ZKÜ Sosyal Bilimler Dergisi, Cilt 4, Sayı 8, 2008, ss. 73–88

PREDICTING THE TREND AND SEASONAL FACTORS OF TOTAL SUSPENDED PARTICLES (TSP) LEVELS IN TRABZON: SEASONAL AND WEIGHTED LEAST SQUARES REGRESSION

Yrd.Doç.Dr. Ali Sait ALBAYRAK
Zonguldak Karaelmas Üniversitesi
İİBF İşletme Bölümü
asalbayrak@karaelmas.edu.tr

ABSTRACT

In winter season the air pollution is the one of the most important environmental problems in Trabzon, located in Eastern Black Sea Region of Turkey. The humid climate as well as the densely populated urbanization cause serious problems for a long time period. One of the main targets of air quality management is to estimate the influence of different factors such as trend and seasonality upon air quality levels in a given area. In this paper, an attempt has been made to identify and estimate the trend and seasonal factors in TSP in Trabzon using accurate and reliable techniques.

The data used in this study is concerned with monthly measurement levels of TSP taken from January 1996 through May 2007. The most accurate WLS regression results show that the adjusted–R^2 is about 83,8% and the monthly averages of TSP do not have a clear trend over the period 1996–2007, the trend estimate is only about 0,08 points per year, which has a statistical significant of 0,681.

Keywords: *Suspended Particles, Air Quality, Weighted Regression, Decomposition Method.*

TRABZON'DA TOPLAM ASILI PARTİKÜLLER DÜZEYLERİNDEKİ TRENDİN VE MEVSİMLİK FAKTÖRLERİN TAHMİN EDİLMESİ: MEVSİMSEL VE AĞIRLIKLI REGRESYON ANALİZİ

ÖZET

Türkiye'nin Doğu Karadeniz Bölgesinde yer alan Trabzon'da kış mevsiminde hava kirliliği en önemli çevre sorunlarından birisidir. Uzun zamandan beri rutubetli iklimin yanında yoğun kentleşme çok önemli çevre sorunlarına neden olmaktadır. Hava kalite yönetiminin temel hedeflerinden birisi de herhangi bir bölgedeki hava kalitesi üzerinde etkili olan trend veya mevsimsellik gibi faktörleri tahmin etmektir. Bu çalışmada Trabzon'da toplam asılı partiküller maddede düzeylerindeki trend ve mevsimlik faktörler en uygun ve güvenilir tekniklerle araştırılmaktadır.

Araştırmada Ocak 1996 ile Haziran 2007 tarihleri arasındaki aylık asılı partiküller serisi kullanılmaktadır. Ağırlıklı regresyon analizi ile elde edilen en uygun sonuçlarda düzeltilmiş-R^2 değerinin %83,8 ve toplam asılı partiküller düzeylerinin Ocak 1996-Haziran 2007 döneminde yıllık trendinin pozitif ve yaklaşık olara 0,08 um/m^3 olduğu ve bunun 0,681 anlamlılık düzeyine sahip olduğu anlaşılmaktadır.

Anahtar Kelimeler: *Asılı Partiküller, Hava Kalitesi, Ağırlıklı Regresyon, Bileşenlere Ayırma.*

1. INTRODUCTION

Air pollution or insufficient air quality is one of the most common urban problems in the world. Epidemiological studies show clearly that both indoor and outdoor air pollution affect human health negatively. In particular, air pollution negatively affects vulnerable groups such as the sick, old and children. Studies of air pollution on health have linked particulate matter with a number of significant health effects (e.g. mortality, morbidity, respiratory and cardiovascular problems, etc.). These include increased mortality and aggravation of existing respiratory and disease of the heart and blood vessels disease, as evidenced by increased hospitalization, school absences and lost work days (Sapan, 2006).

Urban air quality in developing countries has become serious gradually because of population growth, rapid urbanization, industrialization and domestic heating. The air pollution path in the urban atmosphere consists of emissions and transmission of air pollutants resulting in the ambient concentrations. Each part of the path is influenced by different factors. Emissions from domestic heating are very important source group in Turkey, as well as other cold countries of the world (Turanlıoğlu et al., 2005). During transmission, air pollutants are dispersed, diluted and subjected to photochemical reactions (Mayer, 1999).

Most cities worldwide have witnessed serious air quality problems mainly due to industries and vehicles. Urbanization has resulted in high levels of ground level deterioration of air quality. The investigation of air pollution in mega cities showed that the major problem affecting these cities is their high levels of total suspended particles (Mage, et al., 1996). It is well established that high levels of *TSP* are significantly associated with adverse health effects, ecosystem damage and degraded visibility (Goswami et al., 2002).

Total suspended particles (*TSP*) are the collective term used for a mixture of solid particles and liquid droplets found in the air. *TSP* refers to all particles in the atmosphere. *TSP* was the first indicator used to represent suspended particles in the ambient air. *TSP* has wide range of sizes and originates from many different stationary and mobile sources (Aneja, et al., 2001). One of the major characteristics of particulate matter is particle size. Particles can be categorized as *TSP*, PM10, PM2.5, particles less than 0.1 um, condensable particulate matter. Particles ranging in size from 0.1 micrometer to about 30 micrometer in diameter are referred to as *TSP*. Particles less than 2,5 microns in diameter are known as "fine" particles; those larger than 2,5 microns are known as "coarse" particles. Fine particles with diameters of less than 1 um, move like gases. Because of their low settling velocities, fine particles may be transported 1,000 kilometers or more from their source. Under the influence of gravity, larger particles do not remain suspended and tend to settle out of the air, sometimes creating localized areas of high particle disposition.

