

KELEBEK ŞEKLİNDE DAĞILAN ARTIKLAR için GENEL VARYANS DENGELEME YÖNTEMLERİNİN KARŞILAŞTIRILMASI

Reşit ÇELİK*

ÖZET

Basit ve çoklu regresyon yöntemlerinde, bağımlı değişken Y ' nin koşullu dağılımlarının aynı standart sapmaya sahip olduğunu varsayılır. Regresyonda regresyon katsayılarının anlamlılığı ve güven aralıkları bu varsayıma bağlıdır. Regresyon katsayılarının testi için jacknife ve bootstrap gibi yöntemler kullanılsa bile değişen varyanslılığın söz konusu olması halinde sonuçlar çok hatalı olabilir.

İstatistik literatüründe var olan değişen varyans kalıpları genel olarak monoton azalan ya da monoton artan şeklindedir. Buna koşut olarak varyans dengeleme teknikleri de monoton artan ya da azalan varyansı dengelemeye yöneliktir.

Bu çalışmada Sivas ili için aylık ortalama havada asılı partiküller madde miktarı, sıcaklık ve kükürt dioksit için modellenerek kelebek dağılan artıklarda değişen varyanslılığın dengelenmesi konusunda istatistik literatüründe mevcut olan temel yöntemler incelenerek karşılaştırılmış ve sonuçlar irdelenmiştir.

Anahtar Sözcükler: Kelebek dağılan artıklar, değişen varyanslılık, ağırlıklandırma, dönüşümler.

COMPARING GENERAL VARIANCE STABILIZING METHODS FOR BUTTERFLY DISTRIBUTED RESIDUALS

ABSTRACT

In Simple and multiple regression methods, it is assumed that the conditional distributions of the dependent variable Y have the same standard deviation. The significance of regression coefficients, and confidence intervals based on this assumption. Even Jacknife and bootstrap methods are used to test for regression coefficients when there is heteroscedasticity results may be inaccurate.

* Arş.Gör.Dr., Yıldız Teknik Üniversitesi, Fen - Edebiyat Fakültesi, İstatistik Bölümü, İstanbul, rcelik @yildiz.edu.tr

Existing literature on statistical patterns of heteroscedasticity in general, is monotonically increasing or monotonically decreasing. Correspondingly monotonically increasing or decreasing so that the stabilize techniques of the variance are to stabilize monotonic heteroscedasticity.

In this study, the monthly average amount of suspended particulate matter, sulfur dioxide, temperature of the Sivas city are modeled residuals for butterfly distributed errors and then basic methods of stabilizing heteroscedasticity available literature on the statistical results were examined, compared and discussed.

Key Words: *Butterfly distributed residuals, heteroscedasticity, weighting, transformations*

1. GİRİŞ

Klasik doğrusal regresyon analizinde, varsayımlar yerine gelmediğinde ağırlıklandırma ve dönüşümler kullanılarak varsayım bozulmaları düzeltilmeye çalışılır. Değişen varyanslılık bu varsayım bozulmalarından biridir ve kelebek şeklinde dağılan artıklar da değişen varyanslılığın özel bir halidir.

Bu çalışmada, değişen varyanslılık, nedenleri, saptanması ve sonuçları üzerinde durulduktan sonra kelebek şeklinde dağılan artıklar için değişen varyanslılık sorunu ve düzeltilmesi konusunda istatistik literatüründe bulunan bazı yöntemler sınanmıştır.

2. DEĞİŞEN VARYANSLILIK, NEDENLERİ, SAPTANMASI VE SONUÇLARI

Hata terimlerine ilişkin varyansların, bağımsız değişkenin ya da ortalamanın çeşitli düzeyleri için istatistiksel olarak farklı olması değişen varyanslılık olarak adlandırılır. Değişen varyanslı regresyon modellerinde hata terimlerine ilişkin varyanslar açıklayıcı değişkenlerin bir ya da bir kısmının fonksiyonudur (Harvey, 1976).

Değişen Varyanslılığın Nedenleri: Bağımsız değişken değerleri değiştikçe, $V(y_i)$ 'nin de değişmesi sonucu değişen varyanslılık ortaya çıkabilir.

Değişen varyanslılığın bir diğer nedeni modelleme hatasıdır. Değişen varyanslılık modelin fonksiyonel yapısından kaynaklanabilir (Sümer, 2006). Farklı bir model tahmini ya da bazı önemli değişkenlerin model dışı bırakılması değişen varyanslılığın kaynağı olabilir (Gujaratti, 2001).

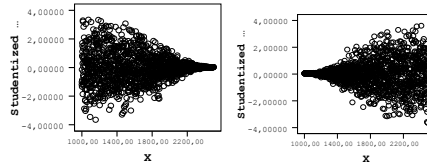
Değişen varyanslılık, aykırı değerlerin örneklemedeki öteki gözlemlerden, çok küçük ya da çok büyük olması sonucunda da ortaya çıkabilir. Bu tip gözlemlerin atılması özellikle örneklem büyüklüğü küçükse regresyon çözümlerinin bulgularını önemli ölçüde değiştirebilir.

Veri derleme teknikleri ya da ölçme araçları geliştikçe varyans küçülebilir. Oberg ve Davidian'a (2000) göre, normal dağılım ve eş varyanslılık varsayımlarının bozulması daha çok tekrarlı ölçümlü verilerde görülmektedir.

Kişilerin herhangi bir konudaki deneyimleri arttıkça davranış hataları da zamanla azalmakta dolayısıyla varyans da küçülmektedir (Gujaratti, 2001).

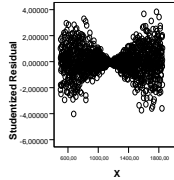
Değişen Varyanslılığın Saptanması: Değişen varyanslılığın saptanması ve modellenmesi konusunda kullanılan pek çok grafiksel yöntem vardır. Değişen varyanslılığın saptanması konusunda kullanılan en yaygın araç, ağırlıklı olmayan en küçük kareler ham ya da Student-türü artıkların dikey ekseninde yer aldığı çizimlerdir

Değişen varyanslılıkta en çok karşılaşılan durum artık saçılım görüntüsünün megafon biçiminde olduğu durumdur (Şekil 1)



Şekil 1. Sağa ve Sola megafon

Megafon biçimindeki iki kalıba ek olarak kelebek ve galaksi şeklinde değişen varyanslılık kalıplarına da rastlanmaktadır. Kelebek şeklindeki bir artık saçılım grafiğinin iki ucu geniş, ortası ise dardır (Şekil 2). Galaksi ise bunun tersine orta kısım geniş ve uçlar dardır (Darlington, 1997).



