

# DEĞİŞEN VARYANS DURUMUNDA EN KÜÇÜK KARELER TEKNİĞİNİN ALTERNATİFİ AĞIRLIKLIL REGRESYON ANALİZİ VE BİR UYGULAMA

*Yrd.Doç.Dr. Ali Sait ALBAYRAK\**

## ÖZET

Bu çalışmanın amacı, en küçük kareler (EKK) tekniğinin en önemli varsayımlarından biri olan sabit varyans varsayımının teorik ve teknik ayrıntılarını ampirik bir uygulama üzerinde tartışmaktır.

Bu çalışmada, Zonguldak'ta toplam asılı partiküller madde düzeylerindeki trend ve mevsimlik faktörler uygun olan ağırlıklı regresyon (AR) yöntemiyle araştırılmaktadır. Araştırmada Ocak 1990 ile Haziran 2007 tarihleri arasındaki aylık toplam asılı partiküller serisi kullanılmaktadır. AR analizi ile elde edilen sonuçlardan, düzeltilmiş- $R^2$  değerinin %91 ve toplam asılı partiküller düzeylerinin ilgili dönemde yıllık trendinin negatif ve yaklaşık olarak  $-1.467 \text{ um/m}^3$  olduğu ve bunun 0.000 anlamlılık düzeyinde anlamlı olduğu anlaşılmaktadır.

**Anahtar Kelimeler:** EKK, Değişen Varyans, Ağırlıklı Regresyon, Hava Kalitesi.

## WEIGHTED REGRESSION ANALYSIS ALTERNATIVE TO THE LEAST SQUARES TECHNIQUE IN THE PRESENCE OF NONE CONSTANT VARIANCE AND AN APPLICATION

### ABSTRACT

The aim of this study is to discuss the theoretical and technical details of constant variance, which is the most important assumption of least squares technique.

In this study, an attempt has been made to identify and estimate the trend and seasonal factors of total suspended particles (TSP) in Zonguldak using accurate and reliable method of weighted least squares regression analysis. The data used in this study is concerned with monthly measurement levels of total suspended particles taken from January 1990 through June 2007. The most accurate model with WLS regression results show that the adjusted- $R^2$  is about 91% and the monthly averages of TSP have a negative trend over the period January 1990-June 2007, the trend estimate is only about  $-1.467 \text{ um/m}^3$  per year, which has a statistical significant of 0.000.

**Keywords:** LS, None Constant Variance, Weighted Regression, Air Quality.

---

\* Zonguldak Karaelmas Üniversitesi, İİBF, İşletme Bölümü, Sayısal Yöntemler Anabilim Dalı.

## 1. GİRİŞ

Bu çalışmanın amacı, istatistikte en küçük kareler (EKK) tekniğinin en önemli varsayımlarından birisi olan sabit varyans varsayımının teorik ve teknik ayrıntılarını ampirik bir uygulama üzerinde tartışmaktır.

EKK tekniği, regresyon analizinde sabit varyans varsayımının sağlanması durumunda kullanılması uygun olan bir tekniktir. Ağırlıklı en küçük kareler (AEKK) tekniği, doğrusal regresyon modelinde değişen varyans problemi olduğunda EKK tekniğine alternatif olarak geliştirilmiştir. Değişen varyans durumunda EKK tahminleri yansız olmakta, ancak varyans ve kovaryans tahminleri etkin (minimum varyanslı) olmadığından istatistik hipotez testleri geçerliliğini kaybetmektedir. Ayrıca belirli bir anlamlılık düzeyindeki tahmin ve öngörü aralıkları genişlemektedir.

Türkiye'nin Batı Karadeniz Bölgesinde yer alan Zonguldak'ta özellikle kış mevsiminde hava kirliliği en önemli çevre sorunlarından birisidir. 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 farklı faktörleri tahmin etmektir.

Bu çalışma dört bölümden oluşmaktadır. Giriş bölümünden sonra gelen ikinci bölümde sırasıyla değişen varyans sorununun tanımı, nedenleri, sonuçları, değişen varyans sorununun saptanması, değişen varyans sorununun çözümü, EKK ve AEKK tekniğinin teknik ayrıntıları tartışılmaktadır. Üçüncü bölümde ise, Zonguldak'ta toplam asılı partiküller (*TAP*) madde düzeylerindeki trend ve mevsimlik faktörler uygun olan ağırlıklı regresyon (AR) yöntemiyle araştırılmaktadır. Araştırmada Ocak 1990 ile Haziran 2007 tarihleri arasındaki aylık *TAP* serisi kullanılmaktadır.

## 2. EKK TEKNİĞİ, DEĞİŞEN VARYANS VE AĞIRLIKLIL REGRESYON ANALİZİ

### 2.1. Değişen Varyans Probleminin Tanımı, Nedenleri ve Sonuçları

EKK tekniğinin en önemli varsayımlarından birisi de sabit varyans varsayımdır (Tarı, 2006; Orhunbilge, 2000). EKK tekniği, bağımsız değişkenlerin birim değerleri değişirken bağımlı değişkenin birim değerlerine ait varyansın sabit kalacağını varsaymakta ve istatistik yazınında bu varsayıma sabit varyans (homoscedasticity) adı verilmektedir (Gujarati, 1995; Orhunbilge, 2000; Tarı, 2006; Yamak ve Köseoğlu, 2006). Diğer bir anlatımla hata terimi varyansı, bağımsız değişkenlerdeki değişimlerden etkilenmeyip hep aynı kalmaktadır. Sabit varyans varsayımı aşağıdaki gibi gösterilebilir (Yamak ve Köseoğlu, 2006):

$$E(\varepsilon_i^2) = \sigma^2 \quad i = 1, 2, \dots, N.$$

$$E(\varepsilon\varepsilon') = \sigma^2 \mathbf{I} = \sigma^2 \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \dots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}_{N \times N} = \begin{pmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \dots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{pmatrix}_{N \times N}$$

Hata teriminin varyansının farklı olması durumuna ise, değişen varyans (heteroscedasticity) adı verilmekte ve regresyon analizinde istenmeyen durumu göstermektedir. Bu durumda regresyon modelinin hatalarının varyansı sabit kalmayıp artan, azalan veya hem artan hem de azalan bir dağılım göstererek değişebilmektedir. Değişen varyans problemi matrisle aşağıdaki gibi gösterilebilir (Yamak ve Köseoğlu, 2006):

$$E(\varepsilon\varepsilon') = \delta^2 \Omega \quad \text{ve} \quad \Omega \neq \mathbf{I}$$

Burada  $\Omega$  matrisi, köşegen dışı elemanları sıfır fakat köşegen üzerindeki varyanslar birbirinden farklıdır.

Değişen varyans sorunu zaman, kesit ve panel verilerinde görülebilmektedir. Fakat değişen varyans sorunu daha çok kesit verilerinde karşılaşılan bir sorun olarak görülebilir. Değişen varyansın ortaya çıkmasının çeşitli nedenleri vardır. Bu nedenler arasında şunlar sayılabilir: (1) Önemli açıklayıcı değişkenlerin model dışında tutulması, (2) kesit verilerinde değişen varyans sorunu daha yaygın bir şekilde görülmesi, (3) mevsimsellik gösteren zaman serisinin modelde bağımlı değişken olarak kullanılması, (4) bağımlı değişkenin ölçümünün veya tanımının yanlış yapılması ve bu hatanın bağımlı değişken(ler)e göre değişmesi, (5) türdeş olmayan anaküteller üzerinde çalışılması.