Twinning Projects of the European Commission were introduced as a tool to achieve the same standards and backgrounds in all countries which are interested in a closer cooperation with the European Union (Müller, 2006). The “Air Quality” Twinning Project of Turkey started in October 2004 and has to fulfill four main tasks: (1) Transposition of the Air Quality Framework Directive 96/62/EC and the Large Combustion Plants Directive 2001/80/EC into Turkish (Draft) Regulation, (2)

Draft Agreed Framework Regulation on Air Quality which defines the roles and the responsibilities of the involved ministries (considering both directives), (3) stringent of the qualification of the administration stringent of the quality management and preparation of the accreditation of the two laboratories-Refik Saydam Hygienic Center (RSHC) and Gölbaşı, (4) agreed strategic Action Plans on further implementation steps of the two directives (Gömer et al., 2006).

A large number of epidemiological studies have established the link between ambient particle concentrations and daily excesses in mortality and morbidity (Pope III., et al., 1995). Although there is some evidence that certain particle properties, such as chemical composition and size, have different importance on human health (Harrison and Yin, 2000), current EU legislation only controls the mass concentration of particles with diameter below 10 um in ambient air (European Council, 1999). This is implemented by imposing two health-based limit values: (1) A 24h mean concentration of 50 um/m^3 not to be exceeded more than 35 times during a calendar year and (2) an annual concentration of 40 um/m^3. More stringed PM10 objectives will have to be achieved by EU countries by 2010: (1) A 24h mean concentration 50 um/m^3 not to be exceeded more than 7 times during a calendar year (2) an annual mean concentration of 20 um/m^3. Furthermore, a concentration cap (25 um/m^3) and an exposure reduction target have been recently proposed for mass concentration of particles with diameters below 2.5 um (PM2.5) in ambient air (Vardoulakis and Kassomenos, 2008).

Hess et al. (2001) presented an overview of statistical approaches available for detecting and estimating linear trend in environmental data. They had evaluated seven methods of trend detection and made recommendations based on a simulation study. They showed t-test adjusted for seasonality and Seasonal Kendall test appear to maintain their stated a levels (false rejection level) as well as maintain high power with different trend functions. Gupta and Kumar (2006) present a set of time series analysis methods t-test adjusted for seasonality, Seasonal Kendall test and Intervention analysis have been applied to identify and estimate the trend in PM10 and *TSP* levels monitored for about 10 years at three monitoring sites at each of the four cities in India (Gupta and Kumar, 2006). Jorquera et al. (2000) used intervention analysis methodology to detect the trend of PM10 and PM2.5.

Seasonal decomposition method, seasonal least squares regression and seasonal weighted least squares regression has been used to detect trends in the monthly average concentration of *TSP* in this paper.

2. ESTIMATION TECHNIQUES

Classical time series decomposition separates a time series into four components: long-range trend, seasonality, cycle, and randomness. The multiplicative decomposition model is (Trend) x (Seasonality) x (Cycle) x (Random) $= X_t = T_t \times S_t \times C_t \times R_t$.

Where X_t denotes the series or, optionally, log of series; T_t denotes the linear trend; C_t denotes cycle; S_t denotes season; R_t denotes random error and t denotes the time period (Makridakis and Wheelwright, 1978).

Note that this model is multiplicative rather than additive. Although additive models are more popular in other areas of statistics, forecasters have found that the multiplicative model fits a wider range of forecasting situations (Hintze and NCSS, 2005). Multiplicative seasonal component is a factor by which the seasonally adjusted series is multiplied to yield the original series. Observations without seasonal variation have a seasonal component of 100. Additive seasonal adjustments are added to the seasonally adjusted series to obtain the observed values. Observations without seasonal variation have a seasonal component of 0.

Decomposition is popular among forecasters because it is easy to understand. While complex ARIMA models are often popular among statisticians, they are not as well accepted among forecasting practitioners (Hintze and NCSS, 2005). For seasonal data, decomposition methods are often as accurate as the ARIMA methods and they provide additional information about the trend and cycle which may not be available in ARIMA methods (Hintze and NCSS, 2005).

Decomposition method has one disadvantage: the cycle component must be input by the forecaster since it is not estimated by the algorithm. You can get around this by ignoring the cycle, or by assuming a constant value (Hintze and NCSS, 2005).

Multiple regression analysis refers to a set of techniques for studying the linear relationship among two or more variables. This relationship for population is described in the following formula:

$$y_i = \beta_0 + \beta_1 x_{1j} + ... + \beta_p x_{1p} + \varepsilon_j$$

Where Y is the value of the dependent scale variable; p is the number of predictors; X is the value of the independent variable; the subscript j represents the observation number. The β's are the unknown regression parameters. Their estimates are represented by b's each represents the original unknown (population) parameter, while b is an estimate of this. The ε is the error of observation j.

Multiple regression analysis studies the relationship between a dependent scale variable and p independent variables. The sample multiple regression equation is

$$\hat{y}_i = b_0 + b_1 x_{1j} + ... + b_p x_{1p}$$

The intercept, b_0, is the point at which the regression plane intersects the Y axis. The b_i *is* the slopes of the regression plane in the direction of x_i. These coefficients are called the partial-regression coefficients. Each partial regression coefficient represents the net effect the ith variable has on the dependent variable, holding the remaining X's in the equation constant. A large part of a regression analysis consists of analyzing the sample residuals, e_j, defined as

$$e_j = y_j - \hat{y}_j$$

Once the β's have been estimated, various indices are studied to determine the reliability of these estimates. One of the most popular of these reliability indices is the correlation coefficient. The correlation coefficient is an index that ranges from -1 to 1. When the value is near zero, there is no linear relationship. As the correlation gets closer to plus or minus one, the relationship is stronger.