Şekil 2. Kelebek şeklinde dağılan artıklar

Kelebek şeklinde dağılan artıklar, regresyon modeli ilgi değişkeni ile aralarında etkileşim olan bazı değişkenler tamamen yok sayıldığında ortaya çıkar (Darlington, 1997).

Meteorolojik verilerin aylara göre alt gruplar şeklinde sıralanarak modellenmesi halinde kelebek şeklinde dağılan artıklar ortaya çıkar.

Değişen varyanslılığın belirlenmesi konusunda özellikle ekonometri ve zaman serileri literatüründe geçen pek çok skor testi vardır. White Testi, bu testlerden biridir. Test, eş varyanslılık varsayımının geçerli olup olmadığına incelenmesinde en çok kullanılan testlerden biridir. Normallik varsayımına dayanmayan yardımcı bir regresyon modelinin tahminini gerektirir. Model belirlenerek, en küçük kareler tahminlerinden artıklar bulunur. Bulunan artıkların karelerinin bağımlı değişken, bağımsız değişkenlerin, kendilerinin, karelerinin ve çarpımlarının bağımsız değişken olduğu bir yardımcı regresyon modeli kurulur:

$$e_i^2 = \alpha_0 + \alpha_1 x_{i1} + \dots + \alpha_p x_{ip} + \gamma_1 x_{i1}^2 + \dots + \gamma_p x_{ip}^2 + \dots + \theta_1 x_{i1} x_{i2} + \dots + \theta_p x_{ip-1} x_{ip}$$

$$H_0 : \forall(\alpha_i, \gamma_j, \theta_l) = 0 \text{ (Eş Varyanslılık)}$$

$$H_1 : \exists(\alpha_i, \gamma_j, \theta_l) \neq 0 \text{ (Değişen Varyanslılık)}$$

k , yardımcı regresyon modelindeki parametre sayısı olmak üzere, yardımcı regresyon modelinin belirtme katsayısı n ile çarpılarak

$$nR_{yard.}^2 \approx \chi_{k, \alpha}^2$$

test istatistiği hesaplanır. $nR_{yard.}^2$ İstatistiğinin $\chi_{k, \alpha}^2$ değerinden küçük olması eş varyanslılık hipotezini; büyük olması ise değişen varyanslılığı öne süren hipotezi doğrulayacaktır (Güriş ve Çağlayan, 2001).

Değişen Varyanslılığın Sonuçları: Değişen varyanslılık durumunda $\hat{\beta}$ kendi parametresinin yansız bir tahminidir, ancak standart hatası ($SE_{\hat{\beta}}$) çok büyük ya da çok küçük olacağından etkinliğini yitirecektir. Eğer $SE_{\hat{\beta}}$ olduğundan çok daha büyük tahmin edilirse, $\hat{\beta}$ kendi parametresinin etkin bir tahmini olmaz ve $t = \hat{\beta} / SE_{\hat{\beta}}$ değeri küçüleceğinden doğru olan alternatif hipotezin reddi ve ya yanlış olan yokluk hipotezinin kabul olasılığı olan II. Tip hata ortaya çıkar. Eğer $SE_{\hat{\beta}}$ olduğundan çok daha küçük tahmin edilirse, $\hat{\beta}$ kendi parametresinin etkin bir tahmini olmaz ve $t = \hat{\beta} / SE_{\hat{\beta}}$

değeri büyüyeceğinden doğru olan bir temel hipotezin reddi veya yanlış olan alternatif hipotezin olasılığı olan I. Tip hata ortaya çıkar (Friel, 2009).

Megafon biçimindeki şekiller “kelebek” veya “galaksi” biçimindeki diğer iki kalıp kadar istatistiksel çıkarsamayı bozmaz (Darlington, 97).

3. DEĞİŞEN VARYANSLILIĞIN DENGELENMESİNDE KULLANILAN BAZI YÖNTEMLER

Bu bölümde değişen varyanslılık sorununun giderilmesinde kullanılan bazı yöntemlere kısaca değinilmiştir.

3.1. Yeniden Ağırlıklandırılmış İteratif En Küçük Kareler

Meloun, Militky, Kupka, ve Brereton (2002) tarafından önerilen yöntemin aşamaları aşağıda verilmiştir.

- En küçük regresyondan elde edilen artıklarının medyanı bulunur.

$$\hat{\sigma} = \frac{\text{medyan}|\hat{e}_i - \text{medyan}(\hat{e}_i)|}{0.6745} \quad (2)$$

veya $\hat{\sigma} = 2.1 \text{ med}|\hat{e}_i|$ alınır. Burada 0.6745 değeri, hataları normal dağılıma sahip büyük örnekler için $\hat{\sigma}$ 'yı artık standart sapmasına eşit kılma sabitidir.

- Ağırlıklar (l) iterasyon sayısını göstermek üzere $w_i^{(l)} = \text{diag}(r_i^{(l)} / \hat{\sigma}_i)$ alınarak $\hat{\beta}^{(l)} = (X'W^{(l)}X)^{-1}X'W^{(l)}Y$ (3) eşitliği ile ağırlıklı en küçük kareler regresyonu uygulanır.
- Ardışık iki parametre tahmin değeri farkı $(\hat{\beta}^{(l)} - \hat{\beta}^{(l-1)})$ istenen düzeyde değilse son adımda elde edilen parametre tahminleriyle bir önceki adımdaki dönülerek parametre tahminleri güncellenir

3.2. Chatterjee ve Machler (1997) İteratif Yöntemi

Bu yöntemde, sıradan en küçük kareler tahminlerine başvurulmaksızın doğrudan iteratif ağırlıklı yöntemlere başlanır.

- İlk adımda, $H = X(X'X)^{-1}X'$ matrisinin asal köşegeni h_{ii} i'nci kaldıraç noktası değeri ve parametre sayısının gözlem sayısına oranı

$\tilde{h} = k / n$ olmak üzere, $w_i^0 = 1 / \max(h_i, \tilde{h})$ başlangıç ağırlık ağırlık tahminleri kullanılarak $\hat{\beta}^{(0)} = (X'W^{(0)}X)^{-1}X'W^{(0)}Y$ ile iteratif sürece başlanır.