Regresyon analizinde modele alınması gereken önemli açıklayıcı değişkenler, çoğunlukla modeldeki değişkenlerle aynı yönde ve büyüklükte değişmektedir. Bu durum, modelde değişen varyans problemine yol açabilmektedir. Örneğin; bireysel gelirin bağımsız, seyahat harcamalarının bağılı olduğu bir regresyon modelinde değişen varyans sorununun ortaya çıkmasının nedenlerinden biri, bireysel seyahat harcamalarındaki grup içi değişkenliğinin düşük gelir düzeylerinde az, yüksek gelir düzeylerinde ise fazla olmasıdır (Yamak ve Köseoğlu, 2006). Diğer taraftan, araştırmalarda kesit verilerinin kullanılması değişen varyans sorununa neden olabilmektedir. Örneğin; gelir, tüketim, tasarruf veya ücret gibi genellikle grup ortalamalarından oluşan kesit verilerlerinden oluşan fonksiyonlarla çalışılması durumunda değişen varyans sorunuyla karşılaşmaktadır (Tarı, 2006; Yamak ve Köseoğlu, 2006). Zira yüksek gelirli birimlerin tüketim alışkanlıklarının değişkenliği, düşük gelirli birimlerin tüketim alışkanlıklarının değişkenliğinden (varyansından) oldukça farklı olabilmektedir. Böylece düşük gelirli ailelerde tahmin hatalarının varyansı ve büyüklükleri daha düşük iken, yüksek gelirli birimlerde daha yüksek çıkmaktadır. Değişen varyans problemi kesit, zaman ve panel verilerde gözlenebilmektedir. Fakat bağımlı ve bağımsız değişkenleri aynı oranda değiştiği zaman serilerinde görülmemektedir. Örneğin; gelir ve tüketim gibi iki değişken zaman içerisinde aynı oranlarda artmaktadır. Fakat mevsimsellik gösteren zaman serilerinde genellikle değişen varyans sorunuyla karşılaşmaktadır.

Değişen varyans probleminin regresyon analizi sonuçları üzerinde olumsuz etkileri vardır. Değişen varyans durumunda EKK tah-

minleri yansızlık (unbiased) ve tutarlılık (consistent) özelliğini korumakta, fakat minimum varyanslı veya etkinlik (efficient) olarak bilinen özelliğini kaybetmektedir. Bu durum güvenilir olmayan istatistik testlere neden olmaktadır. Yani, EKK tahmincilerinin kısmi t ve genel F testleri güvenilirliklerini kaybetmektedir (Yamak ve Köseoğlu, 2006). Ayrıca geliştirilen modellerle gerçekleştirilen öngörüler etkin olmamaktadır.

## 2.2. Değişen Varyans Probleminin Saptanması

Literatürde değişen varyans problemini saptamada kullanılan birkaç istatistik test bulunmaktadır (Gujarati, 1995; Green, 1994; Pindyck ve Rubinfeld, 1991; Tarı, 2006; Webster, 1995; Yamak ve Köseoğlu, 2006). Bu testler şunlardır: Grafik yöntem, Park testi, Goldfeld-Quandt testi, Glejser testi, Spearman sıra korelasyon testi, Breusch-Pagan-Godfrey testi ve White nR-kare testi. Burada sadece araştırmacılar tarafından en yaygın kullanılan White nR-kare testi tanıtılmaktadır.

### 2.2.1. White nR-Kare Testi

White  $nR^2$  istatistiği Webster (1995) tarafından iki bağımsız değişkenli bir model için aşağıdaki üç aşamayla açıklanmaktadır:

- 1. Birinci Adım:** Hataları elde etmek için ilk regresyon modeli çözülür ve  $e_j$  değerleri elde edilir:  $y_j = b_0 + b_1x_{1j} + b_2x_{2j} + e_j$
- 2. İkinci Adım:** Hata birim değerinin karesi alınır ve hataların kareleri orijinal değişkenler ile bu değişkenlerin kareleri ve bu iki değişkenin çarpımları üzerine regresyon modeli kurularak modelin R-kare değeri elde edilir:  $e_j^2 = b_0 + b_1x_{1j} + b_2x_{2j} + b_3x_{1j}^2 + b_4x_{2j}^2 + b_5x_{1j}x_{2j} + v_j$
- 3. Üçüncü Adım:** Daha sonra White  $R^2$  ile n (örneklem hacmi) çarpılarak  $nR^2$  istatistiği hesaplanır.

White nR-kare istatistiği ki-kare dağılımına uymaktadır. Ki-kare dağılımını serbestlik derecesi k-1'e (k yardımcı modeldeki sabit terim dahil modeldeki parametre sayısını göstermektedir) eşittir. Hipotez testleri için istatistik karar ölçütleri ise aşağıdaki gibidir:

$$H_0 : \sigma_i^2 = \sigma^2 \quad \text{ve} \quad nR^2 \leq \chi_{k-1, \alpha}^2 \rightarrow H_0 \text{ Kabul edilir.}$$

$$H_0 : \sigma_i^2 \neq \sigma^2 \quad \text{ve} \quad nR^2 > \chi_{k-1, \alpha}^2 \rightarrow H_0 \text{ Reddedilir.}$$

White  $nR$ -kare istatistiği kritik ki-kare değerinden küçük veya eşitse sabit varyans durumunu ileri süren sıfır hipotezi kabul edilmektedir. Aksi durumda sıfır hipotezi reddedilmektedir.

Örnekleme hacmi 100'den ( $n \geq 100$ ) büyük olması durumunda sabit varyans varsayımı grafik yöntemle de saptanabilmektedir. Modelin önemli açıklayıcı değişkenleri ve tahminleri ile hataları arasındaki serpilme diyagramları ayrı ayrı incelenir. Hatalar ile tahminler (veya önemli bağımsız değişkenler) arasındaki dağılımın şekli artan, azalan veya kum saatine benziyorsa değişen varyans; dikdörtgensel bir dağılıma yakın bir dağılım söz konusu ise sabit varyans varsayımının sağlandığı kabul edilmektedir.

### 2.3. Değişen Varyans Probleminin Çözümü

Varsayımların sağlanmasında değişkenlerin dönüştürülmesi yaklaşımı yaygın bir şekilde kullanılmaktadır. Değişkenlere dönüşüm uygulamanın doğrusallaştırmak, normalleştirmek ve durağanlaştırmak (sabit varyans) gibi üç temel amacı vardır. Aşağıda en yaygın kullanılan dönüşümler kısaca açıklanmaktadır.

Bu dönüşümlerle genellikle son iki amaç, bazen de üçüncü amaç gerçekleştirilebilmektedir. Dönüşümlerle ilgili ayrıntılı tartışmaları birkaç değişik kaynakta bulmak mümkündür. Bulardan en önemlileri Armitage (1971), Draper ve Smith (1981) ve Neter, Wasserman ve Kunter (1983) sayılabilir. En yaygın kullanılan dönüşümler şunlardır (Albayrak, 2006):

- 1. Logaritmik Dönüşüm ( $Y^* = \ln Y$ ):** Bu dönüşüm negatif sayıların logaritması alınmayacağı için sadece pozitif birim değerli değişkenlere uygulanabilmektedir.  $y$  artarken  $y$ 'nin varyansı da artıyorsa varyansı durağanlaştırmak,  $y$ 'nin hataları sağa çarpıksa  $y$ 'nin dağılımını normalleştirmek ve bağımlı değişken ile bağımsız değişken arasında sürekli artan bir eğim söz konusu olması durumunda ise modeli doğrusallaştırmak için kullanılmaktadır.

2. **Karekök Dönüşümü ( $Y^*=Y^{0.5}$ ):** Varyans  $y$ 'nin ortalaması ile orantılı ise varyansı durağanlaştırmak için kullanılmaktadır. Bu dönüşüm özellikle bağımlı değişkenin Poisson dağılımına uyması durumunda kullanılmaktadır.
3. **Hiperbolik Dönüşüm( $Y^*=1/Y$ ):** Varyans  $y$ 'nin dördüncü dereceden kuvvetiyle orantılı ise ( $y$ 'nin ilk değerleri arasında aşırı farklar varsa) varyansı durağanlaştırmak için kullanılmaktadır. Bu dönüşümle serideki sapan değerler etkisiz hale getirilmektedir. Çünkü büyük sayıların tersi sıfıra daha yakın olacağından  $y$ 'deki sapan değerler  $y^*$ 'da önemsiz olmaktadır.
4. **Kare Dönüşümü ( $Y^*=Y^2$ ):** Varyans  $y$ 'nin ortalamasına göre azalıyorrsa varyansı durağanlaştırmak, bağımlı değişkenin hata değerleri sola çarpıksa bağımlı değişkeni normalleştirmek ve bağımsız değişkenlerden bazıları bağımlı değişken ile aşağıya doğru eğrisel bir ilişki (bağımsız değişken artarken eğimin ani düşmesi) göstermesi durumunda modeli doğrusallaştırmak için kullanılmaktadır.
5. **Arcsin Dönüşümü ( $Y^*=\text{Arcsin}Y^{0.5}=\text{Sin}^{-1}.Y^{0.5}$ ):**  $y$ , oran veya göreceli bir büyüklük ise varyansı durağanlaştırmak için kullanılmaktadır.