For the purpose of testing hypotheses about the values of model parameters, the linear regression model also assumes: (1) the error term has a normal distribution with a mean of 0; (2) the variance of the error term is constant across cases and independent of the variables in the model. An error term with non-constant variance is said to be heteroscedastic; (3) the value of the error term for a given case is independent of the values of the variables in the model and of the values of the error term for other cases (Orhunbilge, 1998).

If the data are a random sample from a larger population and the *ε's* are independent and normally distributed, a set of statistical tests may be applied to the *b's* and the correlation coefficient. These t-tests and F-tests are valid only if the above assumptions are met (Orhunbilge, 1998).

When the parameters of a linear regression model are estimated, all observations usually contribute equally to the computations. This called ordinary least squares (OLS) regression. When all the observations have the same variance, this is the best strategy since it results in parameters estimates that have the smallest possible variance (Norusis and SPSS Inc, 1999). However, if the observations are not measured with equal precision, OLS no longer yields parameter estimates with the smallest variance. It is well known that the method of Ordinary Least Squares (OLS) is the most efficient for regression problems under the normal constant variance error. But any violations of the assumptions (non-normality, non-constant variance) may cause OLS estimates be less satisfactory and it is worthwhile to consider alternatives. A modification known as weighted least-squares regression (WLS) analysis may be used as an appropriate method in the presence of non-constant variance. In WLS regression, observations are weighted by the reciprocal of their variances. This means that observations with large variances have less impact on the analysis than observations associated with small variances (Norusis and SPSS Inc, 1999).

In the presence of heteroscedasticity, the OLS regression method leads to unbiased and consistent estimates but leads to ineffective parameters estimates and inconsistent covariance matrix. That is, the variance of estimated parameter is not a minimum. As a result, statistical tests of the significance may lead to incorrect conclusion. Heteroscedasticity usually does not occur in time series data when both dependent and independent variables tend to change in the same magnitude. For instance, income and consumption both change about the same rate. But heteroscedasticity can occur more often in the cross section data (Shin, 1996).

Graphic method, Goldfield-Quandt, Breusch-Pagan-Godfrey, Park, Glejser, Spearman Rank Correlation and White NR^2 are some tests of detecting heteroscedasticity (Gujarati, 1995).

3. RESULTS AND DISCUSSIONS

The series used in this study are concerned with include monthly measurements of *TSP* taken from January 1996 through June 2007, although several observations are missing. The principle aim of this paper is to search whether the *TSP* levels of Trabzon shows a trend or not. To do this, it is built a regression model, expressing the *TSP* as a linear combination of other variables, including time and dummy month variables. If the model is satisfactory, the coefficient of time indicates a trend.

Table 1: Descriptive Statistics for Monthly Suspended Particles Data

							95% CI for Mean			
Month	n	Mean	Std.Dev.	St. Error	Skew.	Kurt.	U. Bound	L. Bound	Min.	Max.
JAN	12	90.75	25.09	7.24	0.45	-0.52	74.81	106.69	55.0	139.0
FEB	12	82.00	20.99	6.06	0.65	0.07	68.66	95.34	53.0	125.0
MAR	12	67.58	11.39	3.29	0.66	-1.54	60.35	74.82	56.0	85.0
APR	12	45.83	7.53	2.17	-0.20	-0.76	41.05	50.62	34.0	58.0
MAY	12	33.08	7.49	2.16	-0.05	-1.40	28.33	37.84	22.0	43.0
JUN	11	21.73	5.52	1.66	-0.39	0.53	18.02	25.43	11.0	31.0
JLY	11	19.46	5.28	1.59	1.01	0.91	15.91	23.00	13.0	31.0
AGT	11	19.82	3.46	1.04	-0.50	-1.36	17.49	22.14	14.0	24.0
SEP	11	24.36	4.74	1.43	1.41	1.60	21.18	27.55	20.0	35.0
OCT	11	31.68	7.99	2.41	0.58	-0.35	26.31	37.05	21.0	46.5
NOV	11	72.39	18.18	5.48	0.71	-0.73	60.18	84.61	51.3	105.0
DEC	11	98.79	31.19	9.41	0.01	-1.36	77.83	119.74	57.0	142.0
Total	137	51.11	32.10	2.74	0.92	0.09	45.68	56.53	11.0	142.0

Table 1 shows the descriptive statistics for *TSP* series. The *TSP* levels have an average of 51.11 um/m^3. Winter months have an average level above the series mean, while summer months have an average level below the series mean. November, December, January, February and March have the highest mean of *TSP* and standard deviation levels. The 95% confidence intervals for January and December are nearly 74.8-106.7 um/m^3 and 77.8-119.7 um/m^3 respectively.