- İkinci ve daha sonraki adımlar için en son regresyondan elde edilen artıklar kullanılarak yeni ağırlıklar hesaplanır:

$$r^{j-1} = y - \hat{y}^{j-1} = y - X\hat{\beta}^{j-1} \quad (4)$$

$$w_i^j = (1 - h_i)^2 / \max\{|r_i^{j-1}|, \text{medyan}_i |r_i^{j-1}|\} \quad (5)$$

Bu işlem istenen yakınsama gerçekleşinceye kadar tekrarlanır. Yakınsama garanti değildir

3.3. Değiştirilmiş Ardışık Artıklar Yöntemi

Bischoff, Heck, Howind, Teusch (2006) tarafından önerilen yöntemde, ilk k gözlem (parametre sayısı) kullanılarak (k+1)'inci artık tahmin edilir ve her adımda bir gözlem eklenerek bu işlem (n-k) defa tekrarlanarak (n-k) sayıda bağımsız artık elde edilir.

$$\hat{\beta}_k = (X_k'X_k)^{-1}X_k'Y_k \rightarrow \hat{Y}_{k+1} = x_{k+1}'\hat{\beta}_k \quad (6)$$

$$r_{k+1} = \frac{Y_{k+1} - \hat{Y}_{k+1}}{1 + x_{k+1}'(X_k'X_k)^{-1}x_{k+1}} \quad (7)$$

Hatalar normal dağılımlı varsayıldığından artık vektörü r de '0' ortalamayla normal dağılıma sahip olur. Bundan dolayı sifıra çok yakın kare artıklar için $\ln(r^2)$ göreceli olarak çok büyük negatif değerler alır. Bu problemten kaçınmak için alternatif Box-Cox dönüşümü kullanılabilir (Bischoff vd. 2006). T(.) dönüşüm göstermek üzere, değiştirilmiş ardışık artıklar,

$T(z) = \ln(z + M)$, $M = (M, \dots, M)^T \in R^n$; $z = r^2$ şeklinde elde edilir. Burada

$$M = (\text{medyan}(r^2))^2 / \max(r^2) \quad (8)$$

dir.

3.4. Logaritmik Dönüşüm

Box ve Cox (1964) tarafından önerilen ve varyansın bağımlı değişkenin artan değerlerine paralel olarak arttığı durumlarda kullanılan bu dönüşüm pozitif 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 ise modeli doğrusallaştırmak için kullanılmaktadır (Albayrak, 2008. Hamasaki ve Goto, 2005).

3.5. Tekrarlı Gözlemler için Ortalama-Varyans Fonksiyonları

Pek çok deney probleminde deney bilimciler öngörü değişkenlerinin her bir değerinde bağımlı olanı yeniden elde eder. Öngörü değişkenleri (x_1, \dots, x_n) ve her bir x_i değerinde m_i adet bağımlı değeri gözlenir. Bu durum için ortalamalar ve varyanslar:

$$E(y_{i,j}) = \mu_i = \mu_i(\beta), \quad st.sapma(y_i) = \sigma_i$$

dir. Tahmini ağırlıkların öngörü değişkenlerinin her bir düzeyindeki örnek varyanslarının tersi olduğu ağırlıklar kullanılarak β en küçük karelerle tahmin edilmeye çalışılır. Burada

$$\hat{w}_i = 1/s_i^2; \quad s_i = (m_i - 1)^{-1} \sum_{j=1}^{m_i} (y_{i,j} - \bar{y}_i)^2 \quad (9)$$

$$\hat{\beta} = (X'WX)^{-1} X'WY \quad (10)$$

dir. \bar{y}_i , x_i öngörü değerindeki örnek ortalaması ve m_i i. kitledeki gözlem sayısıdır (Carroll ve Ruppert, 1988).

4. UYGULAMA

Hava Kalitesi İndeksi (HKİ): Hava kalitesi, belli kirletici konsantrasyonlarını kaydeden ölçüm cihazlarından oluşan bir ağ yardımı ile ölçülür. Bu ham ölçüm değerleri, geliştirilen standart formüller kullanılarak HKİ değerlerine dönüştürülmektedir. HKİ değeri, bölgedeki her bir kirletici

için ayrı ayrı hesaplanır (yer seviyesindeki ozon, partiküller, karbon monoksit, kükürt dioksit ve azot dioksit)¹.

Partiküller madde (PM) terimi, havada bulunan katı partiküller ve sıvı damlacıkları ifade eder. İnsan faaliyetleri sonucu ve doğal kaynaklardan, doğrudan atmosfere karışır. Atmosferde diğer kirleticiler ile reaksiyona girerek PM oluşturur ve atmosfere verilirler².

Uygulama için Sivas iline ilişkin aylık ortalama partiküller madde, kükürt ve sıcaklık ölçüm değerleri kullanılmıştır. Analizler için SAS 9.2 ve SPSS 15.0 programları kullanılmıştır.

4.1. Denklem Kestirimi ve Artıklar

Başlangıç modelinde partiküller madde miktarı bağımlı değişken; taşıt sayısı, sıcaklık, doğal gaz kullanımı ve kükürt miktarı bağımsız değişken alınarak modellenmiştir.

$PM = \beta_0 + \beta_1 * Sıcaklık + \beta_2 * Kükürt + \beta_3 * Taşıt + \beta_4 * dgaz + \varepsilon$ Sıcaklığa ilişkin ortalama aylık veriler devlet meteoroloji müdürlüğünden, kükürt miktarı, taşıt sayısı ve partiküller maddeye ilişkin aylık ortalama veriler ise TÜİK kayıtlarında elde edilmiştir. Taşıt sayısına ilişkin TÜİK kayıtlarının başlangıç yılı 1994 Temmuz; diğer değişkenlere ilişkin kayıtların başlangıcı ise 1990 yılıdır.

Tablo 1. Model Seçimi

Variable	DF	Estimate	s. e.	t Value	Pr > t
Intercept	1	66.26640	8.839	7.50	<.0001
Sıcaklık	1	-2.713	0.305	-8.89	<.0001
Kükürt	1	0.299	0.05462	0.47	<.0001
Taşıt Sayısı	1	0.0002	0.00012476	1.60	0.1105
Doğal gaz	1	0.41	4.58	0.09	0.9283

Seçilen başlangıç modelinde taşıt sayısı ve doğal gaz kullanımı istatistiksel olarak anlamlı çıkmadığından analizin bundan sonraki süreci için bu iki değişken analize alınmamıştır. Bundan sonraki analizler,

$Partiküller_Madde = \beta_0 + \beta_1 * Sıcaklık + \beta_2 * Kükürt + \varepsilon$ modeli için yapılmıştır.

