994 | 0.993 | 0.800 | 17.52749 |
| Education | **0.9844** | **0.983** | **0.382** | **0.058783** |
| Electricity, gas, steam and air conditioning supply | 0.833 | 0.819 | 1.141 | 4.086651 |
| Financial and insurance activities | **0.963319** | **0.960** | **0.230** | **0.002944** |
| Human health and social work activities | 0.985 | 0.984 | 0.377 | 0.108122 |
| Information and communication | **0.986246** | **0.985** | **0.373** | **1.354439** |
| Manufacturing | 0.979 | 0.977 | 0.317 | 27.33123 |
| Mining and quarrying | **0.989608** | **0.989** | **0.878** | **14.72642** |
| Other service activities | 0.988 | 0.986 | 0.357 | 0.463746 |
| Professional, scientific and technical activities | **0.984913** | **0.984** | **0.406** | **0.480524** |
| Public administration and defence; compulsory social security | 0.960 | 0.959 | 0.328 | 1.612866 |
| Real estate activities | **0.989223** | **0.988** | **0.414** | **0.097737** |
| Transport and storage | 0.979 | 0.978 | 0.923 | 2.432461 |
| Water supply; sewerage, waste management and remediation activities | **0.981416** | **0.980** | **0.298** | **0.498354** |

| | | | | |
|---|---|---|---|---|
| Wholesale and retail trade; repair of motor vehicles and motorcycles | 0.981 | 0.979 | 0.519 | 7.283433 |

In the introduction I have speculated about "Real estate activities" industry being the least pollutant in terms of NMVOC emission over the studied period. Since I have at disposal all the necessary statistics, I can say that "Financial and insurance activities" industry is the less pollutant at this matter. But "Real estate activities" industry demonstrates the highest relative decrease of the emissions of this pollutant, with the 14% year-to-year decrease, in the year 2016 this industry emitted only 2% of the volume of the emissions of the year 1990.

## 2.2 Carbon monoxide emissions

### 2.2.1 Overall carbon monoxide emissions in the United Kingdom

In this section it will be described the progression of the carbon monoxide emission in the UK over the same time period. The same way as previously, this time series from the year 1990 to the year 2016 will be visualised with a line chart.

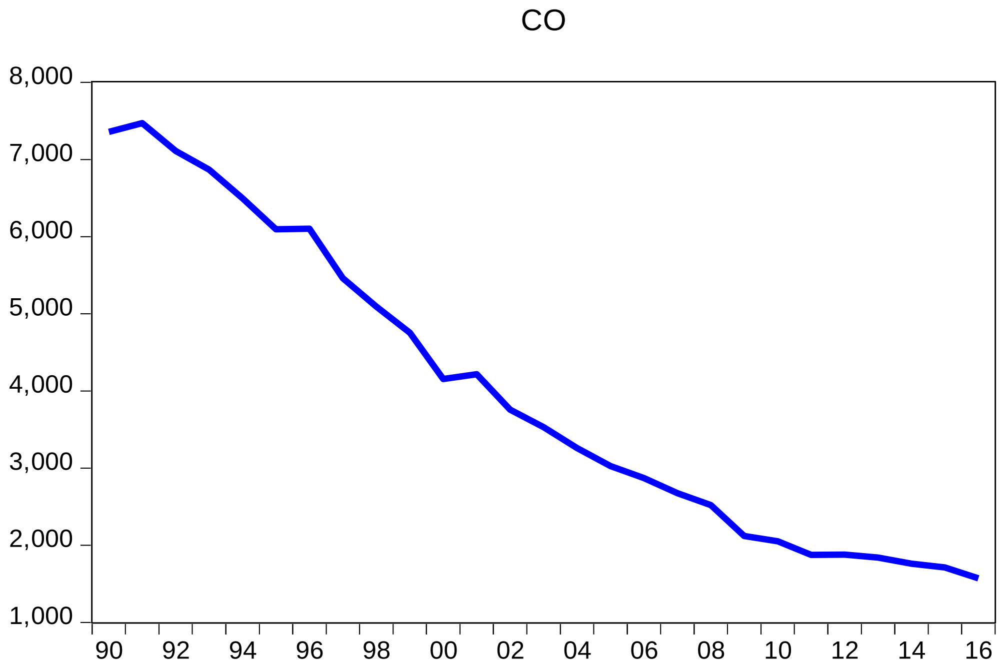


**Figure 5.** Overall CO emissions in the UK 1990-2016

Once again, we can observe a constant decrease of the pollutant emission with the three noticeable breakpoints in the years 1991, 1996 and 2001.

Now the basic characteristics can be calculated.

**Table 8.** Basic characteristics of the overall emissions of CO in the UK

| Arithmetic mean | Mean absolute increase | Mean growth coefficient | Coefficient of variation | Ratio 2016/1990 |
|---|---|---|---|---|
| 3985.266 | -222.589 | 0.942 | 0.493366513 | 0.21336213 |

During the time period from the year 1990 to the year 2016 at average there were emitted 3985.266 thousand tonnes of the carbon monoxide pollutant in the United Kingdom. The average year to year decrease in the emission is 6.8%. CO emissions in the year 2010 have only 21.3% of the emissions volume of the year 1990.

After analysis of the basic characteristics, the appropriate trend function should be chosen. The same approach as before will be used.

**Table 9.** Estimation of the parameters for various trend function, CO emissions

| Trend function | $b_0$ | $b_1$ | $b_2$ |
|---|---|---|---|
| Constant | **3985.266** | | |
| **Linear** | 7447.455 | -247.2992 | |
| Quadratic | **8328.892** | **-429.6656** | **6.513083** |
| **Exponential** | 8814.153 | 0.936228227 | |
| Hyperbolic | **2974.456** | **7013.28** | |
| **Logarithmic** | 9395.53 | -2262.743 | |

Partial t-tests and F-test must be conducted for the verification of those models.

**Table 10.** P-Values of the F-tests and partial t-test for each trend function, CO

| Trend function | Partial t-test of $b_1$ | Partial t-test of $b_2$ | F-test |
|---|---|---|---|
| Constant | | | |
| **Linear** | 0 | | 0 |
| Quadratic | 0 | 0 | 0 |
| **Exponential** | 0 | | 0 |
| Hyperbolic | 0 | | 0.000044 |
| **Logarithmic** | 0 | | 0 |

Unsurprisingly but each model passed those tests. Now the most suitable function must be chosen.

**Table 11.** Criteria statistics for selecting suitable trend, CO

| Trend function | $R^2$ | $R^2$ adj. | DW-statistic | RMSE |
|---|---|---|---|---|
| Constant | 0 | 0 | 0.021361 | 1966.197 |
| **Linear** | 0.959714 | 0.958102 | 0.227655 | 394.644 |
| Quadratic | 0.991889 | 0.991213 | 1.026571 | 177.0826 |
| **Exponential** | 0.989877 | 0.989472 | 0.698875 | 243.747 |
| Hyperbolic | 0.493701 | 0.473449 | 0.287918 | 1399.041 |
| **Logarithmic** | 0.902564 | 0.898666 | 0.39481 | 613.7446 |

Over again, the quadratic trend function has not the only better value of the adjusted Coefficient of Determination, but the lowest value of the Root Mean Squared Error. The parameters estimated by the least square method have following values: $\widehat{\beta_0} = 8328.892$, $\widehat{\beta_1} = 429.6656, \widehat{\beta_3} = 6.513083$. The trend estimation is:

$$\widehat{T_t} = 8328.892 - 429.6656t + 6.513083t^2$$

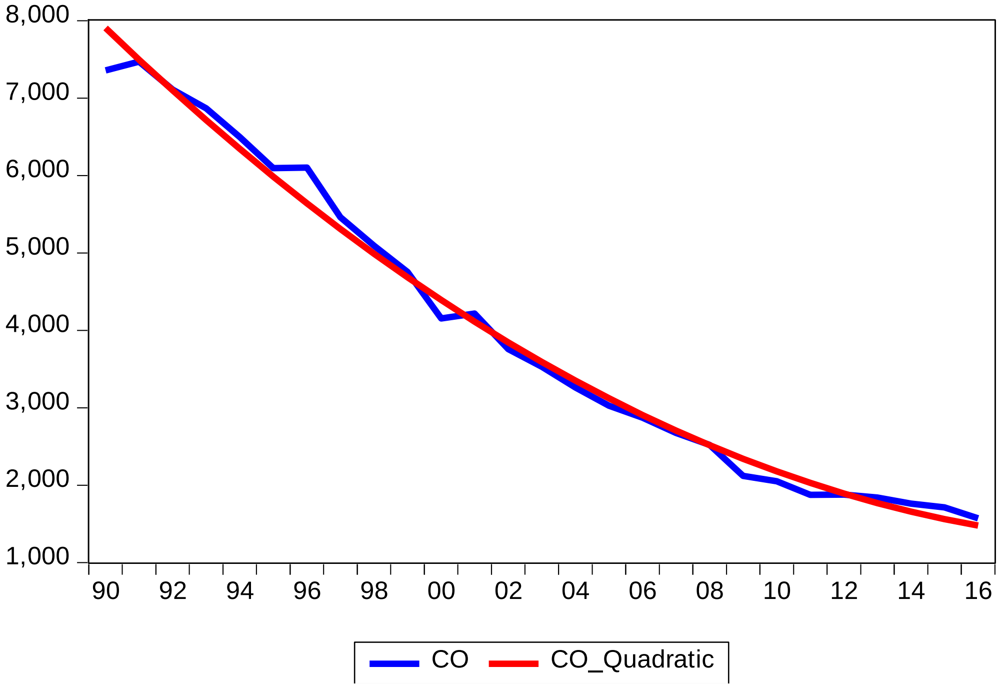


**Figure 6.** Actual progression end quadratic trend function, CO

## 2.2.2 CO emissions over the selected industries

As was mentioned in the introduction the rest of the uncovered emissions (CO, PM2.5, and PM10) will be analysed across the eight most emitting industries. In order to determine those industries arithmetic mean will be used.

**Table 12.** Top eight industries by the CO emission

| **Industry** | **Average** |
|---|---|
| Consumer expenditure | 2284.839 |
| **Manufacturing** | **667.799** |
| Construction | 277.623 |
| **Transport and storage** | **180.782** |
| Wholesale and retail trade; repair of motor vehicles and motorcycles | 118.520 |
| **Electricity, gas, steam and air conditioning supply** | **84.0368** |
| Agriculture, forestry and fishing | 68.210 |
| **Administrative and support service activities** | **54.2081** |

As usual, the basic characteristics such as mean absolute increase, mean growth coefficient and coefficient of variation will be calculated.

**Table 13.** Basic characteristics of the top 8 industries by CO emission

| Industry | Mean absolute increase | Mean growth coefficient | Coefficient of variation | Ratio 2016/1990 |
|---|---|---|---|---|
| Consumer expenditure | -140.0370385 | 0.93297749 | 0.554016309 | 0.164683583 |
| **Manufacturing** | **-28.29326923** | **0.955260984** | **0.380446455** | **0.304209691** |
| Construction | -6.964307692 | 0.977316638 | 0.247381231 | 0.550703205 |
| **Transport and storage** | **-8.317** | **0.947678769** | **0.424700774** | **0.247280537** |
| Wholesale and retail trade; repair of motor vehicles and motorcycles | -11.02988462 | 0.883455407 | 0.851108111 | 0.039884161 |
| **Electricity, gas, steam and air conditioning supply** | **-3.237423077** | **0.958592926** | **0.268577493** | **0.333034872** |
| Agriculture, forestry and fishing | -7.260653846 | 0.944808494 | 0.772992834 | 0.228527526 |
| **Administrative and support service activities** | **-3.294423077** | **0.940359136** | **0.570767269** | **0.20213311** |

Majority of those industries demonstrates the similar year to year decrease of the emission, in the range from 2.3% to 6.8%. But "wholesale and retail trade" industry has 11.7% year to year decrease and in the year 2016 emitted less than 4% of the emissions of the year 1990. So, the assumption about "Transport and storage" industry having the highest relative decrease of the CO pollutant emissions among all the industries is incorrect. Figures below will demonstrate the progressions of the CO emissions in the selected industries.

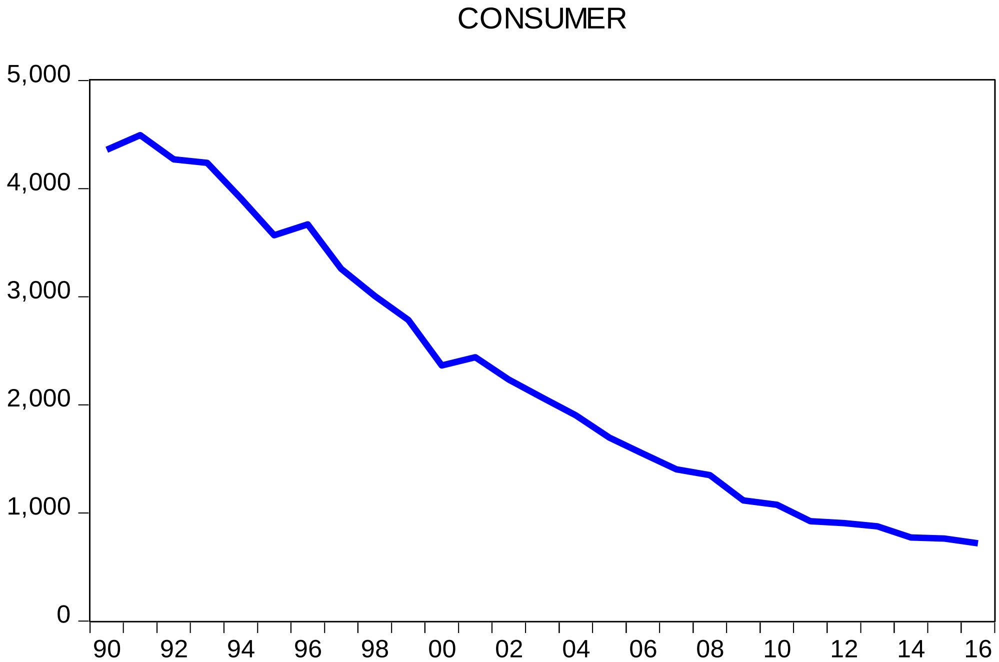


**Figure 7.** Progression of the CO emissions in the "Consumer" industry

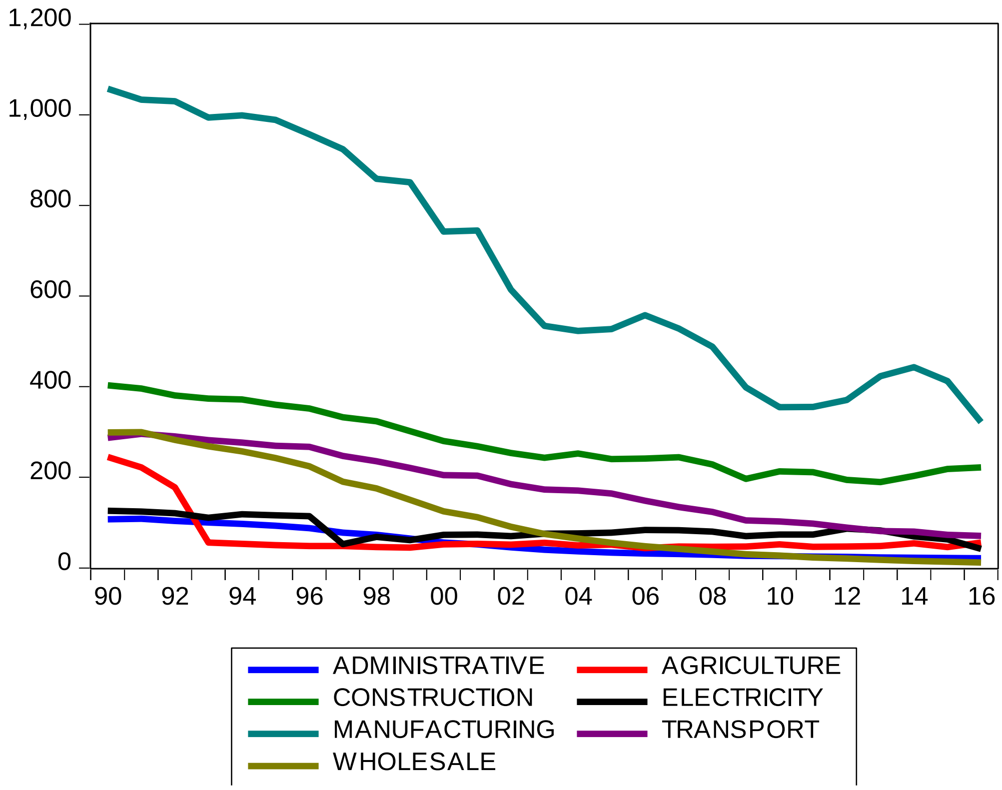


**Figure 8.** The progression of the rest 7 industries by CO emission

Since the “consumer industry” demonstrates much higher emissions, for better visualisation it was displayed in the separate graph.