Ağırlıklı regresyon (AR) analizde, değişen varyans probleminin olduğu  $y_j = b_0 + b_1x_j + e_j$  regresyon doğrusuna aşağıdaki dönüşümlerden en küçük hata karelerini veren dönüşüm ağırlık olarak kullanılır.

1. Varyans terimlerinin bilinmesi durumu: Her bir birimi, bağımlı değişkenin standart sapmasına bölmek:

$$\frac{y_j}{\sigma_j} = \frac{b_0}{\sigma_j} + \frac{b_1x_j}{\sigma_j} + \frac{e_j}{\sigma_j}$$

2. Bağımlı değişkenin beklenen değerine bölmek:

$$\frac{y_j}{E(y)} = \frac{b_0}{E(y)} + \frac{b_1y_j}{E(y)} + \frac{e_j}{E(y)}$$

3. Bağımlı değişkenin tahmin edilen birim değerlerine bölmek:

$$\frac{y_j}{\hat{y}_j} = \frac{b_0}{\hat{y}_j} + \frac{b_1x_j}{\hat{y}_j} + \frac{e_j}{\hat{y}_j}$$

4. Bağımsız değişkenlerden birinin birim değerlerine bölmek:

$$\frac{y_j}{x_{ij}} = \frac{b_0}{x_{ij}} + \frac{b_1 x_{1j}}{x_{ij}} + \frac{b_2 x_{2j}}{x_{ij}} + \frac{e_j}{x_{ij}}$$

5. Bağımsız değişkenlerden birinin kareköküne bölmek:

$$\frac{y_j}{\sqrt{x_{ij}}} = \frac{b_0}{\sqrt{x_{ij}}} + \frac{b_2 x_j}{\sqrt{x_{ij}}} + \frac{e_j}{\sqrt{x_{ij}}}$$

6. Bağımlı ve bağımsız değişkenlere logaritmik dönüşüm uygulamak:

$$\ln y_j = \ln b_0 + b_1 \ln(x_{1j}) + b_2 \ln(x_{2j}) + e_j$$

Yukarıdaki eşitlikler Box-Cox dönüşümleri olarak bilinmektedir. Optimum çözüme hata kareleri toplamını minimum yapan dönüşümle ulaşılmaktadır (Bkz: Neter-Wasserman-Kunter, 1990:149–151).

Regresyon analizinde değişen varyans problemini çözmek için, EKK tekniğinin alternatifini olan ağırlıklı en küçük kareler (AEKK) tekniği kullanılmaktadır. SPSS istatistik programıyla iki ayrı batch dosyasıyla ağırlıklı regresyon çözümü elde edilebilmektedir.<sup>†</sup>

#### 2.4. Teknik Ayrıntılar: EKK ve AEKK Regresyon Yöntemleri

Bu bölümde en küçük kareler (EKK) ve ağırlıklı en küçük kareler (AEKK) regresyon yöntemlerinin matris yaklaşımıyla teknik ayrıntıları özetlenmektedir. Bu bölümdeki açıklamaların çoğu Hintze (2007) kaynağından özetlenerek verilmektedir. Aşağıda verilen eşitlikler hem EKK hem de AEKK regresyon çözümlerinde kullanılabilir. Sabit ağırlıklar kullanılmadığında aşağıdaki eşitlikler AEKK tekniğine, eşit ağırlıklar kullanıldığında ( $w_j=1$ ) ise EKK tekniğine dönüşmektedir. Tüm açıklamalarda aşağıda tanımlanan vektör ve matrislerden yararlanılmaktadır:

---

<sup>†</sup> Bunlardan birincisi “Lineer Regression” diyalog kutusunda WLS komutuyla elde edilmektedir. İkincisi ise “Weight Estimation” diyalog kutusuyla elde edilebilmektedir. Ancak SPSS ile değişen varyans problemini doğrudan test edecek bir prosedür bulunmamaktadır. Sadece White  $nR^2$  istatistiğini hesaplayabilmek için bir makro dosyası yazmak mümkündür. Fakat Eviews ve STATA programıyla White  $nR^2$  ve Breusch-Pagan-Godfrey testleri doğrudan elde edilebilmektedir.



$$\mathbf{Y} = \begin{bmatrix} y_1 \\ \vdots \\ y_j \\ \vdots \\ y_N \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{p1} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & x_{1j} & \cdots & x_{pj} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & x_{1N} & \cdots & x_{pN} \end{bmatrix}, \mathbf{e} = \begin{bmatrix} e_1 \\ \vdots \\ e_j \\ \vdots \\ e_N \end{bmatrix}, \mathbf{1} = \begin{bmatrix} 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_p \end{bmatrix}$$

$$\mathbf{W} = \begin{bmatrix} w_1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \vdots \\ 0 & 0 & w_j & 0 & 0 \\ \vdots & 0 & 0 & 0 & 0 \\ 0 & \cdots & 0 & 0 & w_N \end{bmatrix}$$

Yukarıdaki gösterim kullanılarak genel olarak AEKK tahminleri aşağıdaki eşitliklerle elde edilmektedir.

$$\mathbf{b} = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1} \mathbf{X}'\mathbf{W}\mathbf{Y}$$

Burada eşit ağırlıklar kullanıldığında yukarıdaki eşitlik aşağıdaki eşitliğe indirgenmektedir.

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}$$

Bağımlı değişkenin tahminleri ve hataları sırasıyla aşağıdaki eşitliklerle elde edilir.

$$\hat{\mathbf{Y}} = \mathbf{b}'\mathbf{X} \quad \text{ve} \quad \mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}}$$

Hataların tahmin edilen varyansları ise aşağıdaki gibi hesaplanmaktadır.

$$s^2 = \frac{\mathbf{e}'\mathbf{W}\mathbf{e}}{N-p-1}$$

Regresyon katsayılarının tahmin edilen varyansları aşağıdaki gibi hesaplanmaktadır.

$$V \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_p \end{pmatrix} = s^2 (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}$$

Belirli bir X değeri ( $X_0$  ile gösterilsin) için, Y ortalama tahmininin varyansı ise aşağıdaki eşitlikle hesaplanır.

$$s_{Y_m/X_0}^2 = s^2 (1, \mathbf{X}_0) (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1} \begin{pmatrix} 1 \\ \mathbf{X}_0 \end{pmatrix}$$

Belirli bir X değeri ( $X_0$  ile gösterilsin) için, tek bir Y tahmininin varyansı ise aşağıdaki eşitlikle hesaplanmaktadır.

$$s_{Y/X_0}^2 = s^2 + s_{Y_m/X_0}^2$$

Sabit ve diğer regresyon katsayılarının hipotez sınamaları ve güven aralıkları sırasıyla aşağıdaki gibi gerçekleştirilir.

$$t_{b_i} = \frac{b_i - B_i}{s_{b_i}} \text{ ve } P(b_i - t_{N-p-1, 1-\alpha/2} \cdot s_{b_i} < \beta_i < b_i + t_{N-p-1, 1-\alpha/2} \cdot s_{b_i}) = 1 - \alpha$$

Regresyon modelinin iyi-uyum derecesinin değerlendirilmesinde en yaygın kullanılan  $R^2$  ve düzeltilmiş- $R^2$  değerleri ise aşağıdaki eşitliklerle hesaplanmaktadır:

$$R^2 = 1 - \left( \frac{\mathbf{e}'\mathbf{W}\mathbf{e}}{\mathbf{Y}'\mathbf{W}\mathbf{Y} - \frac{(\mathbf{1}'\mathbf{W}\mathbf{Y})^2}{\mathbf{1}'\mathbf{W}\mathbf{1}}} \right)$$

$$\text{Düzeltilmiş } R^2 = \bar{R}^2 = 1 - \frac{(N-1)(1-R^2)}{N-p-1}$$

Bir sonraki bölümde Zonguldak ilinin hava kalitesi ile ilgili en önemli göstergelerinden birisi olan aylık toplam asılı partiküller (*TAP*) serisinin yıllık trendi ve mevsimlik faktörleri uygun regresyon tekniğiyle araştırılmaktadır.