¹ <http://www.havaizleme.com/r.html>

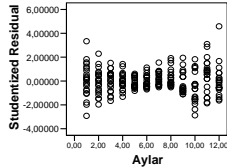
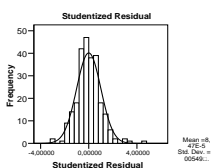
² <http://www.rshm.saglik.gov.tr/hki/hkindex/index5.htm>

4.2. En Küçük Kareler

Partiküller maddenin bağımlı, sıcaklık ve kükürt ölçümlerinin bağımsız değişken olduğu en küçük kareler analiz sonuçları Tablo 2'de verilmiştir.

Tablo 2. En küçük kareler sonuçları

Analysis of Variance						
Source	DF	SS	MS	F Value	Pr > F	
Model	2	831850	415925	418.40	<.0001	
Error	250	248518	994.07317			
Corrected Total	252	1080369				
Dependent Mean	83.43083	Adj R-Sq	0.7681			
Coeff Var	37.79050					
Parameter Estimates						
Variable	DF	Estimate	s. e.	t	Pr > t	V.I.F
Intercept	1	70.128	4.45	15.77	<.0001	0
Sıcaklık	1	-2.407	0.27	-8.79	<.0001	1.43
kükürt	1	0.405	0.022	18.2	<.0001	1.43
AIC	BIC	SBC				
1749.1402	1751.2119	1759.74038				
Test of First and Second Moment Specification						
DF	Chi-Square	Pr > ChiSq				
5	14.33	0.0136				
Durbin-Watson D						
1st Order Autocorrelation	1.249	0.359				

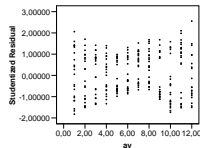
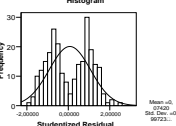
Regresyon çıktısı incelendiğinde kelebek şeklinde dağılan artıklar dolayısıyla değişen varyanslılık sorunu dışında herhangi bir varsayım bozulması yoktur.

4.3. Yeniden Ağırlıklandırılmış İteratif EKK Sonuçları

Kesim 3.1'de verilen algoritma uygulanmış ve Tablo 3'de verilen sonuçlara ulaşılmıştır.

Tablo 3. İteratif ağırlıklandırma sonuçları

Analysis of Variance						
Source	DF	SS	MS	F	Pr > F	
Model	2	2.5E16	1.2E16	1.45E13	<.0001	
Error	250	221905	887.62096			
Corrected Total	252	2.57687E16				
Dependent Mean	184.80360	Adj R-Sq	1.0			
Coeff Var	16.12142					
Parameter Estimates						
Variable	DF	Estimate	s. e.	t	Pr > t	VIF
Intercept1	74.58	6.454E-10	1155717	<.0001	0	
Sıcaklık 1	-2.9	1.0482E-6	-27728	<.0001	191	
Kükürt 1	0.36812	1.02E-3	61768	<.0001	191	
AIC	BIC	SBC				
1720.4783	1722.5500	1731.07843				
Test of First and Second Moment Specification						
DF	Chi-Square	Pr > ChiSq				
3	3.04	0.3851				
Durbin-Watson D						
1st Order Autocorrelation	1.023	0.483				

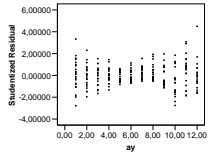
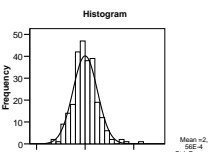
38 iterasyon sonunda elde edilen nihai ağırlıklar kullanılarak ulaşılan regresyon sonuçları incelendiğinde, çok yüksek çoklu belirtme katsayısı (1) elde edilmesine rağmen söz konusu yöntemle değişen varyanslılık dengelenememiş ve bununla birlikte çoklu bağlantı ve normal dağılımdan sapma durumu ortaya çıkmıştır.

4.4. Chatterjee ve Machler (1997) İteratif Yöntemi Sonuçları

Yöntem SAS programında yazılan kodlarla uygulanmış ve Tablo 4’de verilen sonuçlara ulaşılmıştır.

Tablo 4. Chatterjee ve Machler iteratif yöntemi sonuçları

Analysis of Variance						
Source	DF	SS	MS	F Value	Pr > F	
Model	2	5452	2726	403.66	<.0001	
Error	250	1688	6.75404			
Corrected Total		252	7141.24794			
Dependent Mean		82.40007	Adj R-Sq	0.7617		
Coeff Var		3.15395				
Parameter Estimates						
Variable	DF	Estimate	s.e.	t	Pr > t	VIF
Intercept	1	70.04	4.46	15.7	<.0001	0
Sıcaklık	1	-2.4	0.276	-8.78	<.0001	1.4
Kükürt	1	0.405	0.0227	17.78	<.0001	1.4
AIC	BIC	SBC				
486.2477	488.3194	496.84788				
Test of First and Second Moment Specification						
DF	Chi-Square	Pr > ChiSq				
6	17.22	0.0085				
Durbin-Watson D	1.242					
1st Order Autocorrelation	0.364					

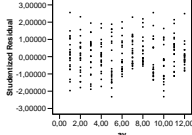
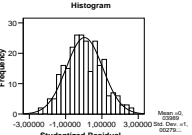
Yakınsama 3 iterasyon sonucunda gerçekleşmiş ancak değişen varyanslılığın düzelmediği, son iterasyondan elde edilen ağırlıklı en küçük kareler Student-türü artıkların aylara karşı çiziminden anlaşılmaktadır.

4.5. Değiştirilmiş Ardışık Artıklar

Yöntemin Bölüm 3.3’de verilen algoritmasına paralel olarak elde edilen değiştirilmiş ardışık artıkların karelerinin tersi ağırlık tahminleri elde edilmesinde kullanılarak ağırlıklı en küçük kareler uygulanmış ve Tablo 5’de verilen sonuçlara ulaşılmıştır.

Tablo 5. Değiştirilmiş ardışık artıklar yöntemi sonuçları

Analysis of Variance						
Source	DF	SS	MS	F	Pr > F	
Model	2	557	278	343.98	<.0001	
Error	248	202	0.81021			
Corrected Total		252	759.95122			
Dependent Mean		57.63	Adj R-Sq	0.7313		
Coeff Var		1.56197				
Parameter Estimates						
Variable	DF	Estimate	s.e.	t	Pr > t	VIF
Intercept	1	70.36	4.5	5.63	<.0001	0
Sıcaklık	1	-2.4	0.27	9.26	<.0001	1.47
Kükürt	1	0.38	0.026	14.95	<.0001	1.47
AIC BIC SBC						
		-50.2645	-48.1928	39.66435		
Test of First and Second Moment Specification						
DF	Chi-Square	Pr > ChiSq				
6	10.43	0.1078				
Durbin-Watson D 1.112						
1st Order Autocorrelation 0.436						

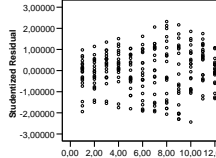
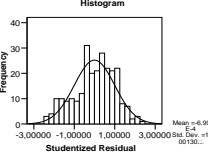
Ardışık artıkların ağırlıklıdırma kullanılması sonucunda değişen varyanslık sorunu giderilmiştir.