The “manufacturing” industry shows significantly different development in comparison with the other industries. The following table contains the key statistics of selected trend functions.

**Table 14.** Selected trend function, parameters and P-Values of the tests

| Industry | Selected trend function | $b_0$ | $b_1$ | $b_2$ | Partial t-test of $b_1$ | Partial t-test of $b_2$ | F-test |
|---|---|---|---|---|---|---|---|
| Consumer expenditure | Quadratic | 5019.535 | -263.0242 | 3.692114 | 0 | 0 | 0 |
| **Manufacturing** | Quadratic | **1192.941** | **-49.03999** | **0.628898** | **0** | **0.0091** | **0** |
| Construction | Quadratic | 438.3465 | -17.34262 | 0.319766 | 0 | 0 | 0 |
| **Transport and storage** | Linear | **317.8435** | **-9.790098** | | **0** | | **0** |
| Wholesale and retail trade; repair of motor vehicles and motorcycles | Quadratic | 358.4411 | -26.11572 | 0.489737 | 0 | 0 | 0 |
| **Electricity, gas, steam and air conditioning supply** | Logarithmic | **134.9498** | **-21.29344** | | **0** | | **0** |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Agriculture, forestry and fishing | Hyperbolic | 33.3417 | 241.9287 | | 0 | | 0 |
| **Administrative and support service activities** | Quadratic | **127.2128** | **-7.899253** | **0.146434** | **0** | **0** | **0** |

Each of the selected trend functions passed the partial t-tests and F-test of the Overall Significance, P-Values of each test are much lower than the selected $\alpha$. CO emission trends of the five out of eight selected industries were described by the quadratic trend function, the most suitable trend functions "Transport and Storage", "Electricity, gas, steam and air conditioning supply" and "Agriculture, forestry and fishing" industries were defined by the linear, logarithmic and hyperbolic functions respectively.

**Table 15.** Criteria statistics of the selected trend functions, CO emissions by industries

| Industry | $R^2$ | $R^2$ adj. | DW-statistic | RMSE |
|---|---|---|---|---|
| Consumer expenditure | 0.989895 | 0.989053 | 1.207186 | 127.2467 |
| **Manufacturing** | **0.946472** | **0.942012** | **0.534713** | **58.77985** |
| Construction | 0.96876 | 0.966157 | 0.775499 | 12.13882 |
| **Transport and storage** | **0.986387** | **0.985843** | **0.558204** | **8.95804** |
| Wholesale and retail trade; repair of motor vehicles and motorcycles | 0.9863 | 0.985158 | 0.378437 | 11.80697 |
| **Electricity, gas, steam and air conditioning supply** | **0.606561** | **0.590824** | **0.990178** | **14.15721** |
| Agriculture, forestry and fishing | 0.816956 | 0.809634 | 1.492908 | 22.55819 |
| **Administrative and support service activities** | **0.980348** | **0.97871** | **0.324959** | **4.337409** |

Unfortunately, for the "Electricity, gas, steam and air conditioning supply" industry, the most appropriate trend function has a low value of the Coefficient of Determination in compare to the other industries, that means that none of the trend functions were able to describe a trend in this case. This is not a surprise, considering the very unusual progression of this industry's time series. Growth coefficient of the CO emission in the "Electricity, gas, steam and air conditioning supply" industry between the years 1996 and 1997 is 0.463, what is 53.7% decrease over one year. Therefore, how the trend component might be estimated in this case? For the decision making in the matter of the trend selection in cases like this, the moving averages method can be used.

For the smoothing of the CO emission progression from the "Electricity, gas, steam and air conditioning supply" industry, simple moving averages method was used, with the length of the moving segment $m = 5$. Note that the selection of the length is rather subjective. The result is in the figure below.

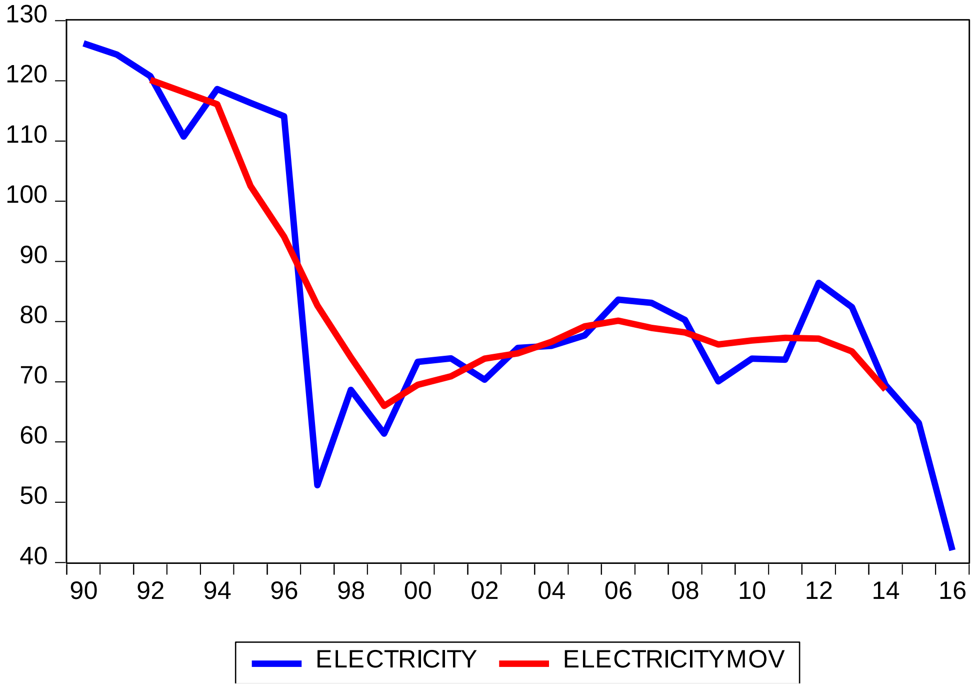


**Figure 9.** CO emission of "Electricity, gas, steam and air conditioning supply" industry, observations vs smoothed data

Now it is rather obvious that trend has the properties closed to the hyperbolic function.

## 2.3 PM2.5 pollutant emissions

### 2.3.1 Overall PM2.5 pollutant emissions in the United Kingdom

For PM2.5 pollutant analysis once again the line chart visualisation will be used.

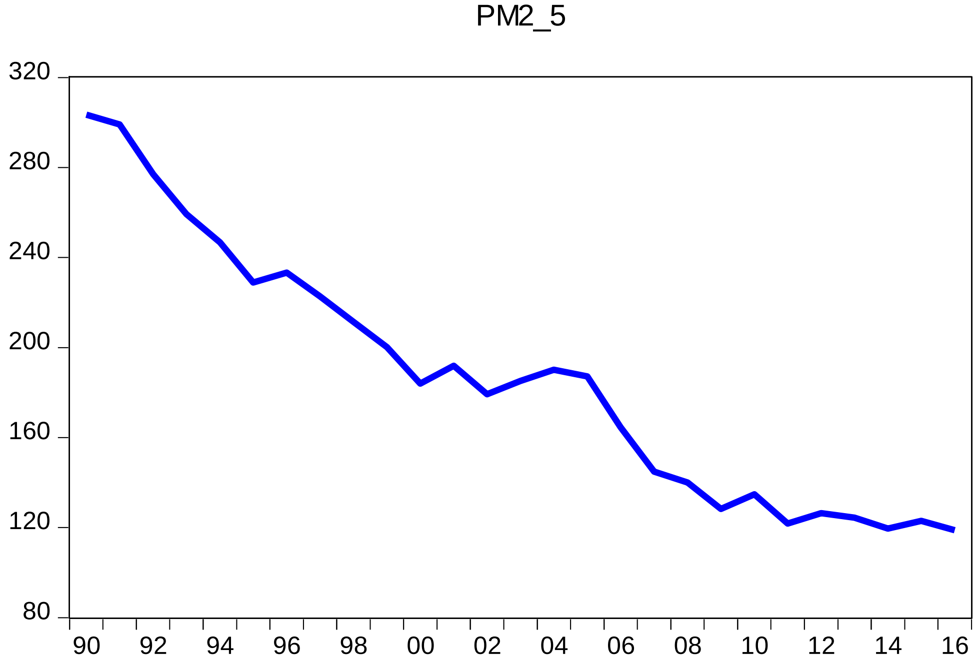


**Figure 10.** Overall PM2.5 emissions in the UK, 1990-2016

Development of the PM2.5 emissions has a noticeable difference from the previous ones due to the noticeable increases between the years 2002 and 2004 with the absolute increase for two years 10.880 thousand tonnes in total.