Figure 1: Total Suspended Particles by Month

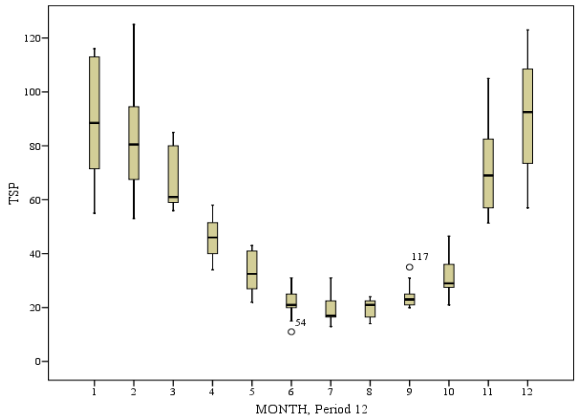
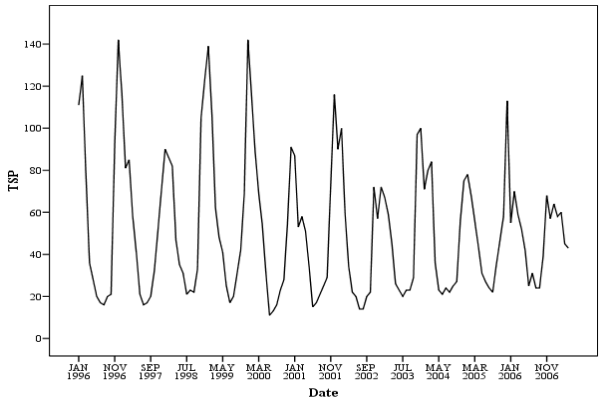


Figure 2 shows the *TSP* level by months. This display is some time called a box-and-whisker plot. For each group of month, the horizontal line in the middle of box marks the median of the sample. Thus, for example, the median for January is around 88.5 um/m^3, while in August roughly 21 um/m^3. The edges of each box, called hinges, mark the 25th and 75th percentiles. The length of the box is called the hspread and corresponds to the interquartile range. The hspread for, April, May, June, July, August, September and October is very short, while that for the January, February, March, November and December is considerably longer. The whisker (vertical lines extending up and down from each box) show the range of values that

fall within 1.5 hspread of the hinges. It is easy to see from Figure 1 that the medians and the spread of 12 boxes vary greatly. The median *TSP* levels for March fall toward the top of its box, indicating that the distribution is left-skewed, while that for the January, February, March, October, November and December is more centered as occurs for symmetric distributions.

In addition to providing a sufficient summary of where the most of the values are concentrated and the shape of each distribution, the box plot is constructed to indicate outliers. Cases that have values more than 3 hspread below the lower hinge or above the upper hinge are marked by an asterisk (*) and called extreme values, while cases that have values between 1.5 and 3 hspread outsides the hinges are marked by an open circle (o) and called outliers. In summary, 54th month value in June and 117 month value in September is an outlier.

Figure 2: Monthly Total Suspended Particles Data



It is always a good idea to have a feel for the nature of data before building a model. Does the data exhibit seasonal variations or trend? Figure 2 shows that the TSP series exhibits numerous peaks, many of which appear to be equally spaced, as well no clear upward or downward trend. The equally spaced peaks suggest the presence of a periodic component to the time series. There are also peaks that do not appear to be part of the seasonal pattern and which represent significant deviations from the neighboring data points. These points may be outliers, which should be also taken into consideration.

Replacing Missing Data of Suspended Particles Monthly Series

Firstly I dealt with missing data. There are two considerations: In time series analysis, you cannot have any missing time periods, since observations must be evenly spaced. In a monthly series like this one, it must have been an observation every month even if the observation contains missing data.

Once the series has a complete set of time periods, the next step is to decide how to deal with any missing values within the series. Some time series methods cannot process a series that contains missing data. Seasonal decomposition method, which is used in this study, is one of them. Before using one of the methods that require valid data, it must have been filled with reasonable values in place of the

missing data, either manually or with the Replace Missing Value procedure on the Transform menu in SPSS. Since *TSP* data are seasonal, a value midway between the preceding and following months is likely to be a better prediction, so linear interpolation is used instead of series mean. In this study a plot of series shows the seasonal variation in the *TSP* (Figure 2).

Calculating a Trend Variable for *TSP* Monthly Series

As mentioned before the aim of this paper is to determine if there is a trend in the *TSP* monthly series. Since the trend per month would be quite small, it is preferable to see the trend per year. To express the trend in part per year, it is needed a variable to indicate how many years each observation is from the beginning of the study. There are several ways to compute such a variable; perhaps the simplest is to use sequential number of each monthly observation in the data file. Then dividing it 12 gives the number of years since the first observation.

Removing Seasonality for Predicting the Real Trend

In order to uncover any real trend in the *TSP* monthly series, firstly it is needed to account for the variation in the *TSP* measures that are due to seasonal effects. For instance, if *TSP* levels are always higher in winter than in the summer, this would confound the estimate of trend. The seasonal decomposition method decomposes a series into a seasonal, trend, cycle and error component (Orhunbilge, 1998). The Seasonal Decomposition Method normally treats the series as the product of the seasonal, trend, and cycle components. This multiplicative method is appropriate when seasonal variation is greater at higher levels of the series. For series, where seasonality does not increase with the level of the series, an alternative additive model is available. This produces the following results shown in Table 2.

Table 2: Seasonal Factors of Total Suspended Particles (*TSP*)

Seasonal	Month											
Factor	JAN	FEB	MAR	APR	MAY	JUN	JLY	AGT	SEP	OCT	NOV	DEC
Additive	40.3	28.9	16.1	-4.4	-18.7	-29.0	-31.7	-30.9	-26.0	-18.6	22.0	52.0
Multiplicative	180.7	155.6	132.1	92.4	63.2	44.4	37.0	39.5	48.2	63.3	142.4	201.2

The seasonal index shown in Table 2 is the average deviation of each month's *TSP* level from the level that was due to the other component that month. In January averaged about 180.7 um/m^3 above the deseasonalized *TSP* level. As it seen from the Table 2, December, January and February have the highest *TSP* levels, while July, June, August and September has the lowest *TSP* levels respectively.