### 3. UYGULAMA: ZONGULDAK İLİNİN ÇEVRE KİRLETİCİLERİNDEN TOPLAM ASILI PARTİKÜLLER SERİSİNİN TRENDİNİN TAHMİN EDİLMESİ

Bu çalışmada Zonguldak ilinin Ocak 1990-Haziran 2007 dönemine ait aylık *TAP* serisi kullanılmaktadır. Çalışmanın temel amacı, *TAP* serisinin uzun dönemli eğilimini araştırmaktır. Bu amacı gerçekleştirebilmek için, *TAP* serisi trend ve mevsimli kukla değişkenlerin doğrusal bir fonksiyonu olarak ifade edilmektedir. Model uygun olması koşuluyla zaman değişkeninin katsayısı uzun dönemli eğilimi gösterecektir.

*TAP* serisi ile ilgili tanımsal istatistikler Tablo 1’de gösterilmektedir. *TAP* serisi 82.65  $\mu\text{m}^3$  değerinde genel bir ortalamaya sahiptir. *TAP* serisi kış aylarında ortalamasının üstünde, yaz aylarında ise serinin genel ortalamasının altında değerler almaktadır. *TAP* serisi Kasım, Aralık, Ocak, Şubat ve Mart aylarında en yüksek ortalama ve standart sapma değerlerini almaktadır. Diğer aylarda ise tersi bir durum gözlenmektedir. Aralık ve Ocak aylarına ait %95 güven sınırları sırasıyla yaklaşık olarak 148–180  $\mu\text{m}^3$  ve 151–182  $\mu\text{m}^3$  tür.

**Tablo 1: Aylık *TAP* Serisi İçin Tanımsal İstatistikler**

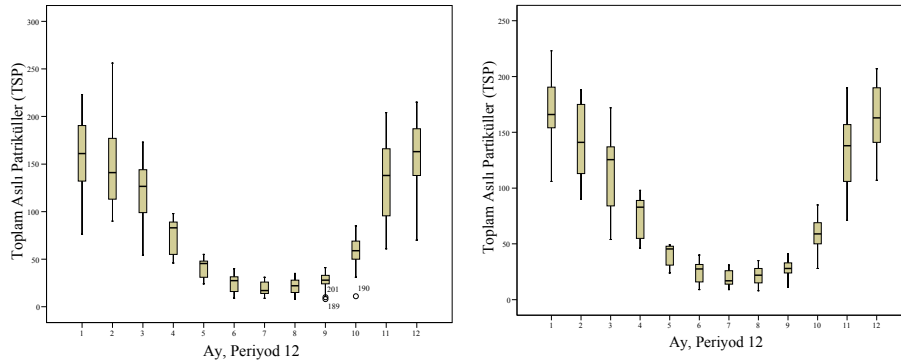
Ay	n	Ortalama	Std. Sapma	Std. Hata	Eğiklik	Basıklık	Ortalama İçin %95 GA		Min	Mak.
							Alt Sınır	Üst Sınır		
OCK	18	166.31	31.057	7.320	-178	-0.368	150.86	181.75	106	223
ŞUB	18	141.72	33.029	7.785	-156	-1.406	125.30	158.15	90	188
MRT	18	116.42	33.514	7.899	-294	-0.685	99.75	133.08	54	172
NIS	18	74.89	18.301	4.314	-552	-1.241	65.79	83.99	46	98
MAY	18	39.78	9.340	2.202	-330	-1.837	35.13	44.42	24	49
HAZ	18	24.89	9.454	2.228	-136	-0.937	20.19	29.59	9	40
TEM	17	19.53	7.451	1.807	.301	-1.305	15.70	23.36	9	31
AĞT	17	21.12	7.721	1.873	-085	-0.917	17.15	25.09	8	35
EYL	17	26.76	8.955	2.172	-458	-0.540	22.16	31.37	11	41
EKM	17	60.06	17.286	4.192	-266	-0.689	51.17	68.95	28	85
KAS	17	132.50	36.879	8.944	-142	-0.920	113.54	151.46	71	190
ARL	17	163.82	31.554	7.653	-273	-0.799	147.60	180.05	107	207
Toplam	210	82.65	60.302	4.161	.533	-1.075	74.45	90.85	8	223

Şekil 1, aylara göre sapan birimler düzeltilmeden önceki ve düzeltildikten sonraki (daha makul değerlerle değiştirilmiş) *TAP* düzeyleri serilerini göstermektedir. Bilindiği gibi, regresyon katsayılarının büyüklükleri ve standart hataları ve bunun bir sonucu olarak anlamlılık

testleri bu tür aykırı değerlerden olumsuz bir şekilde etkilenmektedir. Şekildeki gösterim box-whisker grafikleri olarak da adlandırılmaktadır. Her ay grubu için, kutuların içindeki koyu çizgi ilgili ayın örnek ortanca değerini göstermektedir. Böylece, örneğin Ocak ayının ortanca değeri  $166 \text{ um/m}^3$  iken, Temmuz ayı için yaklaşık olarak  $20 \text{ um/m}^3$  tür.

Şekil 1'deki kutuların alt ve üst sınır çizgileri sırasıyla birinci kartil (yirmi beşinci yüzdelik) ve üçüncü kartil (yetmiş beşinci yüzdelik) değerlerini göstermektedir. Kutuların uzunlukları ise kartillerarası değişim aralıklarını göstermektedir. Mayıs Haziran, Temmuz, Ağustos, Eylül ve Ekim ayları için kartiller arası değişim aralıkları çok dar iken, Ocak, Şubat, Mart, Nisan, Kasım ve Aralık ayları için çok daha geniş olduğu görülmektedir. Kutuların alt ve üst bölgelerindeki dikey çizgiler ise, ayların değişim aralık düzeylerini göstermektedir. Şekil 1'den aylara göre TAP düzeyleri ortalamaları ve değişkenlikleri önemli farklılıklar göstermektedir. Nisan ve Mayıs aylarının ortanca değerleri kutunun üst sınırına yakın olması serini sola çarpık olduğunu göstermektedir. Diğer bir anlatımla bu aylardaki TAP düzeyleri genelde yüksek düzeylerde yığıldığını göstermektedir. Diğer taraftan Ocak, Şubat, Mart, Ekim, Kasım ve Aralık aylarında daha simetrik dağılımlar söz konusudur.

**Şekil 1: Sapan Birimleri Düzeltmeden Önceki ve Sonraki Aylık TAP Serileri**

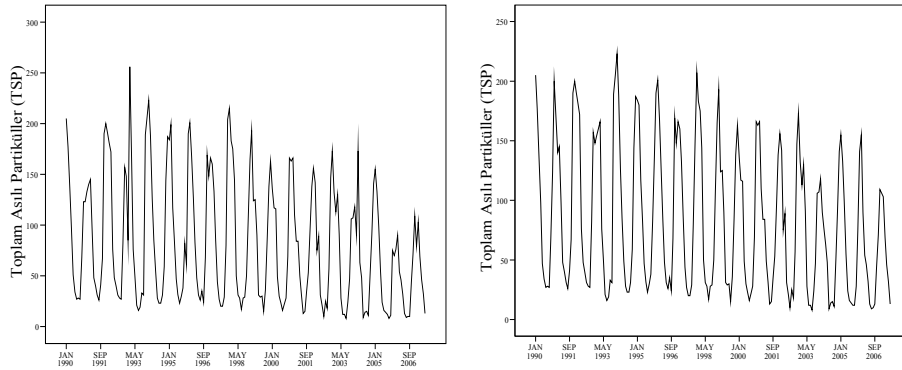


Şekil 1'de ayrıca serinin birim değerlerinin hangi değerlerde toplandığı ve dağılımın şekliyle ilgili yeterli bilgileri sağlamak için

sapan (outliers) ve uç (etremes) birim değerleri de gösterilmektedir. Birinci kartil değerlerinin altından veya üçüncü kartil değerlerinin üstünden kutuların (kartillerarası değişim aralıklarının) üç katından daha uzakta yer alan uç (extreme) birim değeri yıldız (\*) simgesiyle gösterilirken, 1.5 ile 3 katı arasında yer alan sapan birimler çember (o) simgesiyle gösterilmektedir. Kısaca, 189 ve 201 sıra numaralı Ağustos aylarına ait *TAP* düzeyleri ile 190 sıra numaralı Eylül ayına ait *TAP* düzeylerinin birer sapan birim olduğu anlaşılmaktadır.