For better understanding of the data basic characteristics were calculated.

**Table 16.** Basic characteristics of the overall PM2.5 emissions time series

| Arithmetic mean | Mean absolute increase | Mean growth coefficient | Coefficient of variation | Ratio 2016/1990 |
|---|---|---|---|---|
| 186.893 | -7.100 | 0.965 | 0.299 | 0.392 |

At average 186.893 thousand tonnes of the PM2.5 pollutant per year there were emitted into the atmosphere in the UK for the period from 1990 to 2016. Year to year decrease of the PM2.5 emissions is 3.5%. The volume of the PM2.5 emissions in the year 2016 has only 39.2% of the volume of the year 1990. Coefficient of Variation, in this case it can be interpreted this way: values of the PM2.5 emissions time series vary from the arithmetic mean at average of 29.9%.

Next stage is to define the trend function for this PM2.5 emissions time series. Following tables contain the estimations of the parameters of the selected trend functions, P-Values of the partial t-tests and F-Tests of overall significance and the criteria statistic for selecting the best trend function for this time series.

**Table 17.** Estimation of the parameters of the selected trend functions, PM2.5 time series

| Trend function | $b_0$ | $b_1$ | $b_2$ |
|---|---|---|---|
| Constant | **186.8928** | | |
| **Linear** | 284.5196 | -6.97334 | |
| Quadratic | **310.2508** | **-12.297** | **0.190132** |
| **Exponential** | 301.9843 | 0.963259 | |
| Hyperbolic | **155.6345** | **216.8786** | |
| **Logarithmic** | 343.73 | -65.5943 | |

For those estimations the partial t-tests and F-tests were conducted. In the table below P-Values of those tests are recorded.

**Table 18.** Partial t-tests and F-tests of the trend functions, PM2.5 time series

| Trend function | Partial t-test of $b_1$ | Partial t-test of $b_2$ | F-test |
|---|---|---|---|
| Constant | | | |
| **Linear** | 0 | | 0 |
| Quadratic | 0 | 0 | 0 |
| **Exponential** | 0 | | 0 |
| Hyperbolic | 0 | | 0.000004 |
| **Logarithmic** | 0 | | 0 |

All the selected trend functions and their parameters passed those tests as their P-Values are lower than the selected Level of Significance $\alpha$. Now the most appropriate trend function must be selected.

**Table 19.** Criteria statistics for selection, PM2.5 emissions TS

| Trend function | $R^2$ | $R^2$ adj. | DW-statistic | RMSE |
|---|---|---|---|---|
| Constant | 0 | 0 | 0.042194 | 56.05425 |
| **Linear** | 0.938891 | 0.936447 | 0.437712 | 13.85674 |
| Quadratic | 0.972627 | 0.970346 | 0.853292 | 9.274114 |
| **Exponential** | 0.964554 | 0.963136 | 0.8049 | 9.804346 |
| Hyperbolic | 0.580888 | 0.564123 | 0.37081 | 36.2889 |
| **Logarithmic** | 0.933202 | 0.93053 | 0.617568 | 14.48737 |

Once again, the quadratic trend function will be selected. With the Coefficient of Determination 0.972627, it described 97.26% of the time series progression of the PM2.5 emissions over the studied time period. In this case, the equation of the trend function will be the following:

$$T_t = 310.2508 - 12.297t + 0.190132t^2$$

Graphical representation of the selected trend function will be placed below.

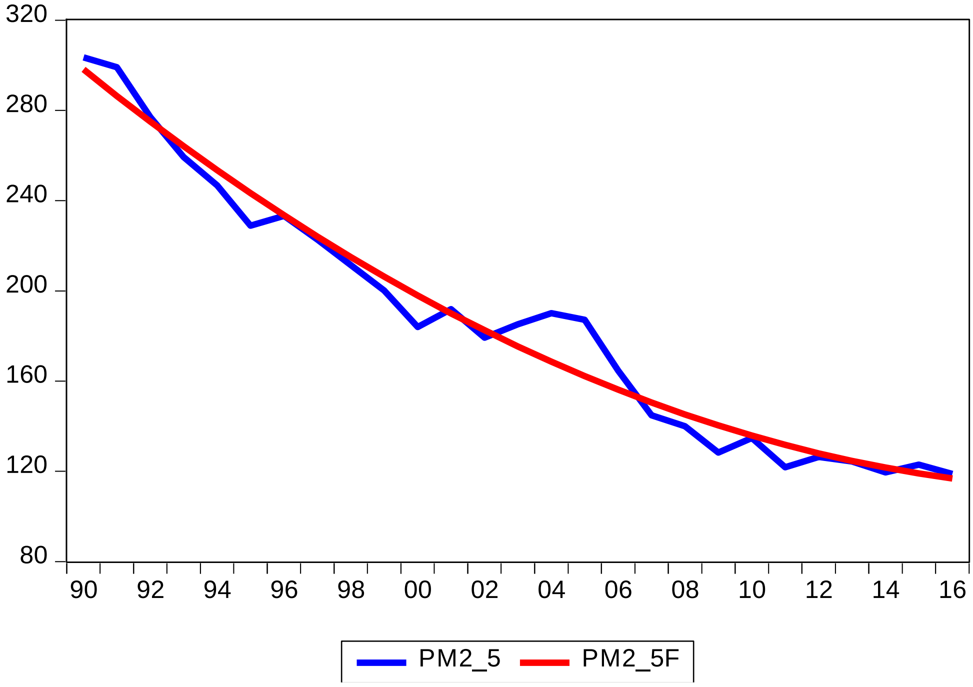


**Figure 11.** PM2.5 emissions TS and the quadratic trend function

## 2.3.2 PM2.5 pollutant emissions over the selected industries

Using the arithmetic mean the eight industries which emitting PM2.5 pollutant the most will be chosen.

**Table 20.** Top 8 industries by the average PM2.5 pollutant emissions

| Industry | Average |
|---|---|
| Consumer expenditure | 51.36740741 |
| **Transport and storage** | **50.23688889** |
| Manufacturing | 31.74403704 |
| **Agriculture, forestry and fishing** | **14.30433333** |
| Electricity, gas, steam and air conditioning supply | 11.43233333 |
| **Construction** | **9.773296296** |

| | |
|---|---|
| Mining and quarrying | 7.171888889 |
| **Wholesale and retail trade; repair of motor vehicles and motorcycles** | **3.543555556** |

Same as previously, the most emitting industry is the "Consumer expenditure" with an average emission of 51.367 thousand tonnes of PM2.5 pollutant per year. In comparison with the CO emissions, "Transport and storage" is the second most polluting industry. Further characteristics will be found below.

**Table 21.** Basic characteristics of the selected industries by PM2.5 emission

| Industry | Mean absolute increase | Mean growth coefficient | Coefficient of variation | Ratio 2016/1990 |
|---|---|---|---|---|
| Consumer expenditure | -0.102769231 | 0.99819459 | 0.108948513 | 0.954103542 |
| **Transport and storage** | **-2.791269231** | **0.935591828** | **0.483571248** | **0.177111562** |
| Manufacturing | -1.017269231 | 0.971811323 | 0.333931658 | 0.475478433 |
| **Agriculture, forestry and fishing** | **-0.905576923** | **0.947643577** | **0.421997906** | **0.247041893** |
| Electricity, gas, steam and air conditioning supply | -1.136884615 | 0.90661134 | 0.809245984 | 0.07815375 |
| **Construction** | **-0.261461538** | **0.970272898** | **0.250590627** | **0.45629049** |
| Mining and quarrying | -0.382576923 | 0.946134999 | 0.413832387 | 0.237017719 |
| **Wholesale and retail trade; repair of motor vehicles and motorcycles** | **-0.125115385** | **0.956761829** | **0.351333767** | **0.316883662** |

It is worth to mention that "Consumer expenditure" industry demonstrated the lowest tempo of the PM2.5 emission decrease, with the only 0.103 thousand tonnes per year decrease at average, in the year 2016 by this industry were emitted 95.4% of the volume of the PM2.5 pollutant of the year 1990. On the contrary, "Electricity, gas, steam and air conditioning supply" industry, with the average decrease of more than 9% per year, in the year 2016 emitted only 7.8% of the volume of the year 1990.