4.6. Logaritmik Dönüşüm Sonuçları

Bağımlı değişken değerlerinin logaritması alınarak regresyon analizi gerçekleştirilmiş ve Tablo 6'da verilen sonuçlar elde edilmiştir.

Tablo 6. Logaritmik dönüşüm sonuçları

Analysis of Variance						
Source	DF	SS	MS	F Pr > F		
Model	2	19.99	9.99225	.85 <.0001		
Error	250	11.06436	0.04426			
Corrected Total		252	31.05502			
Dependent Mean		1.79225	Adj R-Sq	0.6409		
Coeff Var		11.73799				
Parameter Estimates						
Variable	DF	Estimate	s.e.	t Pr > t VIF		
Intercept	1	70.35	4.5	15.63 <.0001	0	
Sıcaklık	1	-2.47	0.27	-9.26 <.0001	1.47	
Kükürt	1	0.38	0.026	14.95 <.0001	1.47	
AIC		-723.1765	BIC	-722.1345	SBC	-716.10972
Test of First and Second Moment Specification						
DF	Chi-Square	Pr > ChiSq				
5	37.21	<.0001				
Durbin-Watson D		0.794				
1st Order Autocorrelation		0.597				

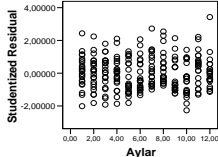
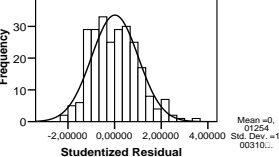
Skor testi ve Student-türü artık çizimleri incelendiğinde, istatistiksel olarak logaritma dönüşümüyle değişen varyanslılık sorununun düzelmediği, ek olarak otokorelasyon sorununun ortaya çıktığı gözlenmektedir.

4.7. Kare Artıklar

Aylara göre alt kitlelere bölünen veri kümesinde en küçük karelerden elde edilen artıkların karesi alınarak yeni bir değişken oluşturulmuştur. Oluşturulan değişken, SPSS paket programının “weight estimation procedure” modülü aracılığıyla $\hat{w}_j = 1 / s_y^2$, $j=1..12$; $i=1,..,n$; şeklinde ağırlık tahminleri elde etmekte kullanılarak Tablo 7’de verilen sonuçlar elde edilmiştir.

Tablo 7. Kare artıklar ile AEKK sonuçları

Analysis of Variance							
Source	DF	SS	MS	F	Pr > F		
Model	2	759	379	6370	<.0001		
Error	250	256	1.02				
Corrected Total		252	1015.32534				
Dependent Mean		58.16094	Adj R-Sq	0.7458			
Coeff Var		1.74015					
Parameter Estimates							
Variable	DF	Estimate	s. e.	t	Pr > t	VIF	
Intercept	1	67.33	4.55	14.80	<.0001	0	
Sıcaklık	1	-2.21	0.256	-8.65	<.0001	1.5	
Kükürt	1	0.41	0.026	15.9	<.0001	1.5	
AIC		9.0623	BIC	11.1341	SBC	19.66252	
Test of First and Second Moment Specification							
DF	Chi-Square	Pr > ChiSq					
6	4.52	0.6060					
Durbin-Watson D							
1st Order Autocorrelation		1.166			0.409		

Skor testi ve Student-türü artıkların aylara karşı çizimleri incelendiğinde, kare artıkların ağırlıklandırılmasında kullanılması sonucunda, kelebek şeklinde dağılan artık sorununun düzeldiği görülmektedir. Diğer istatistik ve kriterler incelendiğinde herhangi bir varsayım bozulması görülmemektedir.

5. YAPAY VERİLER İLE YÖNTEMLERİN SINANMASI

PBC Eletronics Designing'de (2009) yer verilen bir trigonometrik fonksiyon, amaca uygun biçimde geliştirilerek kullanılmıştır.

Bu bölümde, temel olarak kullanılan başlangıç algoritması aşağıda verilmiştir:

$j=0$;

do $i = (n/5)$ to $(n-n/5)$;

β_0 β_1

$j = j + 1$; $y[j] = 800 + 1 * i + 25 * \sin^2(2 * i * 3.14 / n) * \overbrace{\text{RAND('NORMAL')}}^{\varepsilon}$ $x[j] = j$;

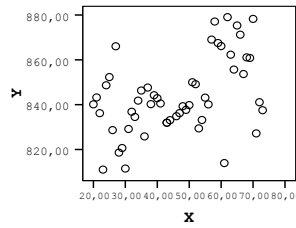
end;

Burada, $y[j]$ bağımlı değişkeni, $x[j]$ bağımsız değişkeni temsil etmektedir. Tekrarlı gözlemlere dayanan yöntemlerin uygulanabilmesi için $g = n / (1 + 3.3 \log(n))$ formülü yardımıyla alt gruplar oluşturulmuştur.

Bu kısıtlar altında, SAS 9.2 Programında yazılan kodlar yardımıyla farklı örneklem büyüklükleri için kelebek şeklinde dağılan artıkları veren yapay veri kümeleri oluşturularak, $Y = \beta_0 + \beta_1 X + \varepsilon$ şeklindeki basit doğrusal regresyon modeli için çalışmada yöntemler farklı büyüklükteki örneklem düzeylerinde kelebek şeklinde dağılan artıklar için incelenmiştir.

5.1. Örneklem Büyüklüğü $n = 53$

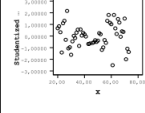
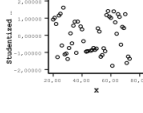
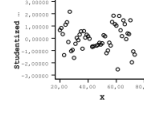
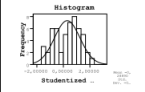
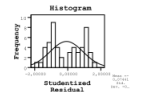
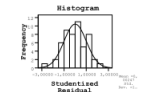
$n = 53$ için model çizimi Şekil 3'de verilmiştir.