If multiplicative model with Seasonal Decomposition Method is used, the seasonal index would be expressed as a percentage. Indexes for high *TSP* months such as November, December, January, February and March would be above 100, while indexes for low *TSP* months such as April, May, June, July, August, September and October would be below 100. Multiplicative and additive seasonal indexes are not directly converted each other, since the type of model used determines how the observation for each month are averaged. The seasonally adjusted or deseasonalized series for multiplicative model can be adequately used to predict whether there is a significant trend in *TSP* level.

Table 3: Regression Analysis Results for Seasonal Adjusted *TSP* Levels

	Unstandardized Coefficients			
	B	Std. Error	t	Sig.
Constant	58.330	2.414	24.157	.000
Trend	-1.335	0.364	-3.663	.000
R=%30.1% mate=14.053	Adjusted R-Square=8.4% DW=1.97	F=13.42 (0.000)	Std. Error of the Esti- DW=1.97	

Table 3 shows regression results for the deseasonalized *TSP* level. The coefficient of *TREND*, about -1.335 um/m^3 represents the annual trend. Deseasonalized *TSP* level declined slowly approximately over this 11-years period. This effect is statistically significant at the 0.000 level. The model does not explain much of the variation. The R^2, adjusted for the number of cases and variables, is only about 8.4%. The standard error of the estimate is around 14.053 um/m^3.

Predicting Trend and Seasonality Simultaneously: Dummy Variable Regression

Seasonally adjusting a series prior to evaluating the model, as done above, was once almost the only practical way of analyzing seasonal data. One way to include effects in a regression model without seasonally adjusting the data is to use dummy variables for the seasons. In this study, we used 11 dummy variables for 11 of the 12 month. The 4^{th} month (April) is reserved as a baseline for comparison. If it used all 12 months, the 12^{th} one would add no information that it couldn't figure out from the first 11 dummy variable months.

The result of dummy variable regression analysis shown in Table 4 shows that the R^2 is much higher than in Table 3. Over 79.3% of the variation in *TSP* level is explained by this model, even after adjusting for the number of variables and cases. This improvement is largely due to the fact that the seasonal variation is included in the model and explained by the seasonal dummy variables, rather than being removed prior to the analysis.

The standard error of estimate in Table 3 (14.053 um/m^3) is slightly less than that in Table 4 (14.598 um/m^3). It was easier to fit the model for the seasonally adjusted *TSP* levels. The dummy-variable regression actually much the same thing as Additive Seasonal Decomposition Method but gave up degrees of freedom in doing so, which led to larger standard errors of estimates.

The *CONSTANT* term is decreased from 58.33 to 53.60 um/m^3. The coefficient of the *TREND* variable has increased slightly to -1.33 um/m^3 and has a statistical significant of 0.001.

Table 4: Regression Analysis Results of Dummy Month Variables

	Constant	TREND	JAN	FEB	MAR	MAY	JUN	JLY	AGT	SEP	OCT	NOV	DEC
B	53.60	-1.33	44.6	36.0	21.6	-12.6	-24.6	-26.7	-26.2	-21.6	-14.2	26.7	53.2
S.E.	4.76	0.38	5.96	5.96	5.96	5.96	6.09	6.09	6.09	6.09	6.09	6.09	6.09
Sig.	.000	.001	.000	.000	.000	.036	.000	.000	.000	.001	.022	.000	.000
R=90.1%	Adjusted R Square=79.3% DW=2.19				F=44.47 (0.000)				Std. Error of the Estimate=14.598				

Each of the dummy month variable show the seasonal effect of that month compared to April, the omitted month. Since the April seasonal effect was quite

small (-4.4 um/m^3) in Table 2 the coefficient of these dummy variables are very close to the effects estimated by Seasonal Decomposition Method.

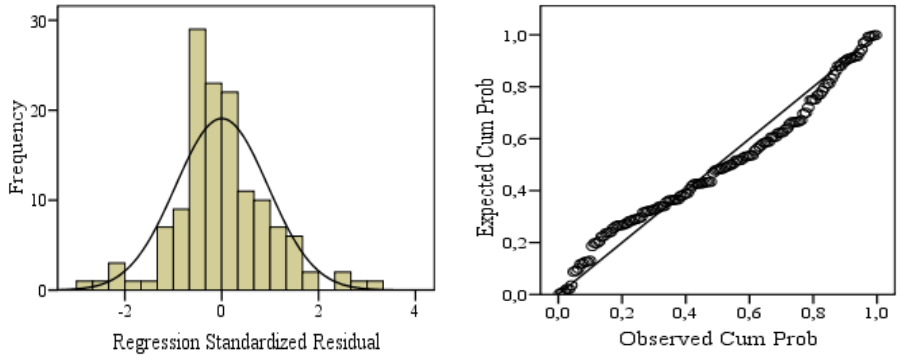
Table 5: Residual Analysis

Case Number	Std. Residual	TSP Level	Predicted Value	Residual
12	2.505	142.0	105.433	36.5673
37	3.078	139.0	94.072	44.9276
48	2.778	142.0	101.446	40.5542
84	-2.772	57.0	97.459	-40.4589

Table 5 shows the residual analysis for the dummy variable regression given in Table 4. It includes a list of outliers, giving their case numbers, standardized residual, *TSP* levels, predicted values and residuals. Four of the residuals are fairly large, greater than 2.5 times 14.598 um/m^3 which is the standard error of estimate in Table 4. This is more than you would expect from only 137 observations. Consequently, the histogram of standardized residual in Figure 3 shows noticeable departure from normality. Specifically, it shows positive kurtosis too many observations in the extreme tails, which therefore inflate the standard deviation and create the impression of too many observations close to the mean.