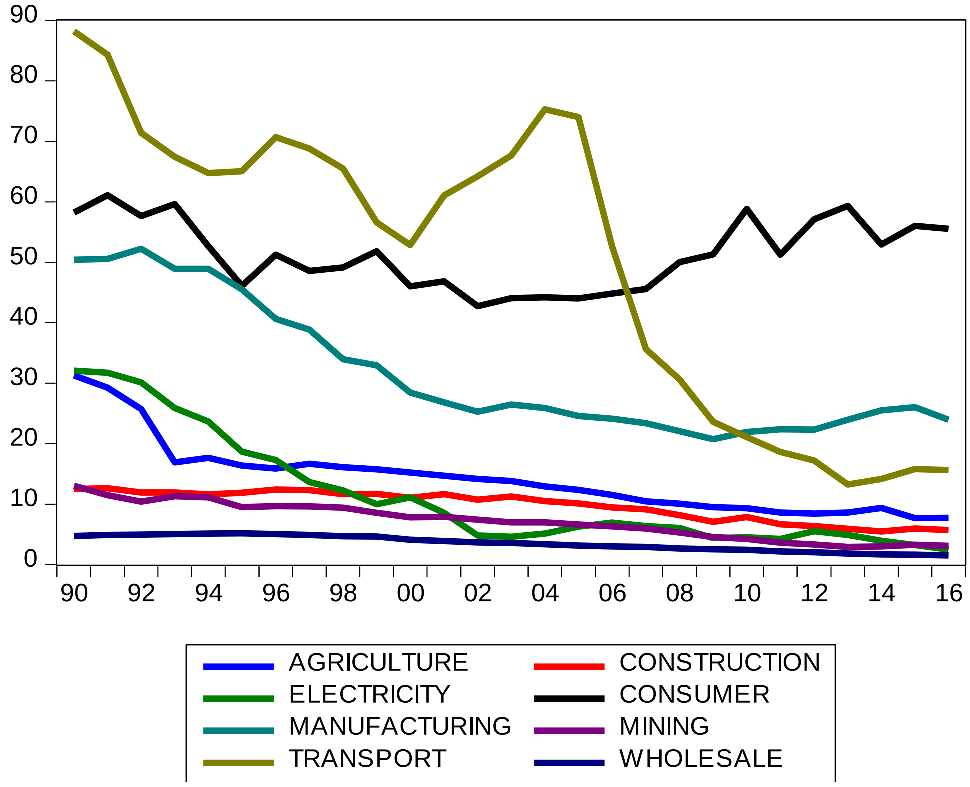


**Figure 12.** Progression of the PM2.5 emission by the selected industries

Figure 12 demonstrates the progression of the times series of the PM2.5 pollutant emission by the selected industries between the years 1990 and 2016. The first thing to notice is a radical increase, then a decrease of the emission from the "Transport and storage" industry. The four-year increase between the years 2000 and 2004, with the total increase of the PM2.5 emissions with an average increase of 5.6155 thousand tonnes per year. Decrease between the years 2005 and 2006, and again between the years 2006 and 2007. For this period, the average decrease in the PM2.5 emissions in this industry was -19.1925 thousand tonnes per year.

In the following table, you will find the estimation of the parameters and the P-Values of the partial t-tests and F-tests of significance for the chosen trend functions. As it was seen in the previous figure, some industries demonstrate different behaviour in terms of the time series development of the PM2.5 emissions, that is why the selection of the trend function differs.

**Table 22.** Estimation of the parameters and P-Values for the t-tests and F-Tests, PM2.5 emissions by the industry

| **Industry** | **Selected trend function** | $b_0$ | $b_1$ | $b_2$ | **Partial t-test of** $b_1$ | **Partial t-test of** $b_2$ | **F-test** |
|---|---|---|---|---|---|---|---|
| Consumer expenditure | Quadratic | 62.96944 | -2.36886 | 0.084008 | 0 | 0 | 0.000002 |
| **Transport and storage** | Linear | **89.35497** | **-2.79415** | | **0** | | **0** |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Manufacturing | Quadratic | 59.88594 | -3.5531 | 0.084162 | 0 | 0 | 0 |
| **Agriculture, forestry and fishing** | Logarithmic | **31.28419** | **-7.10151** | | **0** | | **0** |
| Electricity, gas, steam and air conditioning supply | Quadratic | 36.08137 | -3.11915 | 0.0741 | 0 | 0 | 0 |
| **Construction** | Linear | **13.93725** | **-0.29743** | | **0** | | **0** |
| Mining and quarrying | Linear | 12.44546 | -0.37668 | | 0 | | 0 |
| **Wholesale and retail trade; repair of motor vehicles and motorcycles** | Quadratic | **5.365361** | **-0.08119** | **-0.00267** | **0.0037** | **0.0055** | **0** |

P-Values of each test are close to zero that means all the parameters of the trend functions are significant, and F-Tests of the overall significance are passed, simply speaking, those models make sense. For the "Transport and storage" quadratic model showed a slightly better value of the adjusted Coefficient of Determination, but this model was rejected due to the failure of the t-test of both estimated parameters.

**Table 23.** Criteria statistics of the selected trend functions, PM2.5 emission by the industry

| **Industry** | **$R^2$** | **$R^2$ adj.** | **DW-statistic** | **RMSE** |
|---|---|---|---|---|
| Consumer expenditure | 0.661255 | 0.633027 | 1.493864 | 3.257205 |
| **Transport and storage** | **0.802569** | **0.794672** | **0.36419** | **10.7942** |
| Manufacturing | 0.957852 | 0.95434 | 0.680568 | 2.17625 |
| **Agriculture, forestry and fishing** | **0.943207** | **0.940936** | **1.222132** | **1.438547** |
| Electricity, gas, steam and air conditioning supply | 0.961152 | 0.957915 | 0.625701 | 1.823472 |
| **Construction** | **0.894736** | **0.890525** | **0.457159** | **0.794596** |
| Mining and quarrying | 0.97721 | 0.976298 | 1.325957 | 0.448055 |
| **Wholesale and retail trade; repair of motor vehicles and motorcycles** | **0.96522** | **0.962321** | **0.296422** | **0.23218** |

Selected quadratic trend function covers only 66.12% of the progression of the time series of the PM2.5 emissions of the "Consumer expenditure". In order to verify if the trend function was selected correctly, the smoothing by the moving average method was conducted with the window's length $m = 5$. The result on the figure below demonstrates, that the smoothed function is closer to the quadratic trend function.

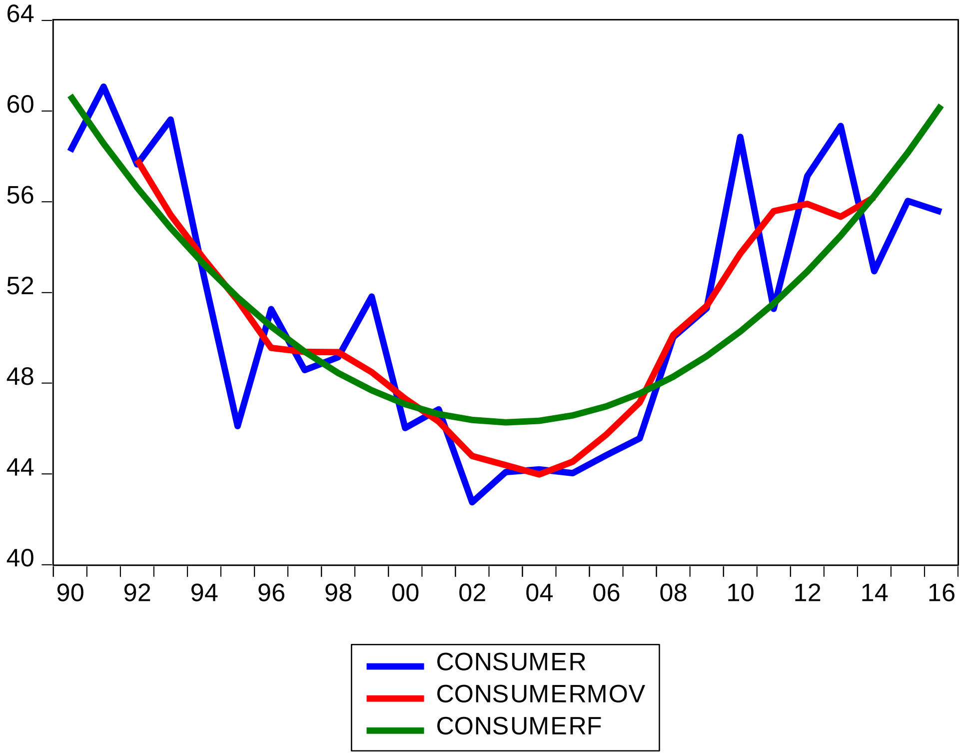


**Figure 13.** Progression of the PM2.5 emissions TS in the "Consumer" industry against the smoothing and the quadratic trend function

## 2.4 PM10 pollutant emissions

### 2.4.1 Overall PM10 pollutant emissions in the United Kingdom

The last pollutant to analyse in this thesis is the PM10 pollutant. As usual, the progression of the overall emissions will be graphically visualised.