Bir model geliştirmeden önce serinin tabiatını, yapısını tanımak her zaman yararlı olmaktadır. Diğer bir anlatımla verilerde mevsimlik hareketler veya uzun dönemli bir eğilim olup olmadığı yüzeysel olarak incelenebilir. *TAP* serisinin sapan birimleri düzeltilmeden önceki ve düzeltildikten sonraki zaman yolu grafikleri Şekil 2’de gösterilmektedir.<sup>‡</sup>

### Şekil 2: Sapan Birimleri Düzeltilmeden Önceki ve Sonraki Aylık *TAP* Serileri



Şekil 2, *TAP* serisi eşit zaman aralıklarında tekrar eden çok sayıda doruk ve dip noktaları gösterirken, çok açık olmayan aşığıya doğru uzun dönemli bir eğilim gözlenmektedir. Eşit zaman aralıklarında tekrar eden serideki periyodik (mevsimlik) hareketleri göstermektedir. Ayrıca serideki üç sapan birim değeri, kendi komşu değerlerinden oldukça uzakta duran birimleri göstermektedir. Bu birimler

<sup>‡</sup> Söz konusu sapan birimler silinip ilgili aydan önceki ve sonraki iki ayın aritmetik ortalaması alınarak tahmin edilen daha makul değerler bu sapan birimler yerine kullanılmıştır.

Şekil 2’de sağ taraftaki grafik üzerinde düzeltilmiş olarak gösterilmektedir.

### ***TAP Aylık Serisindeki Eksik Verilerin Tahmin Edilmesi***

İlk olarak serideki 5 aya ait eksik veriler uygun bir yöntemle tahmin edilmiştir. Bu konuda iki hususun dikkate alınması gerekmektedir: Bazı zaman serisi analiz yöntemleri birim değerlerinin eşit zaman aralıklı olarak gözlenmesini zorunlu kılmaktadır. Üzerinde çalıştığımız *TAP* serisinde ölçülmemiş eksik veri olsa da, tüm aylar bir birim değerine sahip olması gerekmektedir. Seri tüm aylar için tamamlandıktan sonra, seri içindeki eksik verilerin nasıl tahmin edileceği aşamasına gelinmektedir. Burada doğrusal enterpolasyon yöntemiyle eksik ayların verileri, bir önceki ve bir sonraki aya ait verilerin aritmetik ortalaması alınarak tahmin edilmektedir.

### ***Aylık TAP Serisi İçin Bir Trend Değişkeninin Hesaplanması***

Daha önce belirtildiği gibi, bu çalışmanın temel amacı *TAP* serisinin trendini hesaplamaktır. Her ay için trend çok küçük olacağından, trendi yıl bazında incelemek daha anlamlı olmaktadır. Trendi her bir yıl için hesaplayabilmek için, serinin başından sonuna kadar kaç yıl olduğunu gösteren bir değişkene ihtiyaç duyulmaktadır. Bu tür bir değişkeni hesaplamanın değişik birkaç yolu olabilir. Belki bu değişkenin birim değerlerini hesaplamanın en basit yolu, serinin başından sonuna kadar tüm birim değerlerine 1’den başlayarak sıra numarası verip, daha sonra bu değişkenin her bir birim değerini 12’ye bölmektir.<sup>§</sup>

### ***Gerçek Trendi Tahmin İçin Mevsimselliğin Giderilmesi***

Aylık *TAP* serisindeki gerçek trendi tahmin edebilmek için, öncelikle serideki mevsimlik faktörlerden kaynaklanan değişimlerin düzeltilmesi gerekmektedir. Örneğin, kış aylarındaki *TAP* düzeyleri yaz aylarındaki *TAP* düzeylerinden yüksekse, bu gerçek trendin uygun bir şekilde hesaplanmasını zorlaştırmaktadır. Zaman serilerini bileşenlere

---

<sup>§</sup> 1990 yılının Ocak ayından 2007 yılının Haziran ayına kadar toplam 210 birim değeri (ay) vardır. Bu veriler 17.5 yıllık bir veri setine karşılık gelmektedir. 210, 12’ye bölüldüğünde (210/12=17.5) 17.5 yıl değerine ulaşılmaktadır.

ayırma yöntemi (decomposition method), zaman serisini mevsimlik, trend, konjonktür ve hata (tesadüfi) bileşenlerine ayırmaktadır (Orhunbilge, 1998). Genelde bileşenlere ayırma yönteminde zaman serisi mevsimlik, trend, konjonktür ve hata faktörlerinin çarpımı olarak ifade edilmektedir. Çarpımsal model serinin varyansının sabit olması durumunda uygun olmaktadır. Serinin ortalamasına göre varyansın değişmemesi durumunda alternatif bir yöntem olarak toplamsal model kullanılabilir. Bu iki yöntemle elde edilen çarpımsal ve toplamsal mevsimlik endeks sonuçları Tablo 2’de verilmektedir.

**Tablo 2: TAP Serisinin Mevsimlik Faktörleri**

Mevsimlik Faktör	Ay											
	OCAK	ŞUB	MRT	NİS	MAY	HAZ	TEM	AGT	EYL	EKM	KAS	ARL
Toplamsal	84.3	58.6	34.9	-7.9	-43.1	-57.5	-63.4	-61.4	-55.4	-21.9	50.8	82.2
Çarpımsal	201.3	170.9	140.7	92.1	47.4	29.8	23.0	25.2	32.4	74.7	160.4	202.3

Tablo 2, her bir ay için mevsimlik etkilerden arındırılmış TAP düzeylerindeki ortalama sapmayı göstermektedir. Ocak aylarındaki TAP düzeyleri, mevsimlik etkilerden arındırılmış TAP düzeylerine göre ortalama olarak  $84.3 \text{ um/m}^3$  daha yüksektir. Tablo 2’den görüldüğü gibi Kasım, Aralık, Ocak, Şubat ve Mart aylarında TAP düzeyleri en yüksek iken, Mayıs, Haziran, Temmuz, Ağustos ve Eylül aylarında en düşüktür.

**Tablo 3: Mevsimsel Olarak Düzeltilmiş TAP Düzeyleri İçin EKK Sonuçları**

Standart Olmayan Regresyon Katsayıları				
	B	Std. Hata	t	Anl.
Sabit	107.00	3.26	32.82	.000
Trend	-2.90	0.32	-9.03	.000
R=%53.1	Düzeltilmiş- $R^2$ =%27.8	F=81.58 (0.000)	Tahminin Std. Hatası=23.54	

Çarpımsal mevsimlik bileşenlere ayırma yöntemi kullanılması durumunda mevsimlik indeksler yüzdesel olarak ifade dilmektedir. TAP düzeylerinin yüksek olduğu Kasım, Aralık, Ocak, Şubat ve Mart aylarında indeksler 100’ün üstünde iken; TAP düzeylerinin düşük olduğu Mayıs, Haziran, Temmuz, Ağustos ve Eylül aylarında ise 100’ün altında değerler almaktadır. Çarpımsal ve toplamsal modellerde her bir ayın gözlem değerinin nasıl ortalandığını gösterdiğinden, bu değer-

ler doğrudan birbirine dönüştürülemezdir. Verilerin yapısına uygun olan çarpımsal modellerle mevsimlik etkilerden arındırılmış seri kullanılarak, *TAP* düzeylerindeki trend tahmin edilebilir.

Tablo 3, mevsimlik etkilerden arındırılmış *TAP* serisi için regresyon analizinin sonuçlarını göstermektedir. Yıllık *TREND* değişkeninin katsayısı yaklaşık olarak  $-2.90 \text{ um/m}^3$  tür. Mevsimlik faktörlerden arındırılmış *TAP* serisi 17.5 yıllık zaman aralığında çok yavaş olarak düştüğü anlaşılmaktadır. Bu katsayı 0.000 düzeyinde anlamlıdır. Model bağımlı değişkendeki toplam varyansın önemli bir kısmını açıklayamamaktadır. Değişken ve birim sayısına göre hesaplanan düzeltilmiş- $R^2$  değeri yaklaşık olarak %27.8 dir. Tahminin standart hatası ise  $23.54 \text{ um/m}^3$  tür.

#### ***Mevsimlik Faktörlerin ve Trendin Eşzamanlı Tahmini: Kukla Değişkenli Regresyon Analizi***

Mevsimlik bir zaman serisini çözümlemenin pratik yollarından birisi, yukarıda açıklandığı gibi, analiz öncesi seriyi mevsimlik faktörlerden arındırmaktır. Mevsimsel düzeltmeler yapmadan mevsimsel bir zaman serisini çözümlemenin bir yolu regresyon analizinde mevsimlik kukla değişkenlerinin kullanılmasıdır. Bu çalışmada 12 ay için 11 kukla değişken kullanılmaktadır. En düşük ortalama *TAP* düzeyi (19.53) Temmuz ayında gözlemlendiğinden, karşılaştırma yapmak amacıyla bu ay referans ayı olarak seçilmiştir.