The normal probability plot in Figure 3 also shows that observed values of the residuals at the top of the distribution are greater than those expected if the residual normally distributed.

Figure 3: Histogram of Standardized Residual and Normal Probability Plot



Outliers can have a disproportionate influence on trend estimates. Significant tests on regression coefficients depends on the assumption of normally distributed residuals and hence are also sensitive to outliers (SPSS Inc., 1999). Since our primary interest is to estimate the trend and testing the significance, we will smooth the outliers (replace them by less outlying values) and reestimate the regression equation.

Table 6 shows the basic results from the regression analysis after smoothing the outliers (replace them by less outlying values). It is similar to those previous regression analysis (Table 4), but notice that adjusted R^2 for the equation improved markedly, as it is expected when you remove the cases that are farthest from the regres-

sion line. The coefficient of trend is slightly smaller, but its standard error is much smaller. It is still statistically significant at 0.001 levels again.

Table 6: Regression Analysis Results of Smoothed Outliers

	Const.	TREND	JAN	FEB	MAR	MAY	JUN	JLY	AGT	SEP	OCT	NOV	DEC
B	52.16	-1.1	42.7	36.0	21.7	-12.7	-24.5	-26.7	-26.2	-21.6	-14.2	26.7	46.5
S.E.	4.28	0.33	5.33	5.33	5.33	5.33	5.35	5.35	5.35	5.35	5.35	5.35	5.35
Sig.	.000	.001	.000	.000	.000	.017	.000	.000	.000	.000	.009	.000	.000
R=91.5%		Adjusted R Square=82.1% DW=2.24			F=53.08 (0.000)			Std. Error of the Estimate=12.81					

The scatter plot in Figure 4 compares the residuals (on vertical axis) with the predicted values (on horizontal axis). The plot shows a funnel shape that the variance of the points at the right is more than the variance of the points at the left. The shape of the plot of residuals with the predicted values indicates that the residuals for observations with the predicted *TSP* levels have more variance than the residuals for observations with low predicted *TSP* level. Ordinarily least squares regression analysis assumes that the residual have a constant variance. This regression model evidently violates that assumption of constant variance, in technical language, the model shows heteroscedasticity.

The variance of regression errors increases with the predicted values. The components of predicted variables are *TREND* and 11 dummy month variables. We have already seen that *TSP* level vary with the seasons, averaging roughly 71 points higher in December and January than July and August (this is from the coefficients in Table 6). We know from experience that weather conditions are more variable in winter when *TSP* levels are high than in summer. Perhaps the pattern in the scatter plot is due to greater variance in *TSP* levels during the winter months. It is easy to check this.

Figure 4: Residuals with Predicted Values

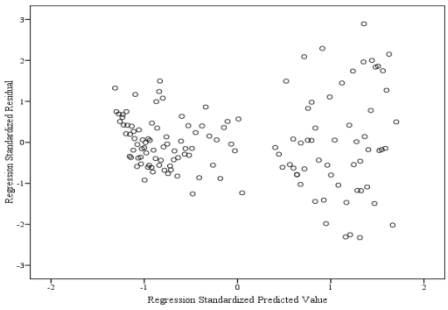


Figure 5: Residual Variance by Month

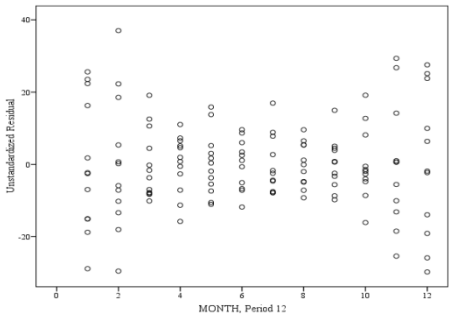


Figure 5 shows the residuals plotted against the month of the observation. This is not a time series plot (sequence chart); all the Januaries are plotted together, all the Februaries, and so on, so that it can be evaluated the variance of the residuals in each month.

Figure 5 shows an impressive sideways hourglass pattern. The residuals are spread out vertically in the early months, squeezed together during the summer

months 5-9, and spread out again at the end of the year. *TSP* levels in Trabzon fluctuated more in the winter when they are generally high than in the summer.

The heteroscedasticity of the residuals violates one of the assumptions of ordinary leas-squares regression, so some of the statistical results of the analysis above may not be reliable. To obtain more reliable and stabilized results, *weighted least squares regression* should be used.

Weighted Least Squares Regression

Standard linear regression models assume that variance is constant within the population under study. When this is not the case (for instance, when cases that are high on some attribute show more variability than cases that are low on that attribute) linear regression using ordinary least squares (OLS) no longer provides optimal model estimates (Norusis and SPSS Inc, 1999).

In the current problem, it is assume that *TSP* levels really are a linear function of *TREND* and other 11 dummy month variable, and the residuals have a different variance in each month due to transient conditions or measurement problems. Observations from July and August, a month with small residual variance, will count more heavily in determining the regression equation than observations from December and January, a month with large residual variance. This is reasonable, since the observations from January and December are likely to be farther from typical January and December value than observations from July and August are from the typical July and August value.