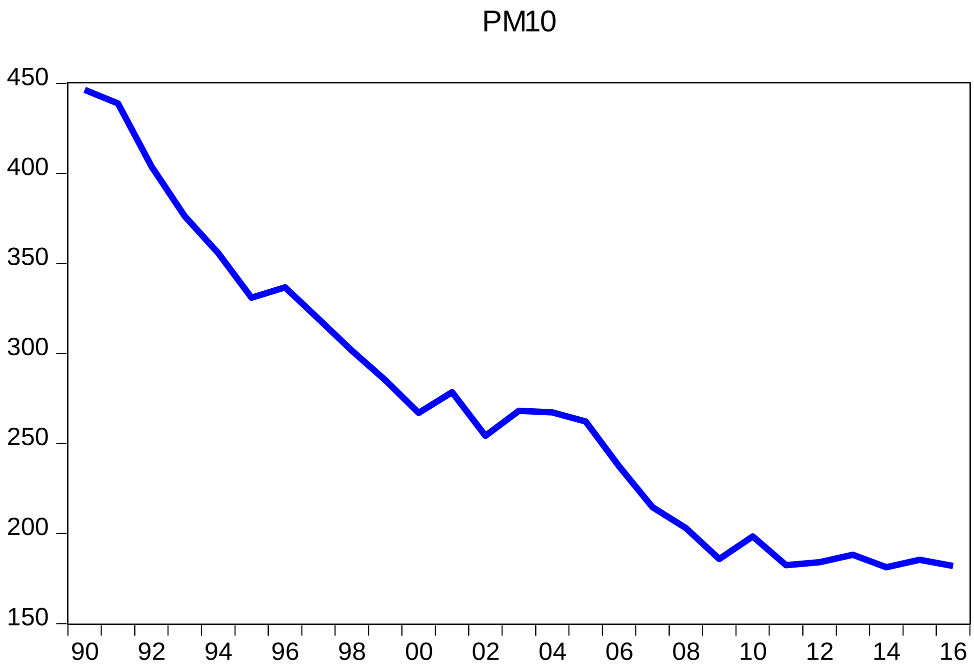


**Figure 14.** Overall PM10 emission in the UK, TS progression

The first thing to be noticed is that PM10 emissions time series development demonstrates very similar behaviour to the PM2.5 emissions.

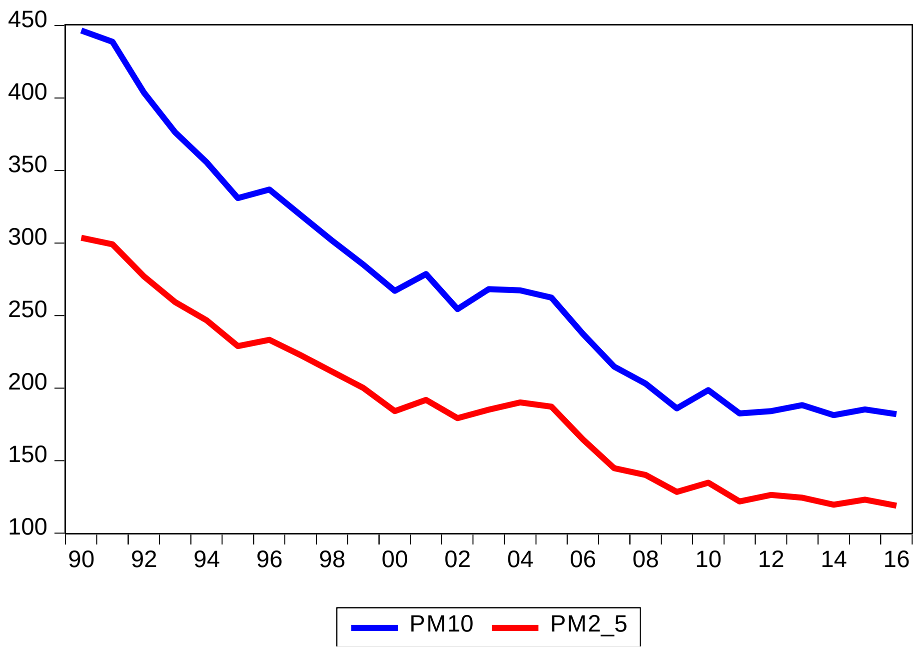


**Figure 15.** Side-by-side comparison of the PM2.5 and PM10 times series

That can be explained by the nature of those pollutants since they are mostly emitted from the same source and have the similar properties but the PM2.5 pollutants are smaller and harder to detect, it is a possibility that the data regarding this pollutant might have some inaccuracies.

The basic properties of the PM10 emission time series will be calculated next, the same way as for the previous series.

**Table 24.** Basic characteristics of the overall PM10 pollutant emissions in the UK

| Arithmetic mean | Mean absolute increase | Mean growth coefficient | Coefficient of variation | Ratio 2016/1990 |
|---|---|---|---|---|
| 271.6636296 | -10.168 | 0.966090676 | 0.295127166 | 0.40781684 |

From those characteristics follows that the average year-to-year decrease of the PM10 pollutant emissions in the period from the year 1990 to the year 2016 is 10.168 thousand tonnes per year with the average emission for the mentioned period 271.664 thousand tonnes per year. The volume of the emissions of PM10 pollutant in the year 2016 is 40% of the volume from the year 1990.

The appropriate trend function should be selected for the overall PM10 emissions. Considering the similarities with the PM2.5 emissions development, we can assume that the most suitable will be the quadratic trend function, but nevertheless, the tests should be conducted as well as comparison with the other trend functions using the stated criteria.

**Table 25.** Selected trend functions and the estimations of their parameters, overall PM10 emissions

| Trend function | $b_0$ | $b_1$ | $b_2$ |
|---|---|---|---|
| Constant | **271.6636** | | |
| **Linear** | 410.0196 | -9.88257 | |
| Quadratic | **456.4388** | **-19.4866** | **0.342999** |
| **Exponential** | 431.1133 | 0.964676 | |
| Hyperbolic | **5.412802** | **1.04138** | |
| **Logarithmic** | 6.346418 | -0.32769 | |

**Table 26.** P-Values of partial t-tests of the parameters F-Tests, PM10 overall emissions

| Trend function | Partial t-test of $b_1$ | Partial t-test of $b_2$ | F-test |
|---|---|---|---|
| Constant | | | |
| **Linear** | 0 | | 0 |
| Quadratic | **0** | **0** | **0** |
| **Exponential** | 0 | | 0 |
| Hyperbolic | **0** | | **0.000026** |
| **Logarithmic** | 0 | | 0 |

**Table 27.** Criteria statistics of the selected trend functions, PM10 overall emissions

| Trend function | $R^2$ | $R^2$ adj. | DW-statistic | RMSE |
|---|---|---|---|---|
| Constant | **0** | **0** | **0.042264** | **80.17532** |
| **Linear** | 0.92174 | 0.91861 | 0.342309 | 22.42901 |
| Quadratic | **0.975407** | **0.973357** | **0.881016** | **12.57335** |
| **Exponential** | 0.957731 | 0.95604 | 0.660698 | 15.65299 |
| Hyperbolic | **0.513672** | **0.494218** | **0.306069** | **61.67726** |
| **Logarithmic** | 0.893284 | 0.889015 | 0.442107 | 31.12643 |

As expected, quadratic function is the most suitable for the modelling of the trend for this time series.