Mevsimlik kukla değişkenli regresyon analizinin sonuçları Tablo 4’te özetlenmektedir. Kukla değişkenli regresyon modelinin düzeltilmiş- $R^2$  değeri Tablo 3’teki değerden daha yüksektir. Modelde kullanılan değişken ve birim sayısına göre düzeltilmiş- $R^2$  değerinin *TAP* serisinin toplam varyansının %84.3’ü model tarafından açıklandığı anlaşılmaktadır.  $R^2$  değerindeki bu iyileşme, analiz öncesi mevsimlik faktörlerin düzeltilmesi yerine, bu faktörlerin modele alınmasından kaynaklanmaktadır.

Tablo 3’teki modelin standart hatası ( $23.54 \text{ um/m}^3$ ), Tablo 4’teki modelin standart hatasından biraz daha düşüktür ( $24.17 \text{ um/m}^3$ ). Bu durum mevsimlik etkilerden arındırılmış *TAP* serisinin daha kolay tahmin edilebileceğini göstermektedir. Kukla değişkenli regresyon analizi, toplamsal bileşenlere ayırma yöntemiyle hemen hemen aynı



şey gerçekleştirilmiştir. Fakat modelin serbestlik derecesini arttırarak, modelin standart hatalarını yükseltmiştir.

**Tablo 4: Kukla Değişkenli Regresyon Analizinin Sonuçları**

	SABİT	TREND	OCK	ŞUB	MRT	NIS	MAY	HAZ	AGT	EYL	EKM	KAS	ARL
B	44.57	-2.91	135.0	128.8	102.7	56.1	21.8	6.6	1.8	7.3	40.3	112.9	137.3
SE	6.51	0.33	8.16	8.16	8.16	8.16	8.16	8.16	8.29	8.29	8.29	8.29	8.29
Anl.	.000	.000	.000	.000	.000	.000	.008	.423	.000	.826	.379	.000	.000
R=%92.3    Düzeltilmiş- $R^2$ =%84.3    F=94.46 (0,000)    Tahminin Std. Hatası=24.17    DW=1.84													

Modelin *SABİT* katsayısı 107 birimden 44.57 birime ( $\text{um/m}^3$ ) düşmüştür. *TREND* değişkeninin katsayısı çok az artarak -2.91  $\text{um/m}^3$  olarak elde edilmiş ve bu katsayı 0.000 düzeyine anlamlıdır.

Kukla değişkenlerinin regresyon katsayıları, ilgili ayın Temmuz (referans ayı) ayına göre ortalama *TAP* düzeylerindeki etkiyi göstermektedir. Tablo 2 incelendiğinde Temmuz ayının mevsimlik etkisinin sıfırdan uzak ( $-63.4 \text{um/m}^3$ ) olduğundan, kukla değişkenlerle hesaplanan katsayılar ile toplamsal Mevsimlik Bileşenlere Ayırma Yöntemiyle elde edilen değerler çok yakın çıkmamıştır.

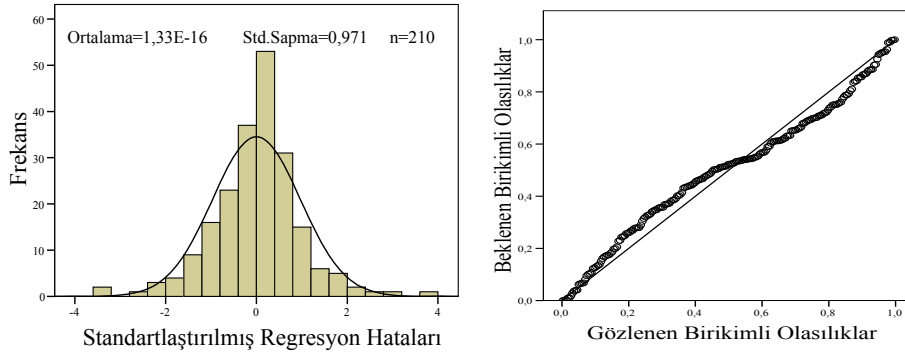
**Tablo 5: Hataların Analizi**

Birim Sıra No	Std. Hata	<i>TAP</i> Düzeyi	Tahmin Değeri	Hata
37	-3.538	85	170.51	-85.510
38	3.802	256	164.09	91.907
71	-3.274	61	140.13	-79.129
95	2.883	204	134.32	69.682
171	2.780	173	105.80	67.204
192	-2.699	70	135.25	-65.247

Tablo 5, Tablo 4'te verilen kukla değişkenli regresyon modelinin hatalarının analizini göstermektedir. Tablo modeldeki sapan birimleri, bu birimlerin sıra numarasını, standartlaştırılmış hatalarını, *TAP* düzeylerini, tahminlerini ve hatalarını göstermektedir. Bu hatalardan 6 tanesi, Tablo 4'te verilen 24.14  $\text{um/m}^3$  olan tahminin standart hastasından 2.5 kat daha büyüktür. 210 tane birim içerisinde bu kadar sapan birim fazla görülebilir. Sonuç olarak, Şekil 3'te gösterilen standartlaştırılmış hataların grafikleri normallikten önemli sapmalar göstermektedir. Daha açık bir ifadeyle hatalar pozitif kurtosis ve uç derlerde çok sayıda gözlem değerinin yer aldığını göstermektedir. Bu durum stan-

dart hataları yükseltmenin yanında, çok sayıda birimin ortalama üzerinde toplandığı izlenimini vermektedir. Şekil 3'teki normal olasılık grafiği ayrıca dağılımın tepesindeki hataların gözlenen değerlerinin beklenen değerlerinden daha büyük olduğunu göstermektedir.

**Şekil 3: Standartlaştırılmış Hataların Histogram ve Normal Olasılık Grafikleri**



Sapan birimler trendin tahminini olumsuz şekilde etkilemektedir. Regresyon katsayılarının anlamlılıkları ayrıca normalite varsayımına dayanmaktadır. Bu nedenle bu anlamlılık testleri de sapan birimlere karşı oldukça duyarlıdır (SPSS Inc., 1999). Bu çalışmada temel amaç *TAP* serisinin trendini ve mevsimlik faktörlerini uygun olarak tahmin etmek olduğundan, daha önce belirtilen yöntemle serideki sapan birimler düzeltilerek regresyon modeli yeniden tahmin edilmiştir.

**Tablo 6: Düzeltilmiş Sapan Birimlerle Elde Edilen Regresyon Analizinin Sonuçları**

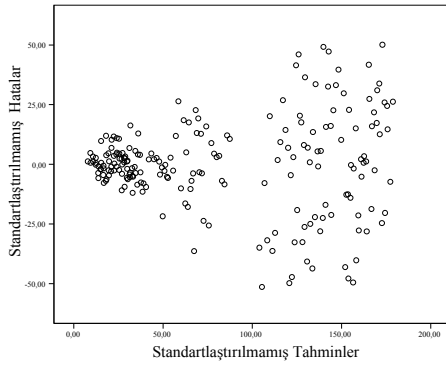
	SABİT	TREND	OCK	ŞUB	MRT	NİS	MAY	HAZ	AGT	EYL	EKM	KAS	ARL
B	44.19	-2.87	146.8	122.4	97.4	56.1	21.2	6.6	1.8	7.7	4.3	113.9	145.5
SE	4.86	0.25	6.10	6.10	6.10	6.10	6.10	6.10	6.19	6.19	6.19	6.19	6.19
Anl.	.000	.000	.000	.000	.000	.017	.000	.284	.768	.214	.000	.000	.000
R=%95.7 Düzeltilmiş-R <sup>2</sup> =%91.0 F=178.1 (0.000) Tahminin Std. Hatası=18.05 DW=1.92													

Tablo 6, sapan birimleri düzeltilmiş *TAP* serisinin regresyon analizi sonuçlarını göstermektedir. Dikkat edilirse, bu sonuçlarla Tablo 4'te elde edilen sonuçlar birbiriyle oldukça örtüşmektedir. Fakat regresyon doğrusundan uzak duran sapan birimler düzeltildikten sonra, beklendiği gibi  $R^2$  değerinde önemli bir artış olmuştur. *TREND* değişkeninin katsayısı biraz küçülerek -2.87 olarak elde edilmiş, fakat bu değişkenin standart hatasının daha fazla küçüldüğü görülmektedir.