Figure 5 shows that the error variance differs according to the month of the observation. WLS is a technique that uses this information, giving more weight to precise observations and less weight to the highly variable observation. To use WLS, you must form a series that shows how much error you expect in each observation. The first step is to calculate how widely the *TSP* levels are spread within each month. This variable will be used as a weight (WGT) variable.

WLS regression results are shown in Table 7. The multiple correlation coefficient of 92.3% is greater than with ordinary least squares. The adjusted R^2 is still about 83.8%. The trend estimate is now only positive and only about 0.08 um/m^3 per year, which has a statistical significant of 0.681. The magnitude of the trend estimates change dramatically from previous regression analysis results. The constant, the estimated value at the beginning of the time period, with seasonal effect removed, is 45.4 um/m^3. There is no first order autocorrelation and multicolinearity problem in the built models. The Durbin-Watson statistics (DW) are 1.97, 2.19, 2.24 and 1.85 respectively. Multicolinearity tests were also conducted with variance inflation factors (VIF), and the VIF values were less than 2 which is well below the problematic level of 10 (Gujarati, 1995).

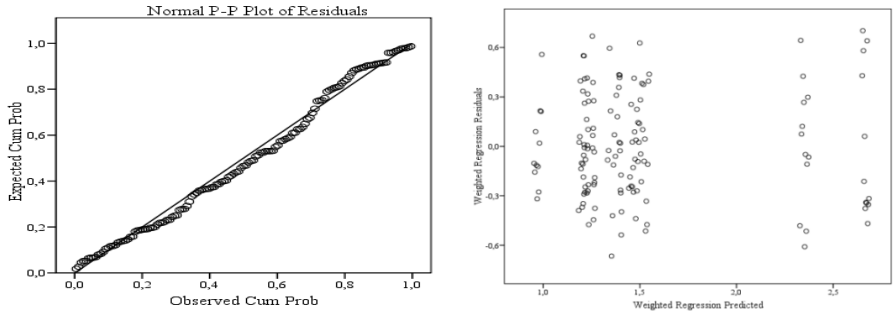
Table 7: Weighted Regression Analysis Results of Dummy Month Variables

	Constant	TREND	JAN	FEB	MAR	MAY	JUN	JLY	AGT	SEP	OCT	NOV	DEC
B	45.4	0.08	42.9	36.2	21.8	-12.8	-24.1	-26.4	-26.0	-21.5	-14.2	26.6	46.4
S.E.	2.18	0.19	6.35	6.49	3.06	2.79	2.46	2.73	2.30	2.49	3.23	6.25	7.71
t	20.78	0.41	6.76	5.57	7.12	-4.58	-9.80	-9.64	-11.31	-8.63	-4.38	4.25	6.01
Sig.	.000	.681	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000

R=92.3% Adjusted R Square=83.8% F=59.77 (0.000) Std. Error of the Estimate=0.331
DW=2.23

The regression estimates have changed again, this time showing a smaller and statistically significant positive trend. Evidently, less reliable observations made in highly variable winter months had contributed to the trend and dummy month variable estimates from ordinary least squares. It should be expected that the weighed least squares estimates to be more reliable and stable ones.

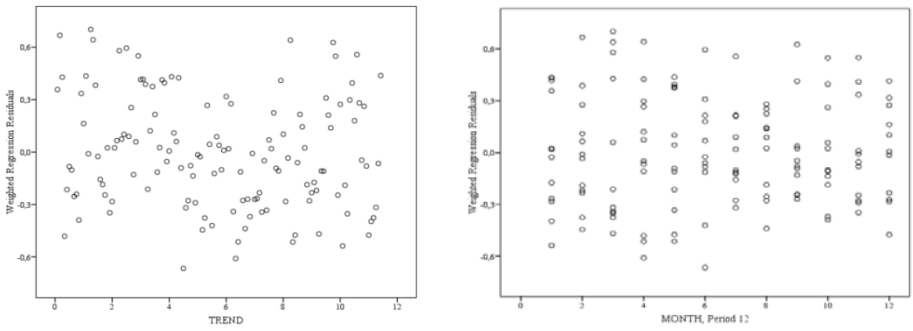
Figure 6: Unstandardized Residual and Normal Probability Plot of WLS



Our conclusion, then, is that over the 11 years periods the *TSP* level in Trabzon was not statistically significant increasing by about 0.08 um/m^3 each year.

The P-P Plot of weighted least squares residuals is shown in Figure 6. It is noticeable better than the plot of residuals from the ordinary least squares analysis shown in Figure 3. Finally, in Figure 7 the scatter plot between predicted values and residuals does not show the heteroscedasticity observed earlier.

Figure 7: Scatter Plot of Residuals against Trend and Residual Variance by Month



It is also a good idea to plot residuals against important independent variables. Figure 7 shows a plot of predicted residuals against trend *(TREND)* variable, whose

effect is being primarily interested in. Once again, there is no apparent pattern in this plot.

4. CONCLUSION

The *TSP* levels in Trabzon is reasonably more in January (90.75 um/m^3), February (82 um/m^3), March (67.58 um/m^3), November (72.39 um/m^3), and December (98.79 um/m^3), especially if the quality target of a daily value of 50 um/m^3 is taken into account, while in April (45.83 um/m^3), May (33.08 um/m^3), June (21.73 um/m^3), July (19.46 um/m^3), August (19.82 um/m^3), September (24.36 um/m^3) and October (31.68 um/m^3) month values respect the quality standard.