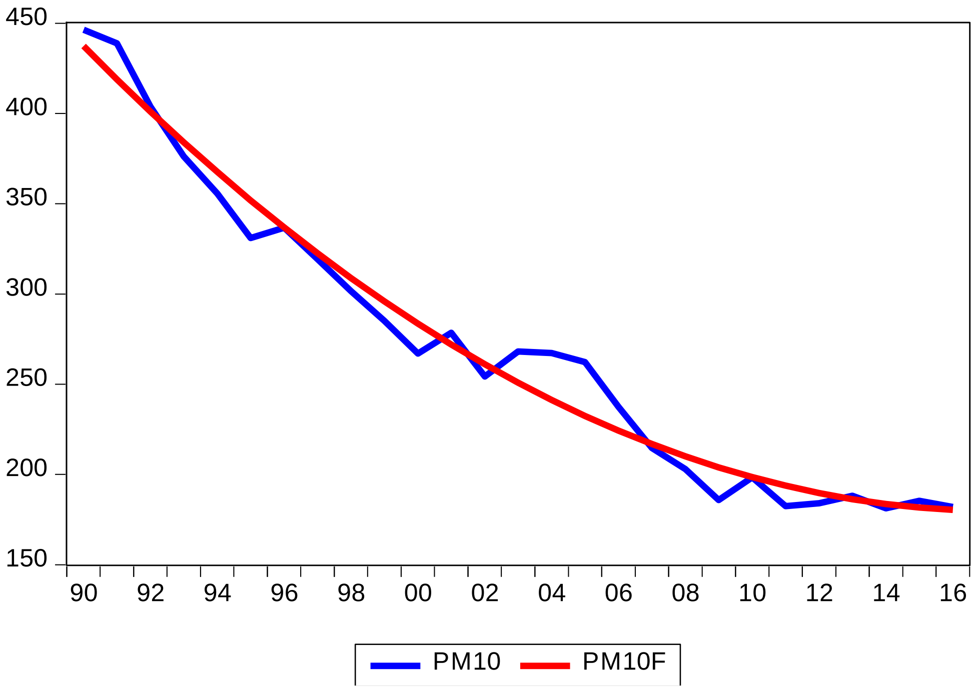


**Figure 16.** PM10 emissions, time series progression vs. quadratic trend function

## 2.4.2 PM10 pollutant emissions over the selected industries

In the last chapter of the practical part of this thesis it will be dedicated to the analysis of the PM10 emissions over the selected industries. Identically as previously, the industries will be selected by the average emission of the pollutant. As was mentioned in the previous section, the PM10 and PM2.5 pollutants share similar properties, so it is expected that the selected eight industries will be the same.

**Table 28.** Top 8 industries by the average PM10 emissions

| Industry | Average |
|---|---|
| Consumer expenditure | 56.44355556 |
| **Transport and storage** | **53.7492963** |
| Manufacturing | 43.73766667 |
| **Construction** | **33.63885185** |
| Agriculture, forestry and fishing | 27.48688889 |
| **Electricity, gas, steam and air conditioning supply** | **22.88411111** |
| Mining and quarrying | 21.08196296 |
| **Wholesale and retail trade** | **4.364111111** |

As you can see, the industries are the same as in the case with PM2.5 pollutants emissions, the only change is the order, "Construction" industry moved from sixth place to the fourth.

The basic characteristics of the selected industries by the PM10 emission are the following:

**Table 29.** Basic characteristics of the selected industries by the PM10 emissions

| Industry | Mean absolute increase | Mean growth coefficient | Coefficient of variation | Ratio 2016/1990 |
|---|---|---|---|---|
| Consumer expenditure | -0.068846154 | 0.998889455 | 0.098060959 | 0.971523116 |
| **Transport and storage** | **-2.923192308** | **0.937539421** | **0.47398882** | **0.18695108** |
| Manufacturing | -1.318538462 | 0.973049725 | 0.301283288 | 0.491485701 |
| **Agriculture, forestry and fishing** | **-0.965** | **0.96980468** | **0.250928664** | **0.450599982** |
| Electricity, gas, steam and air conditioning supply | -2.600153846 | 0.887044346 | 0.238888195 | 0.044317844 |
| **Construction** | **-0.954346154** | **0.975762254** | **0.92741439** | **0.528377556** |
| Mining and quarrying | -0.791961538 | 0.96362631 | 0.28870286 | 0.381614511 |
| **Wholesale and retail trade; repair of motor vehicles and motorcycles** | **-0.138153846** | **0.964318291** | **0.30336929** | **0.388803811** |

"Consumer expenditure" industry is worth to comment since its mean growth coefficient almost equals to 1, which means that the year to year decrease of the PM10 emissions by this industry is insignificant. The Coefficient of variation is 0.098, so values of the PM10 emissions time series vary from the arithmetic mean at an average of 9.8%.

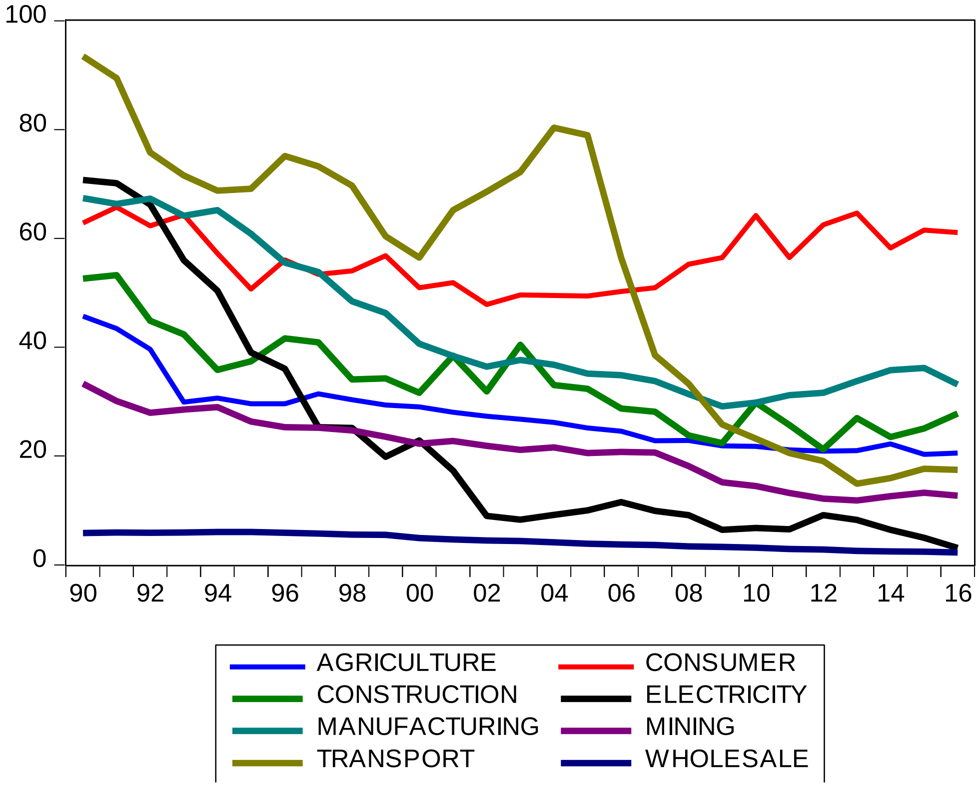


**Figure 17.** PM10 pollutant emission TS by the selected industries

The following table contains the estimations of the parameters of the trend functions. As before, the P-Values of both partial t-tests and F-tests of overall significance presented.

**Table 30.** Selected trend functions, estimation of the parameters, P-Values of the testes, PM10 emission by the industries

| Industry | Selected trend function | $b_0$ | $b_1$ | $b_2$ | Partial t-test of $b_1$ | Partial t-test of $b_2$ | F-test |
|---|---|---|---|---|---|---|---|
| Consumer expenditure | Quadratic | 67.30509 | -2.13487 | 0.082845 | 0 | 0 | 0.000003 |
| **Transport and storage** | Linear | **82.73326** | **-0.45258** | | **0** | | **0** |
| Manufacturing | Quadratic | 77.95529 | -4.20425 | 0.096008 | 0 | 0 | 0 |
| **Construction** | Logarithmic | **55.91864** | **-9.31811** | | **0** | | |
| Agriculture, forestry and fishing | Logarithmic | 45.86081 | -7.68455 | | 0 | | 0 |
| **Electricity, gas, steam and air conditioning supply** | Quadratic | **79.88214** | **-7.26164** | **0.174019** | **0** | **0** | **0** |
| Mining and quarrying | Linear | 31.77898 | -0.76407 | | 0 | | 0 |
| **Wholesale and retail trade** | Quadratic | **6.461937** | **-0.1173** | **-0.00178** | **0.0001** | **0.0491** | **0** |

Pay your attention to the P-Value of the $b_2$ estimation of the quadratic trend function for the "Wholesale and retail trade" industry. With the previously stated Level of Significance

$\alpha = 5\%$, this trend function can be used for our purposes, but if the Level of Significance were lower e.g. 1%, another trend function should be selected.