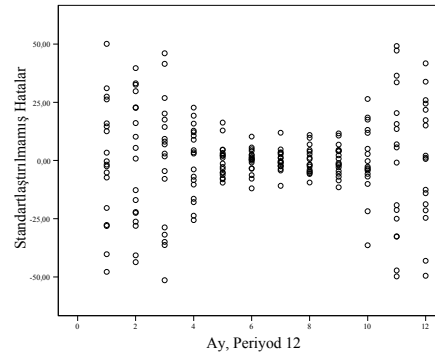
*TREND* değişkeni hala 0.000 düzeyinde anlamlı olduğu anlaşılmaktadır.

Şekil 4'te verilen serpilme diyagramı, modelin hataları (yatay eksen) ile tahmin değerlerini (dikey eksen) karşılaştırmaktadır. Hataların grafiği sağdaki hatların varyansının soldaki hataların varyansından daha büyük olduğunu gösteren bir huniye benzemektedir. Daha önce de belirtildiği gibi, EKK tekniği hataların sabit varyanslı olduğunu varsaymaktadır. Burada regresyon modelinin sabit varyans varsayımının sağlanmadığı açıkça görülmektedir. Daha teknik bir ifadeyle modelde değişen varyans (heteroscedasticity) sorunu vardır.

Şekil 4: Hatalar ve Tahminler



Şekil 5: Aylara Göre Hata Varyansı



Şekil 4 incelendiğinde regresyon hatalarının varyansı tahminlerle birlikte artmaktadır. Modelde *TREND* ve 11 aylık kukla değişkenin katsayıları tahmin edilmektedir. Zonguldak'ta *TAP* düzeyleri mevsimlere göre önemli farklılıklar göstermektedir. Örneğin, Tablo 6'daki regresyon katsayıları incelendiğinde Aralık ve Ocak aylarındaki *TAP* düzeylerinin Haziran ve Ağustos aylarına göre ortalama olarak 142 birim daha yüksek olduğu anlaşılmaktadır. Gözlemlerden *TAP* düzeylerin yüksek olduğu Ekim, Kasım, Aralık, Ocak, Şubat, Mart ve Nisan aylarındaki hava koşullarının diğer aylara göre daha fazla değişken olduğu bilinmektedir. Aylara göre *TAP* serisindeki değişimleri grafik üzerinde kolayca gösterilebilir.

Şekil 5, hataların aylara göre dağılımını göstermektedir. Şekil 5, bir zaman yolu grafiği (sequence chart) değildir. Şekil 5, aylara göre

hataların dağılımının açık olarak bir kum saatine benzediğini göstermektedir. Hatalar yılın ilk ve son aylarında önemli değişkenlikler gösterirken, yaz aylarında (5-9) daha türdeş (homojen) bir dağılım sergilemektedir.

Hatalarda gözlenen değişen varyans problemi, EKK tekniğinin en önemli varsayımlarından biri olan sabit varyans varsayımının sağlanamadığını göstermektedir. Bu nedenle yukarıda verilen bazı istatistik sonuçlar güvenilir olmayabilir. Daha güvenilir ve durağan sonuçlar elde edebilmek için *ağırlıklı regresyon analizinin* kullanılması gerekmektedir.

### ***Ağırlıklı Regresyon (AR) Analizi***

Standard doğrusal regresyon modelleri, üzerinde çalışılan anakütlenin varyansının sabit olduğunu varsaymaktadır. Aksi durumda en küçük kareler regresyon analizi, en iyi tahminlerini sağlayamamaktadır (Norusis ve SPSS Inc, 1999).

Bu çalışmada, aylık *TAP* düzeylerinin *TREND* ve 11 aylık kukla değişkenlerinin doğrusal bir fonksiyonu olduğu ve hataların aylar itibariyle farklı varyanslara sahip olduğu varsayılmaktadır. AR analizinde regresyon modeli belirlenirken küçük varyanslı Temmuz ve Ağustos ayları gibi birimlere daha fazla ağırlık verilirken, daha büyük varyanslı Aralık ve Ocak ayları gibi birimlere ise daha az ağırlık verilmektedir. Burada böyle bir varsayımda bulunmak makul görülmektedir. Çünkü Aralık ve Ocak ayları gibi birim değerlerinin tipik Aralık ve Ocak ayı değerlerinden daha çok farklılık gösterirken, Temmuz ve Ağustos ayları gibi birim değerlerinin tipik Temmuz ve Ağustos değerlerinde daha az farklılık göstermektedir.

Şekil 5 aylara göre hata varyanslarının nasıl değiştiğini göstermektedir. AR analizinde birim değerlerine varyanslarıyla ters orantılı ağırlıklar verilmektedir (Norusis ve SPSS Inc, 1999). AR analizini kullanabilmek için, her bir birim için hangi büyüklükte bir hata beklendiğini gösteren bir serinin (ağırlık değişkeninin) hesaplanması gerekmektedir. Bu değişkeni uygun bir şekilde hesaplayabilmek için ilk olarak her aya ait birim değerlerinin varyansın hesaplanması gerekmektedir. Bu varyanslar regresyon analizinde kullanılacak en uygun

ağırlık değişkeninin birim değerlerinin hesaplanmasında kullanılmaktadır.

**Tablo 7: Kukla Değişkenli Ağırlıklı Regresyon Analizinin Sonuçları**

	SABİT	TREND	OCK	ŞUB	MRT	NİS	MAY	HAZ	AGT	EYL	EKM	KAS	ARL
B	32.12	-1.47	146.8	122.3	97.1	55.7	20.7	5.9	1.7	7.5	40.9	113.5	144.9
SE	1.71	0.12	6.37	6.32	6.33	3.20	2.20	1.79	2.08	2.06	3.82	9.58	6.62
Anl.	.000	.000	.000	.000	.000	.000	.000	.001	.411	.000	.000	.000	.000
R=%95.7 Düzeltilmiş-R <sup>2</sup> =%91.5 F=176.64 (0.000) Tahminin Std. Hatası=0.088 DW=1.91													

AR analizinin sonuçları Tablo 7’de gösterilmektedir. Modelin düzeltilmiş çoklu belirlilik katsayısı %91.5 olarak elde edilmiştir. Tahmin edilen *TREND* değişkeninin katsayısı her bir yıl için ortalama -1.47 um/m<sup>3</sup> ve 0.000 düzeyinde anlamlıdır. *TREND* değişkeninin katsayı büyüklüğü önceki analizlerden önemli düzeyde daha düşük elde edilmiştir. Tahmin döneminin mevsimlik etkilerden arındırılmış başlangıç değerinin tahmin edilen değerini gösteren sabit terim katsayısı 32.12 um/m<sup>3</sup> olarak hesaplanmıştır.

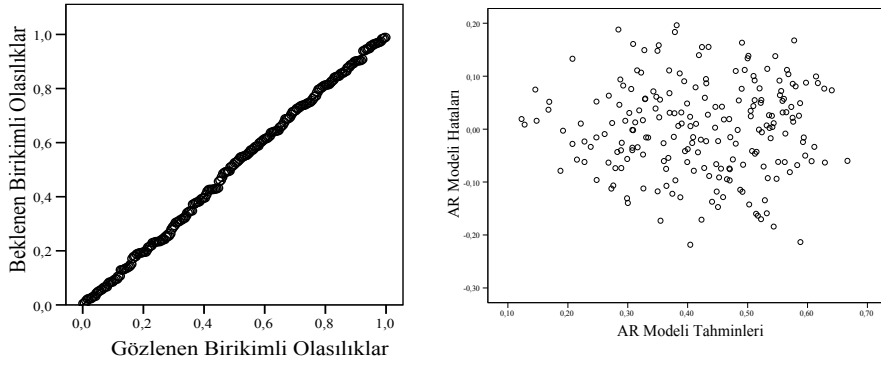
Geliştirilen modellerde birinci dereceden otokorelasyon ve çoklu doğrusal bağlantı sorunu bulunmamaktadır. Geliştirilmiş modellere ilişkin Durbin-Watson (DW) istatistikleri sırasıyla 1.98, 1.84, 1.92 ve 1.91 olarak elde edilmiştir. Çoklu doğrusal bağlantı testleri için varyans şişirme faktörleri (VIF) hesaplanmış ve tüm VIF değerlerinin problemlili kritik değer olan 10’dan çok daha küçük (hepsi 2’den küçük) olarak elde edilmişlerdir (Gujarati, 1995).