This study has demonstrated that the most appropriate trend prediction model for *TSP* series can be estimated by using weighted least squares (WLS) regression. Weighted least squares regression results showing a smaller and statistically insignificant positive trend. Evidently, less reliable observations made in highly variable winter months had contributed to the trend and dummy month variable estimates from ordinary least squares. It should be expected that the weighed least squares estimates to be more reliable ones.

There is no first order autocorrelation problem in all models, the Durbin-Watson statistics are 1.97, 2.19, 2.24 and 2.23. Multicolinearity tests were also conducted with variance inflation factors (VIF) and the VIF values were well below the problematic level of 10.

Our conclusion is that over the 11 years periods the *TSP* level in Trabzon was not statistically significantly increasing by about 0.08 um/m^3 each year.

The P-P Plot of weighted least squares residuals is noticeable better than the plot of residuals from the ordinary least squares analysis. Finally, in weighted least squares regression the scatter plot between predicted values and residuals does not show the heteroscedasticity observed earlier. It is also the plot of predicted residuals against trend *(TREND)* variable, whose effect is being primarily interested in, is not showing an apparent pattern.

REFERENCES

Aneja, V. P., A. Agarwal, P. A. Roelle, S. B. Phillips, Tong, Q. Watkins (2001), "Measurements and Analysis of Criteria Pollutants in New Delhi", India, *Environmental International*, 27, 35–42.

Chapman, R. S., W. P. Watkinson, K. L. Dreher and D. L. Costa (1997), "Ambient Particulate Matter and Respiratory and Cardiovascular İllness in Adults: Particle-Borne Transition Metals and The Heart–Lung Axis", *Environmental Toxicology and Pharmacology*, 4, 331–338.

European Council (1999), Air Quality Daumhter Directive 1999/30/CE.

Goswami, E., T. Larson, T. Lumley and L. Lieu (2002), "Spatial Characteristics of Fine Particulate Matter İdentifying Representative Monitoring Locations in Seattle, Washington", *Journal of Air and Waste Management Association,* 52, 324–333.

Gömer, D., F. Somunkıranoğlu and E. Tok (2006), "Introduction into the Twinning Project Air Quality of The European Commission", *City and Health Symposium*, Bursa: Uludağ University.

Gujarati, D. N. (1995), *Basic Econometrics (Third Edition)*, McGraw-Hill, New Jersey.

Gupta, I. and R. Kumar (2006), "Trends of Particulate Matter in Four Cities in India", *Atmospheric Environment, 40*, 2552–2566.

Harrison, R. and J. Yin (2000), "Particulate Matter in the Atmosphere: Which Particle Properties are Important for its Effect on Health", *The Science of the Total Environment*, 249, 85-101.

Hess, A., H. Iyer and M. Willium (2001), "Linear Trend Analysis a Comparison of Methods", *Atmospheric Environment*, 35, 5211–5222.

Hintze, J. L. and NCSS (2005), *User's Guide-II: Regression and Curve Fitting,* NCSS Inc., Kaysville.

Hintze, J. L. and NCSS (2005), User's Guide-IV: Multivariate Analysis, Clustering, Meta Analysis, Forcasting, Time Series, Operation Research and Mass Appraisal, NCSS Inc., Kaysville.

Jorquera, H., W. Palma and J. Tapia (2000), "An Intervention Analysis of Air Quality Data at Santiago Chile", *Atmospheric Environment*, 34, 4073–4084.

Mage, D., G. Ozolins, P. Peterson, A. Webster, R. Ortherfer, V. Vandeveerd (1996), "Urban Air Pollution in Megacities of the World", *Atmospheric Environment*, 30, 681–686.

Makridakis, S. and Wheelwright (1978), *Iterative Forecasting,* Holden-Day, California.

Mayer, H. (1999), "Air Pollution in Cities", *Atmospheric Environment* , 4029-4037.

Müller, W. J. (2006), "Exceeding of EU-Air Quality Limit Values Starts Clean Air Planning Examples From Germany and Europe", *City and Health Symposium,* Bursa: Uludağ University.

Norusis, M. J. and SPSS Inc (1999), *SPSS Regression Models 10.0,* SPSS Inc., Chicago.

Orhunbilge, N. (1998), *Zaman Serileri Analizi ve Tahmin Yöntemleri,* Avcıol-Basım Yayın, İstanbul.

Pope III., C., M. Thun, M. Namboodiri, D. Dockery, J. Evans, F. Speizer (1995), "Particulate Air Pollution is a Predictor of Mortality in a Prospective Study of US Adults", *Am. J.Respir. Crit. Care Med.*, 669–674.

Sapan, N. (2006), "Effects of Air Pollution On Child Health", *City and Health Symposium,* Uludağ University, Bursa.

Shin, K. (1996), SPSS Guide for DOS Version 5 and Windows Versions 6 and 6.1.2, Irwin, Chicago.

Turanlıoğlu, F. S., A. Nuhoğlu and H. Bayraktar (2005), "Impacts of Some Meteorological Parameters on SO_2 and TSP Concentrations in Erzurum,Turkey", *Chemosphere*, 59, 1633-1642.

Vardoulakis, S. and P. Kassomenos (2008), "Sources and Factors Affecting PM10 Levels in two European Cities: Implication for Local Air Quality Management", *Atmospheric Enviroment*, 42, 17, 58-67.