**Table 31.** Criteria statistics of the selected trend functions, PM10 emission by the industry

| Industry | $R^2$ | $R^2$ adj. | DW-statistic | RMSE |
|---|---|---|---|---|
| Consumer expenditure | 0.657632 | 0.629102 | 1.562808 | 3.238596 |
| **Transport and storage** | **0.798732** | **0.790681** | **0.361947** | **11.42954** |
| Manufacturing | 0.958643 | 0.955196 | 0.603093 | 2.679832 |
| **Construction** | **0.830489** | **0.823708** | **1.75537** | **3.475286** |
| Agriculture, forestry and fishing | 0.933379 | 0.930714 | 1.164258 | 1.694833 |
| **Electricity, gas, steam and air conditioning supply** | **0.965923** | **0.963083** | **0.750279** | **3.917764** |
| Mining and quarrying | 0.956082 | 0.954325 | 0.749729 | 1.275508 |
| **Wholesale and retail trade** | **0.970567** | **0.968114** | **0.310303** | **0.227135** |

Once again, the trend function of the "Consumer expenditure" demonstrated the worst value of the Coefficient of Determination. The moving average method can be applied for a better understanding of the trend.

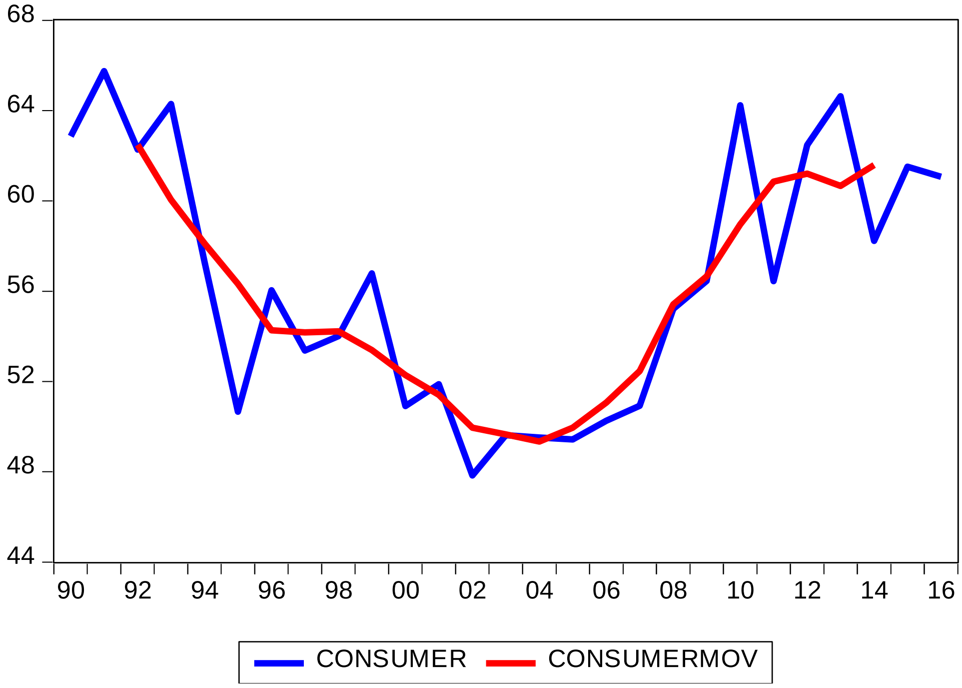


**Figure 18.** Consumer TS and smoothed consumer TS, PM10 emissions

Exactly as before, the smoothed time series demonstrates the outlines of the quadratic trend.

# Conclusion

In this bachelor thesis, by conducting the time series analysis of the NMVOC, CO, PM2.5 and PM10 pollutants emission in the UK overall and by this country's industry sectors, a few interesting discoveries have been made. This chapter is dedicated to highlighting the most important of them.

The average emissions of each pollutant in the UK in the given period 1990-2016 were as follows: CO – 3985.266 tonnes per year, NMVOC – 1665.813 tonnes per year, PM10 – 271.664 tonnes per year, PM2.5 – 186.893 tonnes per year. Each of the overall emission analysis in the UK supports the hypothesis of the long-term decreasing trend of emissions of the studied pollutants in the UK.

The most significant decrease in the analysed period demonstrated the emission of CO pollutant, as CO emissions in the year 2016 had only 21.3% of the emissions volume of the year 1990. Next is the NMVOC emission with an average year-to-year decrease of 4.4%. Emissions of the PM10 and PM2.5 pollutants both demonstrated the same decrease with an approximate rate of 3.5% per year.

Across all the industries the most pollutant one is the "Consumer expenditure", this industry caused the highest average per year emission of the all four studied pollutants. It also has the worst value of the relative decrease of PM2.5 and PM10 pollutants emission. On the other hand, this industry is the second best across the 8 most CO emitting industries in terms of CO emissions decrease with the average 6.7% decline per year.

As was mentioned in the introduction, the United Kingdom has the obligations on the emission reduction, ruled by the environmental frameworks. According to 2018 European Environmental Agency report on the National Emission Ceilings Directive, since 2011 the UK attained all the pollution emission ceiling and reduction commitments [8].

# Bibliography


1. KOUIMTZIS, T., SAMARA, C. (1995). Airborne Particulate Matter. Springer Berlin Heidelberg. ISBN 978-3-540-49145-3

2. Department for Environment Food & Rural Affairs (2017). *Air Pollution in the UK 2017*. [online]. Available at https://uk-air.defra.gov.uk/assets/documents/annualreport/air_pollution_uk_2017_issue_1.pdf (cit. 02 May 2019)

3. Office for National Statistics (2017). *Emissions from other pollutants in the United Kingdom, 1990 to 2016 and provisional 2017*. [online]. Available at https://www.ons.gov.uk/file?uri=/economy/environmentalaccounts/datasets/ukenvironmentalaccountsatmosphericemissionsemissionsofotherpollutantsbyeconomicsectorandgasunitedkingdom/provisional/atmosphericemissionsotherpollutantsprov2017.xls (cit. 05 May 2019)

4. Raaschou-Nielsen, O., Andersen Z. J. (2013). *Air pollution and lung cancer incidence in 17 European cohorts: prospective analyses from the European Study of Cohorts for Air Pollution Effects (ESCAPE)*. The Lancet. [online]. Available at https://www.thelancet.com/journals/lanonc/article/PIIS1470-2045(13)70279-1/ (cit. 06 May 2019)

5. Omaye, S.T. (2002). *Metabolic modulation of carbon monoxide toxicity*. ScienceDirect. [online]. Available at https://www.sciencedirect.com/science/article/pii/S0300483X02003876?via%3Dihub (cit. 06 May 2019)

6. Driving Electric (2019). *Electric car sales UK: electric car sales up 81.1%*. [online]. Available at https://www.drivingelectric.com/news/678/electric-car-sales-uk-2019-stars-110-jump (cit. 07 June 2019)

7. European Environmental Agency (2010). *Non-methane volatile organic compounds (NMVOC) emissions*. [online]. Available at https://www.eea.europa.eu/data-and-maps/indicators/eea-32-non-methane-volatile-1 (cit. 10 May 2019)

8. European Environmental Agency (2018). *NEC Directive reporting status 2018*. [online]. Available at https://www.eea.europa.eu/publications/nec-directive-reporting-status-2018#tab-related-publications (cit. 12 May 2019)

9. ŘEZANKOVÁ, H., MAREK, L., VRABEC M. IASTAT – Interaktivní učebnice statistiky. [online]. Available at https://iastat.vse.cz (cit. 12 May 2019)

10. ŘEZANKOVÁ, H., LÖSTER, T. Úvod do statistiky. Praha: Oeconomica, 2009. ISBN 978-80-245-1514-4.

11. HINDLS, R. a kol. Statistika pro ekonomy. Osmé vydání. Praha: Professional publishing, 2007. ISBN 978-80-86946-43-6.

12. ARLT, J., ARLTOVÁ, M. (2007). Ekonomické časové řady. Praha: Grada Publishing, 2007. ISBN 978-80-247-1319-9

13. EVERITT, B. (1998). The Cambridge Dictionary of Statistics. Cambridge, UK New York: Cambridge University Press. ISBN 978-0521593465

14. ARLT, J., ARLTOVÁ M., RUBLÍKOVÁ, E. Analýza ekonomických časových řad s příklady. 1.vyd.Praha, 148 s. 2002. Vysoká škola ekonomická. ISBN 80-245-0307-7