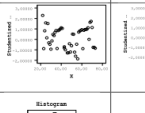
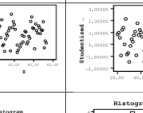
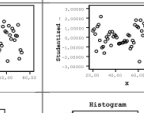

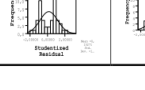
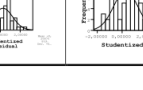
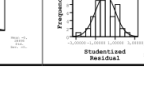

Şekil 3. $n = 53$ için model çizimi

Çizimi Şekil 3'de verilen veri kümesine en küçük kareler ve incelenen yöntemler uygulanmış ve sonuçları Tablo 8'de verilmiştir.

Tablo 8. n =53 için en küçük kareler ve incelenen yöntemlerin sonuçları

$\beta_0 = 800$ $\beta_1 = 1$	n = 53		
	En Küçük Kareler	Yeninden Ağırlıklandırılmış İteratif EKK İterasyon sayısı: 41	Chatterjee ve Mahler İteratif Yöntemi İterasyon sayısı 2
$\hat{\beta}_0; s_{\hat{\beta}_0}$	820.086 (6.5289)	817.687 (0.0009)	819.6333 (6.58013)
$\hat{\beta}_1; s_{\hat{\beta}_1}$	0.5143 (0.13289)	0.59327 (0.0000236)	0.523 (0.13412)
$t_{\hat{\beta}_0} \quad t_{\hat{\beta}_1}$	125.61; 3.87	889411; 25122.2	124.56; 3.90
$F; p$ Model	14.98; 0.0003	6.311E8; <.0001	15.21; 0.0003
RMSE	15.21363	16.78	2.373
R^2_{adj}	0.2119	1.0	0.229
Coeff Var	1.80251	1.99612	0.2812
AIC	290.5	300.9	93.57
DW-IAC	1.758; 0.1	1.534; 0.205	1.756; 0.12
White χ^2 :p	7.16; 0.10	2.12; 0.5486	10.07; 0.102
Artık Çizimi			
Normal Eğri ile Artık Sıklık Çizimi			

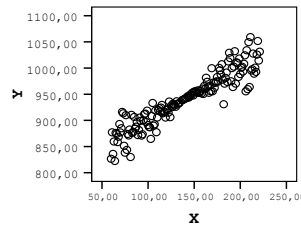
Tablo 8. (devam...)

$\beta_0 = 800 \quad \beta_1 = 1$	n = 53			
	Değiştirilmiş Ardışık Artıklar	Karesel Artık St. Sapmaları ile Ağırlıklandırma	Bağımlı Değişken St. Sapmaları ile Ağırlıklandırma	Box-Cox Log λ Dönüştürümü
$\hat{\beta}_0; s_{\hat{\beta}_0}$	809.73784 (4.29231)	819.18086 (6.54812)	816.75257 (6.41333)	2.91402 (0.00335)
$\hat{\beta}_1; s_{\hat{\beta}_1}$	0.65513 (0.08423)	0.40675 (0.14394)	0.45510 (0.13935)	0.00026324 (0.00008824)
$t_{\hat{\beta}_0} \quad t_{\hat{\beta}_1}$	188.65; 7.78	125.10; 2.83	127.35; 3.27	869.14; 3.86
$F; p$ Model	60.49 <.0001	7.99; 0.0067	10.67; 0.0020	14.88; 0.0003
RMSE	1.12834	0.28356	1.06746	0.00781
R^2_{adj}	0.5433	0.1354	0.1567	0.2107
Coeff Var	0.13398	0.03147	0.12747	0.26698
AIC	14.2762	-139.3852	9.8616	-512.3533
DW-IAC	1.179; 0.360	1.389; 0.298	3.47; 0.3244	1.784; 0.097
White χ^2 :p	5.09; 0.1657	5.18; 0.1589	1.410; 0.284	7.15; 0.0280
Artık Çizimi				
Normal Eğri ile Artık Sıklık Çizimi				

n = 53 örneklem düzeyi için incelenen yöntemlerin sonuçları irdelendiğinde, tekrarlı gözlemler için kare artıklarının ve bağımlı değişken standart sapmalarının ağırlık elde edilmesi, kelebek şeklinde dağılan artıklar için değişen varyanslılığın dengelenmesi konusunda istenen sonucu vermiştir. Bischoff ve diğ. (2006) tarafından önerilen değiştirilmiş ardışık

artıklar yöntemi, uygulama verisinin aksine burada iyi sonuç vermemiş ve artık dağılımında bozulmaya yol açmıştır. Diğer yöntemler istenen sonucu vermemiştir.

5.2. Örneklem Büyüklüğü n = 160



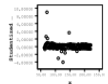
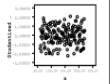
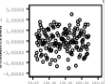
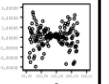
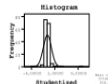
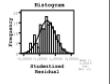
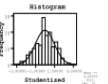
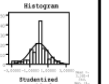
Şekil 4. n = 160 için model çizimi

Çizimi Şekil 4'de verilen veri kümesine incelenen yöntemler uygulanmış ve Tablo 9'da verilen sonuçlar elde edilmiştir.

Tablo 9. n = 160 için en küçük kareler ve incelenen yöntemlerin sonuçları

$\beta_0 = 800$ $\beta_1 = 1$	n = 160		
	En Küçük Kareler	Yeninden Ağırlıklandırılmış İteratif EKK İterasyon sayısı: 77	Chatterjee ve Mahler İteratif Yöntemi İterasyon sayısı: 2
$\hat{\beta}_0; s_{\hat{\beta}_0}$	797.5525 (4.65406)	800.78687 (0)	797.54577 (4.65738)
$\hat{\beta}_1; s_{\hat{\beta}_1}$	1.01721 (0.03140)	0.99813 (0)	1.01733 (0.03144)
$t_{\hat{\beta}_0}; t_{\hat{\beta}_1}$	171.37; 32.40	Infnty; Infnty	171.24; 32.36
F; p Model	1049.59; <.0001	Infnty <.0001	1047.35 <.0001
RMSE	1.98686	0	2.55367
R^2	0.8683	1	0.8681
Coeff Vaz	1.98686	0	0.27151
AIC	838.8942	889.1228	301.9877
DW-LAC	2.429; -0.225	2.388; -0.183	2.426; -0.223
White χ^2 :p	22.24; 0001	2.03; 0.5672	27.06; <.0001
Artık Çizimi			
Normal Eğri ile Artık Sıklık Çizimi			

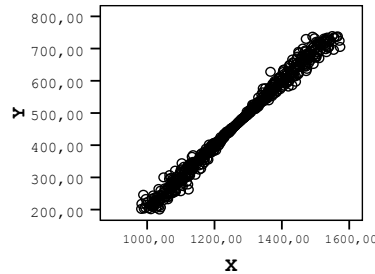
Tablo 9. (devam...)