Regresyon tahminleri tekrar değişerek negatif, daha küçük ve istatistik açıdan anlamlı *TREND* değişkeni katsayısı elde edilmiştir. Sonuç olarak, tüm birim değerlerine eşit ağırlıklar veren EKK regresyon analiziyle daha az güvenilir regresyon katsayıları elde edilmiştir. AR analiziyle elde edilen sonuçlar daha durağan ve güvenilirdir. Sonuç olarak, Zonguldak’ta *TAP* düzeylerinin 17.5 yıllık bir dönemde yılda ortalama olarak -1.47 um/m<sup>3</sup> olarak azaldığı ve bu azalışın istatistik olarak oldukça anlamlı olduğu anlaşılmaktadır.

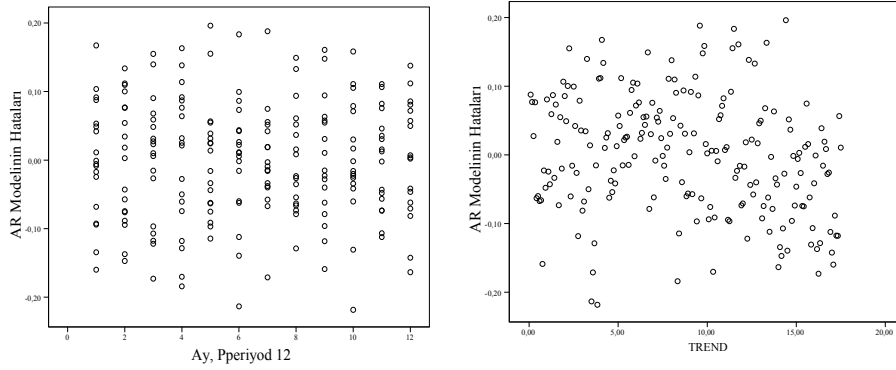
AR analizine ait normal olasılık (P-P) grafiği Şekil 6’da gösterilmektedir. AR analizine ait hataların grafiği, Şekil 3’te gösterilen EKK regresyon analizine ait hataların grafiğinden çok daha iyi du-

rumdadır. Sonuç olarak, AR analizinin tahminleri ile hataları arasında daha önce gözlenen değişen varyans sorunu gözlenmemektedir.

**Şekil 6: Normal Olasılık Grafiği ve Tahminler ile AR Modelinin Hataları**



**Şekil 7: Aylara ve Trende Göre Ağırlıklı Regresyon Modelinin Hataları**



Ayrıca modelin uygunluğunun değerlendirilmesinde önemli bağımsız değişkenlerle (*TREND*) hatalar arasındaki serpilme diyagramlarının kontrol edilmesi yararlı olmaktadır. Şekil 7’de AR modelinin tahmin edilen hataları ile temelde ilgilendiğimiz *TREND* değişkeni arasındaki serpilme diyagramı gösterilmektedir. Grafikte iki değişken arasında sistematik bir ilişki gözlenmemektedir. Ayrıca Şekil 7’de AR modelinin hatalarının aylara göre dağılımı verilmektedir. Grafikte hatalarda gözlenen kum saatine benzeyen dağılım şekli (değişen varyans sorunu) gözlenmemektedir.

#### 4. SONUÇ

Özellikle  $50 \text{ um/m}^3$  günlük ortalama hedef kalite değeri dikkate alındığında, Zonguldak'ta, *TAP* düzeylerinin Şubat ( $142 \text{ um/m}^3$ ), Mart ( $116 \text{ um/m}^3$ ), Nisan ( $74 \text{ um/m}^3$ ), Ekim ( $60 \text{ um/m}^3$ ), Kasım ( $133 \text{ um/m}^3$ ) ve Aralık ( $164 \text{ um/m}^3$ ) aylarında hedef kalite standartlarını sağlayamadığı; Mayıs ( $40 \text{ um/m}^3$ ), Haziran ( $25 \text{ um/m}^3$ ), Temmuz ( $20 \text{ um/m}^3$ ), Ağustos ( $21 \text{ um/m}^3$ ) ve Eylül ( $27 \text{ um/m}^3$ ) aylarında hedef kalite standartlarını sağladığı anlaşılmaktadır.

Bu çalışma, Zonguldak'ta *TAP* serisi için en uygun trend tahmin modelinin ağırlıklı regresyon (AR) analiziyle elde edilebileceğini göstermektedir. AR analizinin sonuçları *TREND* değişkeni katsayısının negatif, daha küçük ve istatistik olarak anlamlı olduğunu göstermektedir.

Geliştirilen modellerde birinci dereceden otokorelasyon ve çoklu doğrusal bağlantı sorunu bulunmamaktadır. Geliştirilen modellere ilişkin DW istatistikleri sırasıyla 1.97, 1.97 ve 1.94'tür. Modeldeki bağımsız değişkenlere ilişkin tüm VIF değerleri kritik değer olan  $10^3$ 'ün oldukça altında (2'den küçük) kalmaktadır.

AR analizinin normal olasılık grafikleri EKK regresyon analizinin grafiklerinden çok daha iyi bir düzeydedir. Son olarak, AR modelinin hataları ile tahmin edilen değerler arasında EKK tekniğiyle geliştirilen modellerde görülen değişen varyans sorunu gözlenmemiştir. Özellikle regresyon katsayısı tahmin edilmeye çalışılan *TREND* değişkeni ile AR modelinin hataları arasında sistematik bir ilişki gözlenmemiştir.

Sonuç olarak, Zonguldak'ta *TAP* düzeylerinin 17.5 yıllık bir dönemde yılda ortalama olarak  $-1.47 \text{ um/m}^3$  olarak azaldığı ve bu azalışın istatistik olarak oldukça anlamlı olduğu anlaşılmaktadır.

## KAYNAKÇA

- Albayrak, A. S. (2006). *Uygulamalı Çok Değişkenli İstatistik Teknikleri*, Ankara: Asil Yayın Dağıtım Ltd. Şti.
- Armitage, P. (1971); *Statistical Methods in Medical Research*, Blackwell Publications: Oxford.
- Draper, N. R. ve H. Smith (1981); *Applied Regression Analysis*, Willey: New York.
- Green, W. H. (1994). *Econometric Analysis*, 2<sup>nd</sup> Ed., Prentice Hall: N J.
- Gujarati, D. N. (1995). *Basic Econometrics* (Third Edition), McGraw-Hill: New Jersey.
- Hintze, J. L. ve NCSS (2007). *User's Guide-IV: Multivariate Analysis, Clustering, Meta Analysis, Forecasting, Time Series, Operation Research and Mass Appraisal*. Kaysville, Utah: NCSS Inc.
- Netter, J., W. Wasserman ve M. H. Kunter (1983). *Applied Linear Regression Models*, Illinois.
- Norusis, M. J. ve SPSS Inc. (1999). *SPSS Regression Models 10.0*, Chicago: SPSS Inc.
- Orhunbilge, N. (2000). *Uygulamalı Regresyon ve Korelasyon Analizi*, İstanbul: Avcıol Basım Yayın.
- Orhunbilge, N. (1998). *Zaman Serileri Analizi ve Tahmin Yöntemleri*, Avcıol-Basım Yayın: İstanbul.
- Pindyck, R. S. ve Rubinfeld, D. L. (1991). *Econometric Model and Econometric Forecast*, 3<sup>rd</sup> Ed.
- Shin, K. (1996). *SPSS Guide for DOS Version 5.0 and Windows Versions 6.0 and 6.1.2*, Chicago: Irwin.
- SPSS Inc. (1999). *SPSS Trends 10*, Chicago: SPSS Inc.
- Tarı, R. (2006). *Ekonometri* (4. Baskı), İstanbul: Avcı Ofset.
- Webster, A. (1995). *Applied Statistics for Business and Economics*, 3<sup>rd</sup> Ed.
- Yamak, R., ve Köseoğlu, M. (2006). *Uygulamalı İstatistik ve Ekonometri* (3. Baskı), Celepler Matbaacılık: Trabzon.