$\beta_0 = 800 \quad \beta_1 = 1$	n = 160			
	Değiştirilmiş Ardışık Artıklar	Karesel Artık St. Sapmaları ile Ağırlıklandırma	Bağımlı Değişken St. Sapmaları ile Ağırlıklandırma	Box-Cox LogY Dönüşümü
$\hat{\beta}_0; s\hat{\beta}_0$	795.13222 (0.63953)	800.36757 (3.92622)	800.78167 (4.03302)	2.90653 (0.00217)
$\hat{\beta}_1; s\hat{\beta}_1$	1.03691 (0.00438)	1.00404 (0.02779)	0.99899 (0.02848)	0.00047094 (0.00001462)
$t_{\hat{\beta}_0} \quad t_{\hat{\beta}_1}$	1243.3; 236.54	203.85; 36.13	198.56; 35.06	1341.54; 32.22
$F; p$ Model	55952.1 <.0001	1305.28 <.0001	1230.61 <.0001	1038.15 <.0001
RMSE	2.07660	0.91135	0.92962	0.00870
R_{adj}^2	0.9972	0.8920	0.8855	0.8671
Coeff Var	0.21963	0.09683	0.09881	0.29264
AIC	232.898	-27.7184	-21.3661	-1516.2581
DW-IAC	1.760; 0.119	2.318; -0.165	2.308; -0.161	2.370; -0.198
White χ^2 :p	3.88; 0.2749	5.80; 0.1219	5.65; 0.1299	21.50; <.0001
Artık Çizimi				
Normal Eğri ile Artık Sıklık Çizimi				

n = 160 örneklem büyüklüğü için yöntemlerin sonuçları incelendiğinde, tekrarlı gözlemler için karesel artıkların ve bağımlı değişken standart sapmalarının ağırlık elde edilmesi, kelebek şeklinde dağılan artıklar için değişen varyanslılığın dengelenmesi konusunda istenen sonucu vermiştir. Bischoff ve diğ. (2006) tarafında önerilen yöntem de kelebek dağılan artıklar sorununu çözmüştür. Buna karşılık söz konusu yöntem aykırı değerlere ve artık dağılımında bozulmaya neden olmuştur.

5.3. Örneklem Büyüklüğü n = 533

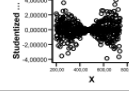
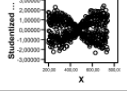
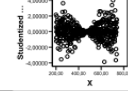
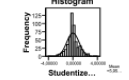
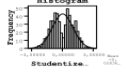
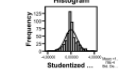
Bu örneklem büyüklüğü için model çizimi Şekil 5.'de verilmiştir.



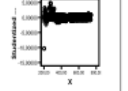
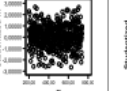
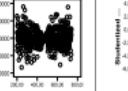
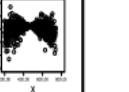
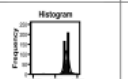

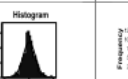

Şekil 5. n = 533 için model çizimi

Şekil 5’de model çizimi verilen veri kümesine incelenen yöntemler uygulanmış ve Tablo 10’da verilen sonuçlar elde edilmiştir.

Tablo 10. n = 533 için En Küçük Kareler ve İncelenen Yöntemlerin Sonuçları

$\beta_0 = 800$ $\beta_1 = 1$	n = 533		
	En Küçük Kareler	Yeniinden Ağırlıklandırılmış İteratif EKK İterasyon sayısı: 35	Chatterjee ve Mahler İteratif Yöntemi İterasyon sayısı: 2
$\hat{\beta}_0; s\hat{\beta}_0$	800.80005 (2.29933)	800.58718 (0.0005015)	800.891 (2.26014)
$\hat{\beta}_1; s\hat{\beta}_1$	1.00169 (0.00458)	1.00103 (7.744451E-8)	1.00170 (0.00458)
$t_{\hat{\beta}_0}; t_{\hat{\beta}_1}$	354.49; 219.60	1.596E7; 1.293E7	354.35; 219.50
$F; p$ Model	48226.1 <.0001	1.67E14 <.0001	48178.7 <.0001
RMSE	16.49260	13.07464	2.02531
R^2	0.9890	1	0.9890
Coeff Var	1.29684	0.90977	0.15925
AIC	3006.7176	2757.7653	758.5331
DW-IAC	1.864; 0.063	1.963; 0.033	1.864; 0.063
White χ^2 :p	36.02; <.0001	2.01; 0.5710	37.74; <.0001
Artık Çizim			
Normal Eğri ile Artık Sıklık Çizimi			

Tablo 10. (devam...)

$\beta_0 = 800$ $\beta_1 = 1$	n = 533			
	Değiştirilmiş Artık Artıklar	Karesel Artık St. Sapmaları ile Ağırlıklandırma	Bağımlı Değişken St. Sapmaları ile Ağırlıklandırma	Bay-Cox LogY Dönüştürme
$\hat{\beta}_0; s\hat{\beta}_0$	797.13302 (0.59919)	799.01745 (2.18435)	800.84478 (2.13978)	2.93882 (0.00085514)
$\hat{\beta}_1; s\hat{\beta}_1$	1.00550 (0.00120)	1.00392 (0.00462)	1.00064 (0.00452)	0.00034510 (0.00000173)
$t_{\hat{\beta}_0}; t_{\hat{\beta}_1}$	1330.36; 834.83	365.79; 217.12	374.27; 221.23	3436.66; 199.88
$F; p$ Model	696036; <.0001	47141.8 <.0001	48943.4 <.0001	39057.2 <.0001
RMSE	1.30203	0.81919	0.86338	0.00624
R^2	0.9992	0.9888	0.9882	0.9868
Coeff Var	0.10153	0.06444	0.06833	0.20130
AIC	283.8696	-211.7972	-155.4812	-5439.9040
DW-IAC	1.636; 0.068	1.892; 0.049	1.812; 0.088	1.529; 0.234
White χ^2 :p	12.60; 0.0056	12.18; 0.0068	45.33; <.0001	31.32; <.0001
Artık Çizimi				
Normal Eğri ile Artık Sıklık Çizimi				

n = 533 örneklem büyüklüğü için uygulanan yöntemlerin Tablo 10’da verilen analiz sonuçları incelendiğinde, yalnızca kare artıkların değişen varyanslılık sorununu tamamen düzelttiği gözlenmektedir. Bu

yöntem artık sıklık dağılımını da normal dağılıma uydurmuştur. Değiştirilmiş ardışık artıkların ağırlıklandırılmasında kullanılması da kelebek şeklinde dağılan artıklar sorununu ortadan kaldırmıştır. Ancak söz konusu yöntem aykırı değerlere ve artık dağılımının bozulmasına neden olmuştur.

Burada dikkat çeken bir diğer konu ise değişen varyanslılığın belirlenmesinde kullanılan skor testidir. Yeniden ağırlıklandırılmış iteratif en küçük karelerin son iki yapay veri kümesi sonuçları için White testi sonucuna göre değişen varyanslılık sorunu yoktur. Ancak bu yöntem için artık çizimlerine bakıldığında kelebek şeklinde dağılan artıklar dolayısıyla değişen varyanslılık sorunu görülmektedir.

6. İNCELEME VE SONUÇ

Monton olmayan artık varyanslarının dengelenmesi için aynı veri setine farklı 6 yöntem uygulanmıştır. Yöntemlerin üçü tanesi iteratif algoritmalara, biri dönüşümlere kalan iki yöntem ise en küçük kareler analizi sonuçlarına dayanmaktadır. Uygulanan üç iteratif yöntemden biri (Bischoff vd. 2006) değişen varyanslılığı dengeleme noktasında işe yaramıştır. Ancak söz konusu yöntem yapay veriler için değişen varyanslılık sorununu düzeltmesine rağmen aykırı değerlere ve artık dağılımının bozulmasına neden olmuştur. Uygulanan diğer iki iteratif yöntem, hem uygulama verisi içi hem de çeşitli örneklem büyüklüğündeki yapay veriler için varyansı dengeleme açısından iyi sonuç vermemiştir. Özellikle Meloun vd. (2002) tarafından önerilen yöntem normal dağılımı bozma ve çoklubağlantı sorunu yaratmıştır. Bağımlı değişkene uygulanan logaritmik dönüşüm de hem uygulama verisinde hem de yapay verilerde istenen sonucu vermemiştir. Öte yandan alt gruplar bazında bağımlı değişken standart sapmalarının ağırlık tahmini elde etme sürecinde kullanılması, uygulama verisi için iyi sonuç vermesine rağmen yapay veri kümelerinde iyi sonuç vermemiştir.

İncelenen yöntemler arasında hem uygulama verisi hem de çeşitli düzeylerdeki örneklem büyüklükleri için kare artıkların standart sapmalarının ağırlık tahmininde kullanılması, değişen varyanslılık sorununu çözmekle birlikte artık dağılımının normal dağılıma daha çok yaklaşmasını sağlamıştır.

İncelenen yöntemler kolaylık ve işlevsellik açısından incelendiğinde, iteratif yöntemler uygulanabilirlik açısından zor ve istenen sonucu vermekten uzaktır. İteratif olmayan ağırlık tahmin süreçleri, mevcut pek çok istatistik yazılımında bulunduğundan bu anlamda daha tercih edilebilirdir.

Bu çalışmanın önemli bulgularından biri de değişen varyanslılığı belirlemede kullanılan skor testlerinden biri olan White testinin bazı durumlarda yanılgılara neden olduğudur.

KAYNAKÇA

Aitkin, M., “Modelling Variance Heterogeneity in Normal Regression Using GLIM”, *Applied Statistics*, 36, pp. 332-339, 1987.

Albayrak, A. S., “Değişen Varyans Durumunda En Küçük Kareler Tekniğinin Alternatifi Ağırlıklı Regresyon Analizi ve Bir Uygulama”, *Afyon Kocatepe Üniversitesi, İ.İ.B.F. Dergisi (C.X ,S II)* ,s. 111-138, 2008.

Basu, A., “Extended generalized linear models: Simultaneous estimation of flexible link an variance functions” *The Stata Journal*, N 4, pp. 501-516, 2005.

Bischoff, W., Heck, B., Howind, J., Teusch., A., “Aprocedure for estimating the variance function of linear models and for checking the appropriateness of estimated variances a case study of GPS carrier-phase observations”, *J Geod.*, 79, 694–704, 2006.

Box, G. E. P. and Cox, D. R., “An analysis of transformations (with discussion)”, *Journal of the Royal Statistical Society, series B*, 26, 211–246, 1964.

Carroll, R.J., Ruppert, D., *Transformation and Weighting in Regression*, Chapman and Hall, 1988.

Chatterjee, S., Machler, M., “Robust Regression: A Weighted Least Squares Approach”, *Commun. Statist.-Theory*, 26(6), 1381-1394, 1997.

Darlington, R. B., “Correcting the standard errors of regression slopes for heteroscedasticity,” <http://www.psych.cornell.edu/darlington/heterosc.htm> (1997).

Friel, C., M., *Weighted Least-squares Regression*, Houston State University, 2009.

Hamasaki, T. and Goto ,M., “Power-Transformation Model to Additivity in Regression”, *Behaviormetrika*, Vol.32, No.2, 111–125, 2005.

Gujaratti, D., *Basic Ekonometrics*, 2001.

Güriş, S. ve Çağlayan, E., 2001. *Ekonometri*, Der Yayınları, 2001.

Harvey, A.C., “Estimating Regression Models with Multiplicative Heteroskedasticity”, *Ekonometrica*, 44, 461-465, 1976..

Meloun, M., Militky, J., Kupka, K., Brereton, R.G.,. “The effect of influential data, model and method on the precision of univariate calibration”, *Talanta*, 57, 721-740, 2002.

Oberg, A. and Davidian, M., “ Estimating Data Transformations in Nonlinear Mixed Effects Models”, *Biometrics* 56, 65-72, 2000.

Pardoe, I. and Cook, D., “ A Graphical Diagnostic For Variance Functions” *Aust. & N. Z. J. Stat.* 49(3), 241–250, 2007.

PBC Electronics Designing, “The fixed geometry structure’s FFT algorithm and FPGA realize-www.51rd.net”, 2009.

SAS/IML 9.2 User’s Guide,. SAS Institute Inc., Cary, NC, USA, 2008.

Sümer, K. K., “White’ın Heteroskedastisite Tutarlı Kovaryans Matrisi Tahmini Yoluyla Heteroskedastisite Altında Model Tahmini”, *İstanbul Üniversitesi İktisat Fakültesi Ekonometri ve İstatistik Dergisi*, Sayı 4 2006-12-28.

<http://www.rshm.saglik.gov.tr/hki/hkindex/index5.htm>

<http://www.havaizleme.com/r.html>